ISSN: 2088-6578
ISBN: 978-1-4799-0423-5

50th ANNIVERSARY
DEPARTMENT OF ELECTRICAL ENGINEERING AND INFORMATION TECHNOLOGY UGM

IEEE
Advancing Technology for Humanity

ICITEE 2013

Proceedings of

2013 International Conference on Information Technology and Electrical Engineering

7-8 October 2013

The Sahid Rich Jogja Hotel

Yogyakarta, Indonesia

"Intelligent and Green Technologies for Sustainable Development"

Organized by

Department of Electrical Engineering
and Information Technology
Universitas Gadjah Mada, Indonesia

ICITEE 2013 Sponsorships

Technically Co-Sponsored by

Supported by

LPPM UGM

Welcome Message from the General Chair

In Honor of the celebration of the 50th Anniversary of the Department of Electrical Engineering and Information Technology, Faculty of Engineering, Universitas Gadjah Mada (UGM), it is my great pleasure to welcome you to Yogyakarta City for the 5th International Conference on Information Technology and Electrical Engineering (ICITEE 2013) on 7-8 October 2013.

ICITEE 2013 is intended as an International Forum for those who wish to share their latest research results, innovative ideas, and experiences in the fields of Information and Communication Technology (ICT) as well as Electrical Engineering. Nowadays, modern technology makes our lives easier. Yet this progress is affecting our climate as a result of the increased carbon dioxide (CO_2) emissions. Under the theme of **"**Intelligent and Green Technologies for Sustainable Development," the conference is expected to provide opportunities to explore emerging green and intelligent technologies that can contribute to environmental sustainability.

In addition, the conference committee has invited three renowned Keynote speakers, Professor Dr. Tadashi Matsumoto of JAIST, Japan, Emeritus Professor Dr. Susumu Yoshida of Kyoto University and Dr. Eng. Khoirul Anwar of JAIST, Japan. The conference committee also invited Professor Dr. Ramesh Kumar Pokharel of Kyushu University, Japan as Invited speaker to present his current research activities.

This conference is technically co-sponsored by IEEE Indonesia Section and supported by Department of Electrical Engineering and Information Technology UGM.

As a General Chair, I would like to take this opportunity to express my deep appreciation to the organizing committee members for their hard work and contribution throughout this conference. I would also like to thank authors, reviewers, all speakers, and session chairs for their support to ICITEE 2013.

I hope that participants will have a fruitful experience to enjoy the cultural heritage, natural beauty of Yogyakarta, and the taste of traditional Javanese cuisines, coupled with the friendliness of its people.

Finally, I would like to welcome you to ICITEE 2013 and wish you all an enjoyable stay in Yogyakarta.

Sincerely,

Dr. I Wayan Mustika, S.T., M.Eng.
General Chair of ICITEE 2013
IEEE UGM Student Brach Counselor

Welcome Message from the TPC Chair

On behalf of the technical program committee (TPC), it is my pleasure to welcome you to the 5th International Conference on Information Technology and Electrical Engineering (ICITEE 2013). As an annual International conference, ICITEE provides excellent platform to share innovative idea and experiences, exchange information, and explore collaboration among researchers, engineers, and scholars the field of information technology, communications, and electrical engineering.

This year, the ICITEE 2013 Technical Program Committee received 190 paper submissions from about 14 countries throughout the world. All the submitted papers were thoroughly and independently reviewed by at least three reviewers in accordance with standard blind review process. Based on the results of the rigorous review process, 92 papers have been selected. These papers have been grouped into 5 technical sessions, ranging from information technology, communications, power systems, electronics, and control systems. Besides those regular sessions, ICITEE 2013 also features world-class keynote/plenary speeches and distinghuish-invited speaker that reflect the current research and development trends in green and intelligent technology to achieve environmental sustainability.

We are deeply indebted to all of our TPC members, as well as our volunteer reviewers, who have greatly contributed to the success of the ICITEE 2013. Many thanks should be given to our keynote and invited speakers who will present their work in this conference. In addition, our sincere gratitude should be given to all authors who submitted their works to ICITEE 2013 and hope you will enjoy a wonderful experience in this small traditional city of Indonesia.

Welcome to Yogyakarta, explore a thousand years old temples, enjoy its traditional arts and cultures, taste the varieties of traditional Javanese cuisines, and bring them back with your memories of Yogyakarta and new collaboration opportunities.

With best regards,

Eka Firmansyah, ST., M.Eng, Ph.D
TPC Chair

ICITEE 2013 Committee

Advisory Board Committee

Panut Mulyono, Universitas Gadjah Mada, Indonesia
Tumiran, Universitas Gadjah Mada, Indonesia
Dadang Gunawan, Indonesia University, Indonesia
Yanuarsyah Haroen, Bandung Institute of Technology, Indonesia
Hidekazu Murata, Kyoto University, Japan
Lukito Edi Nugroho, Universitas Gadjah Mada, Indonesia
Son Kuswadi, EEPIS, Indonesia
Sarjiya, Universitas Gadjah Mada, Indonesia
Risanuri Hidayat, Universitas Gadjah Mada, Indonesia
F. Danang Wijaya, Universitas Gadjah Mada, Indonesia
Litasari, Universitas Gadjah Mada, Indonesia
Isnaeni, Universitas Gadjah Mada, Indonesia

Organizing Committee

General Chair
I Wayan Mustika, Universitas Gadjah Mada, Indonesia

General Co-Chair
Hanung Adi Nugroho, Universitas Gadjah Mada, Indonesia

Technical Program Chair
Eka Firmansyah, Universitas Gadjah Mada, Indonesia

Technical Program Co-Chair
Noor Akhmad Setiawan, Universitas Gadjah Mada, Indonesia

General Secretary
Indriana Hidayah, Universitas Gadjah Mada, Indonesia

Treasurer
Eny Sukani Rahayu, Universitas Gadjah Mada, Indonesia

Publication
Prapto Nugroho**,** Universitas Gadjah Mada, Indonesia

Publicity
Sri Suning Kusumawardhani, Universidad Politécnica de Cataluña, Spain
Sunu Wibirama, Tokai University, Japan
Fikri Waskito, Nanyang Technological University, Singapore
Canggih Puspo Wibowo, KMITL, Thailand

Sponsorship
Avrin Nur Widiastuti, Universitas Gadjah Mada, Indonesia
Iswandi, Universitas Gadjah Mada, Indonesia

Proceeding
Sigit Basuki Wibowo, Universitas Gadjah Mada, Indonesia

Local Arrangement
Budi Setiyanto
Teguh Bharata Aji
Warsun Najib
Selo Sulistyo
Dani Adhiptha
Silmi Fauziati
Adistya Erna Permanasari
Bimo Sunarfrihartono
Ridi Fediana
Yusuf Susilo Wibowo
Lesnanto Multa Putranto

Technical Program Committee

Chair

Eka Firmansyah, Universitas Gadjah Mada, Indonesia

Co-Chair

Noor Akhmad Setiawan, Universitas Gadjah Mada, Indonesia

Member

Abhishek Tomar, G B Pant University of Agriculture & Technology, India
Addy Wahyudie, United Arab Emirates University, UEA
Adha Imam Cahyadi, Universitas Gadjah Mada, Indonesia
Agya Utama, Surya University, Japan
Alagan Anpalagan, Ryerson University, Canada
Amirthalingam Ramanan, University of Jaffna, Sri Lanka
Andy Warner, Google
Anto Satriyo Nugroho, BPPT, Indonesia
Anton Satria Prabuwono, Universiti Kebangsaan Malaysia, Malaysia
Armein Z. R. Langi, Bandung Institute of Technology, Indonesia
Awinash Anand, Kyushu University, Japan
Azwirman Gusrialdi, University of Central Florida, USA
Boonprasert Surakratanasakul, KMITL, Thailand
Chanboon Sathitwiriyawong, KMITL, Thailand
Chotipat Pornavalai, KMITL, Thailand
Cuk Supriyadi Ali Nandar, BPPT, Indonesia
Dhany Arifianto, Institute of Technology Sepuluh Nopember, Indonesia
Ekachai Leelarasmee, Chulalongkorn University, Thailand
Esa Prakasa, LIPI, Indonesia
F Danang Wijaya, Universitas Gadjah Mada, Indonesia
Gunawan Wibisono, Indonesia University, Indonesia
Gusti Agung Ayu Putri, Udayana University, Indonesia
Haiguang Wang, Institute for Infocomm Research, Singapore
Haruichi Kanaya, Kyushu University, Japan
Hatta Fudholi, Islamic Indonesia University, Indonesia
Heroe Wijanto, Institut Teknologi Telkom, Indonesia
Husni Rois Ali, Universitas Gadjah Mada, Indonesia
I Ketut Gede Dharma Putra, Udayana University, Indonesia
I Nyoman Satya Kumara, Udayana University, Indonesia
I Putu Agung Bayupati, Udayana University, Indonesia
Ida Ayu Dwi Giriantari, Udayana University, Indonesia
Issarachai Ngamroo, KMITL, Thailand
Ivanna Timotius, Satya Wacana Christian University, Indonesia
Iwan Setiawan, Satya Wacana Christian University, Indonesia
Jaziar Radianti, University of Agder, Norway
Kang-Hyun Jo, Ulsan University, Korea
Kazunori Hayashi, Kyoto University, Japan
Kazuto Yano, ATR, Japan
Khoirul Anwar, Japan Advanced Institute of Science and Technology, Japan
Kitsuchart Pasupa, KMITL, Thailand
Koji Yamamoto, Kyoto University, Japan
Kuntpong Woraratpanya, KMITL, Thailand
Mahmoud A. Abdelghany, Minia University, Egypt
Maleerat Sodanil, KMUNB, Thailand
Mamiko Inamori, Tokai University, Japan
Manop Phankokkruad, KMITL, Thailand
Marco Aiello, University of Groningen, Netherland
Mauridhi Hery Purnomo, Institute of Technology Sepuluh Nopember, Indonesia
Muh Niswar, Hasanuddin University, Indonesia
Natapon Pantuwong, KMITL, Thailand

Nawat Kamnoonwatana, Mahidol University, Thailand
Nidapan Sureerattanan, Thai-Nichi Institute of Technology, Thailand
Nihan Tran, University of Melbourne, Australia
Nol Premasathian, KMITL, Thailand
Noopadol Maneerat, KMITL, Thailand
Nopporn Chotikakamthorn, KMITL, Thailand
Olarn Wongwirat, KMITL, Thailand
Pattaracha Lalitrojwong, KMITL, Thailand
Poramote Wardkein, KMITL, Thailand
Ramesh Pokharel, Kyushu University, Japan
Rohana Sapawi, University Malaysia Sabah, Malaysia
Rukmi Sari Hartati, Udayana University, Indonesia
Ruttikorn Varakulsiripunth, KMITL, Thailand
Sakchai Thipchaksurat, KMITL, Thailand
Saiyan Saiyod, Khon Kaen University, Thailand
Selo Sulistyo, Universitas Gadjah Mada, Indonesia
Singha Chaveesuk, KMITL, Thailand
Sisdarmanto Adinandra, Eindhoven University of Technology, Netherland
Somjet Suppharangsan, Burapha University, Thailand
Sompong Valuvanathorn, Ubon Ratchathani University, Thailand
Sooksan Panichpapiboon, KMITL, Thailand
Soradech Krootjohn, KMUNB, Thailand
Sorawat Chivapreecha, KMITL, Thailand
Sumei Sun, Institute for Infocomm Research, Singapore
Sumet Prabhavat, KMITL, Thailand
Supakit Nootyaskool, KMITL, Thailand
Supawan Amanab, KMITL, Thailand
Supot Nitsuwat, KMUNTB, Thailand
Surapan Airphaiboon, KMITL, Thailand
Suvapadee Aramvith, Chulalongkorn University, Thailand
Taworn Benjanarasuth, KMITL, Thailand
Teerapong Leelanupab, KMITL, Thailand
Thanisa Numnonda, KMITL, Thailand
Thippaya Chintakovid, KMUNB, Thailand
Tomoaki Ohtsuki, Keio University, Japan
Tony Dwi Susanto, Institute of Technology Sepuluh Nopember, Indonesia
Uke Kurniawan Usman, ITTelkom, Indonesia
Umar Khayam, Bandung Institute of Technology, Indonesia, Indonesia
Virach Sornlertlamvanich, NECTEC, Thailand
Wawta Techataweewan, Srinakharinwirot University, Thailand
Wayan Gede Ariastina, Udayana University, Indonesia
Wimol San-Um, Thai-Nichi Institute of Technology, Thailand
Widyawan, Universitas Gadjah Mada, Indonesia
Yasushi Kato, Tsuruoka National College and Technology, Japan
Yi Ren, National Chiao Tung University, Taiwan
Yoshimitsu Kuroki, Kurume National College of Technology, Japan
Ziaul Haq Abbas, Ghulam Ishaq Khan Institute, Pakistan

TABLE OF CONTENTS

Technical Sessions

➢ Session 1. Software Engineering, Services, and Information Technology

➢ Session 2. Wireless Communications, Networking and Vehicular Technology

➢ Session 3. Power Systems

➢ Session 4. Electronics, Circuits, and Systems

➢ Session 5. Control Systems

Session 1

Software Engineering, Services, and Information Technology

A Comparison of Effectiveness of Risk Data Clustering Method in Psychiatric Patient Service

Khaengkai Compapong and Sumonta Kasemvilas

Department of Computer Science,
Faculty of Science, Khon Kaen University
Khon Kaen, 40002, Thailand
khaengkai@gmail.com, sumkas@kku.ac.th

***Abstract* — In this paper, we clustered clinical risk data of a mental health service, Khon Kaen Rajanagarindra Psychiatric Hospital. This study aims to compare performance values of cluster (k) in k-means clustering algorithm and hierarchical clustering algorithm. The result shows that for k-means clustering algorithm, sum of squared error (SSE) is 32.68, minimum of distance (MD) is 1.38, mean squared error (MSE) is 2.95 and values of k is 11. Therefore, we found that k-means clustering algorithm is the most appropriate method for using in cluster the risk group of the Psychiatric Patient Service. The result also suggests that the most risky age is between the ages of 32 and 36. The result can be a guideline for further research about data prediction. The implications of this study can assist medical staff to be knowledgeable about what should beware of when they treat psychiatric patients and this can be basic planning medicate guidelines for medical staff.**

Keywords—k-means clustering algorithm; hierarchical clustering algorithm; squared Euclidean distance; risk data; hospital; mental health service; data mining

I. INTRODUCTION

From the statistics conducted by Department of Mental Health, Ministry of Public Health revealed the number of Thai psychotherapy patients during fiscal year 2011 that averagely there are 4,456 outpatients and 33.18 hospital patients per day. In the fiscal year 2012, the average number of outpatients was 4,308 patients and the average number of hospital patients was 30.23 patients per day. Psychotherapy patients are those who tend to risk depressive disorder and suicide. When considering the statistics in northeastern part of Thailand, Khon Kaen Rajanagarindra Psychiatric Hospital, the case study, founded that the average number of outpatients in the fiscal year 2011 was 2,721 and of hospital patients was 370 per day. In the fiscal year 2012, the average numbers of outpatient and hospital patient were 2,965 and 394 per day, respectively. Male patients are more than female patients in both fiscal years. The most frequent age span was during 30-34 years old [1]. Since the psychotherapy patients do not show obvious symptoms, the health care providers must be careful and try their best to reduce the risk. The main variables making the patients receive different cares are genders and age [2], [3], [4]. Thus, the risks found in the hospitals are derived from the risk from the patients themselves, the risk occurring during the treatment, and the risk from the environment surroundings. If the risks are collected by the reliable method, we will be able to elicit a plenty of them. Several sectors have attempted to solve the problem, especially the large health care service providers where a lot of service information are stored.

It is difficult to acquire the clinical risk data to analyze, depending on the size of the health care organization; the bigger, the harder [5]. Knowing the clusters facing with the risk in psychiatric hospital done by grouping genders and ages will affect the service quality resulting in the patient satisfaction [2], [3]. Therefore, this study used k-means clustering algorithm, which is an appropriate data mining method for medical research [5], [6], [7]. However, k-means clustering algorithm has its limitation which is deviation when grouping in terms of the cluster size, density, and unformed clustering data. This limitation can be solved by analyzing the appropriate cluster to decrease the deviation when presenting the data [4], [5].

This study brought the clinical risk data of a mental health service to examine and compare the precision during the analysis by k-means clustering algorithm. Some data were analyzed by using hierarchical clustering algorithm to obtain the highest precise value, leading to the obvious clusters of risk patients. This also shows the performance of the two methods of which one will be the most suitable method for analyzing the clinical risk data.

II. THEORIES AND LITERATURE REVIEW

A. Clinical Data Clustering Analysis

Analyzing the clustering of clinical data is the process of categorizing data into appropriate groups; that is to say, the data with the same characteristics must be in the same group. Clustering is different from classification; in other words, the data are divided by their similarities without specifying the types. This can be called unsupervised classification. Clinical researches employ this technique to cluster patients by severity of symptoms and diseases, by the factors of disease and by dispensation [8], [9]. The quantitative data will be brought to find the groups' values and built the work system responding to the medical supplies department. Furthermore, the data can be clustered due to the statistics of people who came to receive the services, and then the result can be utilized in terms of public health management [6], [10], [11], [12], [13]. The main purpose of clustering clinical groups as previously mentioned

is to use those data when planning the clinical service for the patients.

B. Risk Management in Mental Health Service

Many hospital and health care service providers place importance on risk management focusing on patient service. The highest risk variables are staff and service receivers. The risks may happen during these processes: seeing the doctors, evaluating mental status, and admitting to give treatment [8]. These steps are related with patients, medical personnel, and strategic risk in terms of health care provider's management. In risk assessment, we examine the probability of risk event since the patient is admitted until they leave the hospital.

C. K-Means Clustering Algorithm

K-means clustering algorithm is an unsupervised technique used when the number of data is large. Clustering algorithm divides data into groups [7], [8] which will be represented by mean. Mean from each group is a cluster centroid used for measure the distance among the data of the same group. First step of clustering can be done by setting up the required group value and finding the mean. Then, determining the cluster centroid at N point. Most importantly, when determining the cluster centroid, the method must be appropriate, because the different beginning point of cluster centroid will lead to different results.

When using k-means clustering algorithm, users should choose the numbers of cluster in advance. In case of the number of cluster still be doubted, the analyzer may employ one of these techniques 1) Analyze by using k-means clustering algorithm several times. Choose the numbers of cluster differently in each time, and then find the most appropriate number; 2) Analyze some data by using hierarchical clustering algorithm to find the possible number, then adopt k-means method to those data.

D. Steps of K-Means Clustering Algorithm

Step 1: Find the optimize values of group or randomly choose k object as the initial clusters identify the starting centroid at N point.

Step 2: Find the biggest distance between data and centroid. The data are arranged into groups according to the nearest centroid they are from. Calculate distance by using the distance function between two points $a = (x1, y1)$ and $b = (x2, y2)$ is defined as :

$$Distance\ (a, b) = |x2 - x1| + |y2 - y1| \quad (1)$$

where, a is the number of point, b is the number of mean, distance value$) = x1, y1)$ and centroid= $)x2, y2)$

Step 3: Find mean for each group and make it as a new centroid by using data distance formula with the centroid as the following:

The Minkowski distance is a metric on Euclidean space which can be considered as a generalization of both the Euclidean distance and the Manhattan distance. Function of Minkowski distance is:

$$d(i,j) = \sqrt[q]{(|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + ... + |x_{in} - x_{jn}|^q)} \quad (2)$$

q is dimension relevant of dataset, when $q = 1$, d called Manhattan distance; when $q \leq 2$, d called Euclidean distance. The distance between points i and j is the length of the line segment connecting them. In Cartesian coordinates, if $i = (x_{i1}, x_{i2}, ..., x_{in})$ and $j = (x_{j1}, x_{j2}, ... x_{jn})$ are two points in Euclidean n space, then the distance is from p to q or q *to* p .

Step 4: Repeat to step 2 until the centroid is unchanged or (re) assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster.

Step 5: Evaluate k-means clusters group whether it has SSE for each point which goes to the group by calculating from

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} dist^2(m_i, x) \quad (3)$$

Where x is the data point in the cluster C_i , m_i is the centroid in the cluster C_i.

E. Steps of Hierarchical Clustering Algorithm

This algorithm can be divided without knowing the exact values of groups and without knowing each case grouped by:

Step 1: Create data file which contains cases or many variants which can be used for clustering case group or variant group.

Step 2: Choose the case or randomly choose the case for analyzing.

Step 3: Cluster the data using hierarchical clustering algorithm.

III. METHODS AND EXPERIMENT

A. Experimental Design

This study employed the risk mental health services data from Khon Kaen Rajanakarin Psychiatric Hospital, Mental Health Department, Ministry of Public Health, Thailand which is a big specialized hospital providing mental health services in the northeastern of Thailand. The researcher asked for a permission from the research committees of the hospital to conduct the study in order to collect risk data and also requested the risk data from a patient service system which kept an historical data by collecting the attributes as follows: patient code, gender, age of person in risk, career, risk

situation, risk, and medicine allergy. Gender and age of patients are key factors that affect mental health services. We will focus on gender and age in the experiment. [2], [3], [4]. The data will be analyzed by using data mining which the data has to be clustered.

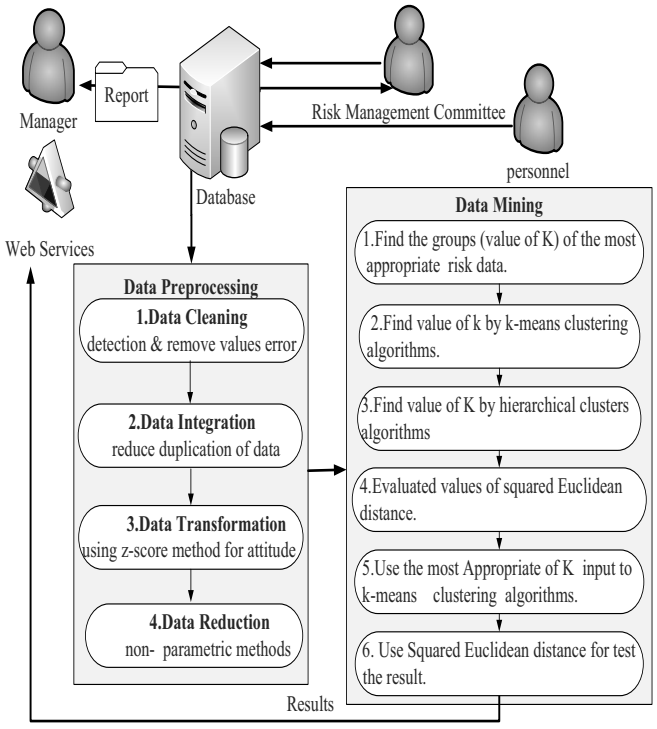


Fig. 1. The framework of risk mental health services data

Fig.1 shows the framework of risk mental health services data in this study. First, the staffs found the risk data from the health care service, then they recorded it via the risk management program. When the data from risk management committee were obtained, they would be collected in risk database. If the data were not completed or missing, they would be deleted from the examination by the user. When risk data were ready, the preprocessing step was launched and then data mining began.

B. Dataset

This research had developed the original model from information technology system and obtained the raw data by screening using the work system of case study-organization. The results appeared in a form of 270 rows of risk data and 1,877 rows from patient management program within 3 months. All data were processed in data preparation.

1) Data Preprocessing: The data were completely cleaned, but not be ready to be used for the data mining. Thus, the data were stored for adjusting. In this study, clustering was used to detect and delete outliers. The data then were ready to be used in the algorithm. The data were divided into two sets:

- The risk data set collected from the model and risk data from patient management program altogether 2,147 cases.
- The risk data set collected from random **f**unction in SPSS which gave different and unequal numbers of case in each round.

TABLE I. RESULTS OF DATA PREPROCESSING

Attribute	Number	Min	Max	Mean	SD
age	1,875	3	91	41.87	18.283
sex	2,005	1	2	1.42	.494

Attributes from the data set used in the experiment consisted of quantitative variables as shown in Table I. The detecting and managing outliers by clustering results found 2,005 rows of absolute value in attributes consisting of the male = 1 , the female = 2 and age = 3, 4, ..., 91. All numeric value will be displayed in the forms of mean, standard deviation, and missing value which rows were cut off 327 rows. In terms of redundancy, in this set of data, the redundant data were not found since the preprocessing via risk management program was thoroughly checked. Function Z-score in SPSS was used to change the variable data to have the minimum value at 0 and maximum at 1 since the weight caused by different clustering must be eliminated the accuracy when grouping reduced data. In this study, the two attributes were put into clustering model. Then, the histogram was examined and it was found that the size of data set was not too complicated.

In this data set, attribute was not duplicate, because there was the thoroughly step by step data input process through the risk data managing program. Z-score function in SPSS is used to transform the value of attribute to 0 or more than 0 but less than 1 (0 is the lowest value, and 1 is the highest value), because the weight is used to limit the clustering group to make the most accurate in reducing data group clustering. In this study, there are two attributes imported into the clustering model. From the Histogram, the size of the data set does not contain too much complexity.

Input : data set of case
Output : the key of output is partition , Number of cluster
1. Input data set // case of data 2. Make initial guesses for the means $m_1, m_2, ..., m_k$ 3. Until there are no changes in any mean 4. a. Use the estimated means to classify the samples into clusters 5. b. For i from 1 to k 6. Replace m_i with the mean of all of the samples for cluster i 7. End_for 8. End_until

Fig. 2. Clustering Algorithm for clinical risk data of a mental health service

2) Data Mining: According to Fig.1, it is the process of risk data group clustering in mental health service by using k-means clustering algorithm without knowing data group, so

there are the comparison of the accuracy between several group value specification and the use of some data by hierarchical clustering algorithm, and then use squared Euclidean distance and evaluate the group to find the least SSE. This study employed medical data group clustering. The researcher emphasized the technique in clustering relationship of mental health services data into small groups. The values of attribute in the same group or the same factor have to be related and they will be brought to the cluster analysis. When the results come, they will be sent to the web service module. According to Fig.2, the appropriate of *k* from comparison of the effectiveness of clustering method risk in Psychiatric Patient Service dataset is *n* sample feature, vector values are *x1, x2, ..., xn,* all from the same class, and we know that they point into *k* compact clusters, when $k < n$. Let m_i be the mean of the vectors in cluster i. If the clusters are well separated, we can use a minimum distance classifier to separate them. That is, *x* is in cluster *i* if || *x* - *mi* || is the minimum of all the *k* distances [14].

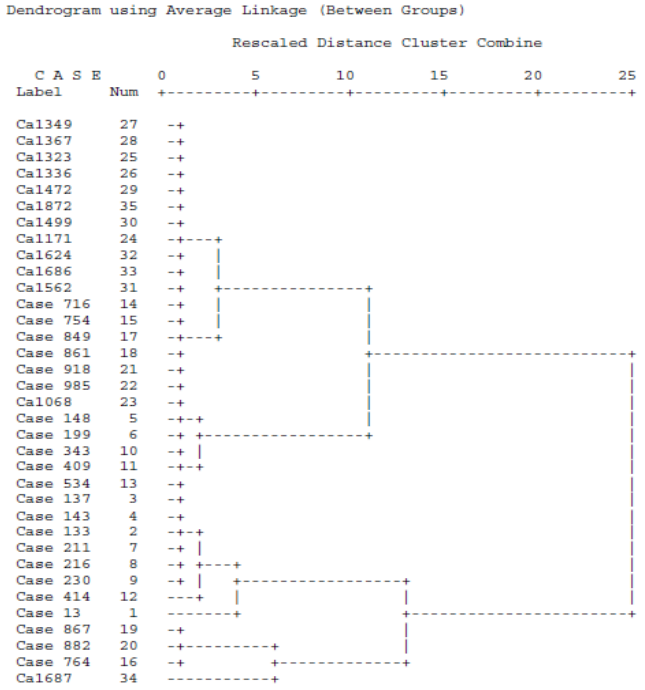


Fig. 3. Analyzing results by using hierarchical clustering algorithm

TABLE II. RESULTS OF K BY HIERARCHICAL CLUSTERING ALGORITHM

Item	Number of Case	Value of k
Test 1	451	10
Test 2	78	9
Test 3	45	8
Test 4	70	8
Test 5	20	8

C. Hierarchical Clustering Algorithm to find k

Some parts of data were analyzed to find the clustering value by using hierarchical clustering algorithm to grouping the data of risky patients. When the groups were clustered, the dendrograms were adopted as shown in Fig. 3. Group analysis helped create appropriate numbers of clusters. The data set was randomized by SPSS program, which gave the unequal and different numbers of case in each round. This was how to prepare the second set as shown in Table II. It was found that the appropriate number of clusters was 8 after 3-5 times of clustering experiment. The difference among groups was statistically significant at level 0.05.

D. K-Means Clustering Algorithm to find k

K-means clustering algorithm organized data into appropriate groups by finding the minimum distance between the data and the finding the centroids of each cluster. With this method, the numbers of cluster would be variously determined such as 7, 8, or 20 and the most appropriate number would be chosen. The researcher brought risk data obtained from the model to combine with those from risk management program, altogether 2,147 cases. All data were standardized whose two variables were analyzed: genders and ages or risk group. The squared Euclidean distance was applied to measure. Then, means of each group were examined and made to be new centroids using the formula that find the distance between data and centroids. The researcher determined the cluster value at 20 and the maximum number of round at 100 rounds.

E. Clustering Efficiency Comparison

We used SPSS to check cluster centroid analyzing. If the program found the stable value, the program will stop calculating. The result obtained from the experiment is that the 34^{th} cycle has no centroid cluster change at 0.000 with the least distance between centroid at 1.38. According to the experiment, it is found that the new centroid in each cluster, and the optimal values that have the distance with the closest point value were cluster 9 and cluster 11 with distance value of 1.38 as seen in Table III.

IV. DISCUSSIONS

The experiments had proved the efficiency of each method. When finding the cluster value in a big data set, the appropriate beginning is important because the result can be presented to the administrator who manages the hidden risk in the hospital. The result of first method was conducted by randomizing the risk data's rows or case and repeated five rounds through hierarchical clustering algorithm. When the experiment ran to the 3^{rd}, 4^{th} and 5^{th} round, k-value was equal at 8. Then, the process of finding clusters was done by using k-means clustering algorithm with the first data set. K-value was set at 8, and the result was the first cluster. There were 321 service receivers (in the age of 36) who involve the service risk. The nearest distance, which was 7.77, between centroids was from cluster 3 and 8. The k-means clustering was evaluated to see the errors in clustering as shown in Fig.4 and Mean square of age was 7.49. The first data set was taken

to data mining process and was analyzed by k-means clustering algorithm. In each time, the numbers of cluster were determined differently such as 7, 8 or 20. Then, the appropriate number of cluster was considered by looking at the nearest distance between centroids. In this case, the nearest distance was between cluster 9 and 11 (1.38). The result pointed at cluster 5. The service receivers who were 32 and involved the service risk were the highest population (263 persons). The experiment found the most appropriate k-value which was 11. K-means clusters were evaluated to find the errors as shown in Fig.5. Means square error of age was 2.95.

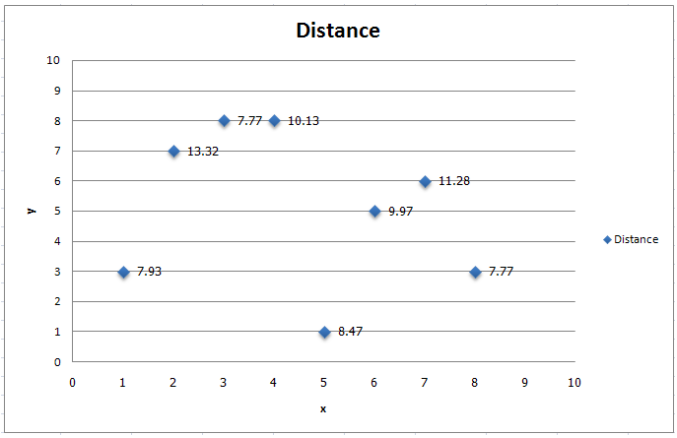


Fig. 4. Trends values of distance obtained, where *k* is 8 and aims to minimum distance

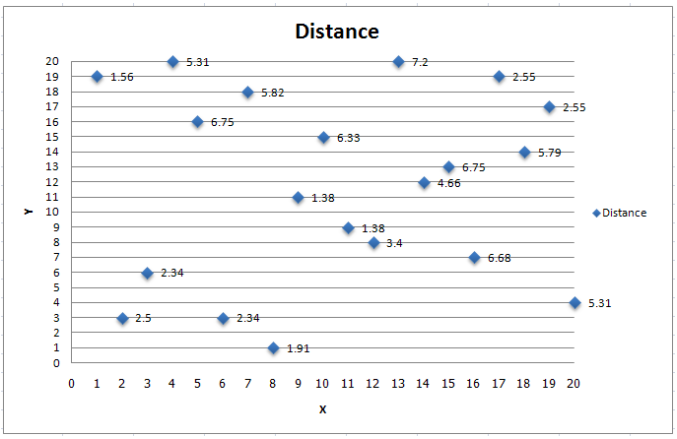


Fig. 5. Trends values of distance obtained, where *k* is 11 and aims to minimum distance

TABLE III. EXPERIMENTAL RESULTS

Method	Value of k	Loop of Cluster	SSE	MSE	MD
1	8	21	87.49	7.49	7.77
2	11	34	32.68	2.95	1.83

Table III shows the comparison of *k* from two methods between 1) k-means clustering algorithm and 2) hierarchical clustering algorithm. They were clustered to find out the service receivers who were in a mental health service risk group. The first method obtained *k*-value at 8. The value was clustered to find the risk group. After calculation, it was found that the distance between points was constant at the 21st round. The SSE was 87.49; the MSE was 7.49; and the minimum distance was 7.77. In terms of k-value calculated by the second method, k was 11. The round that the distance between points is stable was the 34th round. The SSE was 32.68, MSE was 2.95 and MD was 1.83. Therefore, the most appropriate method for using in cluster the risk group of Psychiatric Patient Service was k-means clustering algorithm. The finding was concluded by considering the SSE, MSE and MD. To find SSE, the errors were squared and summed.

V. CONCLUSIONS

This paper employed data mining to report the hidden risk group in order to utilize the finding in planning the mental health service work. The data from the analysis could be applied to hospital management to decrease the risk in terms of service work due to the efficient data mine. The finding also indicated the most effective clustering method which was fast and accurate. The analysis of cluster value should be done by determining value several times; the value will be more accurate, considering from the lower mean square error. Hierarchical clustering algorithm method has its limitation. It spends more time. The measurement used in clustering risk data analysis was Squared Euclidean Distance. The result of measuring was the minimum distance was 7.77 as shown in Fig 4 and 1.38 as shown in Fig 5. This measurement indicated the appropriate distance when finding the proper cluster value in each cluster. The measurement of efficiency by finding the lowest mean square error resulted in cluster evaluation which was 11. The experiment reported the correct cluster value. It was found that the most risky age was during 32-36. This study still has a limitation, which is about the service risk on risk collecting. However, the research needed to collect the risk data from the case study organization that has the restriction in providing data because some of them can affect the patients' rights. Also, the evaluation of the overall image of risk management cannot be conducted due to some restrictions of the organization. Furthermore, this study had divided the patients only in mental health department, which has a lot of variables affecting the patient service.

The findings from this study will be close to the statistics from the medical record department of the hospital. Its scope can be a guideline for further research about data prediction. The risk data can be further improved by clustering the data coming from other sectors such as psychiatrist, nurses and psychologist who record the details of the patients' behaviors. This will make the data become more trustworthy. Moreover, the research can be useful when its result from the analysis is used as a tool to monitor the risk occurring in mental hospital. The findings can also be applied to risk management to make the mental health management become more effective.

ACKNOWLEDGMENT

In this research, the author would like to thank you to Mr. Prapat Ukranan, the Psychiatrist of Khon Kaen Rajanakarin Psychiatric Hospital as well as gratitude to the risk management committees and officers in the hospital who provided all useful information.

REFERENCES

[1] Department of Mental Health, Ministry of Public Health has revealed a number of Thai psychotherapy patients [Online]. Available: http://www.dmh.go.th/report/report1.asp, Access 2012.

[2] Rungdaw dedokmai, Kaiwit Sutameang, "The satisfaction with quality of service as a private in an emergency hospital in Bangkok and vicinity", Journal of Finance and Investment, Marketing and Business Administration Volume **2** Issue, 2011.

[3] Cattleya Nutavong, "Factors Influencing the use of Health-Check up Service Provided by a Hospital in Bangkok and Metropolitan Area", Thailand, 2010.

[4] Antidepressant used in hospitals, [Online]. Available: http://www.hiso.or.th/hiso/picture/reportHealth/hosp_antidepressants_use-PhaReD.pdf Access 2009.

[5] Shusaku Tsumoto, Shoji Hirano, and Yuko Tsummoto, "Towards Data-Oriented Hospital services: Data mining -based Hospital Management", 2010 IEEE International Conference on Data Mining Workshops, Japan , 2010. pp.1078-1068.

[6] Chin-Tsai Lin, Hui-Yin Tsai, " Hierarchical Clustering Analysis Based on Grey Relation Grade", Information and Management Sciences Volume 16, Number 1, pp.95-105, 2005

[7] Yu-Sheng Li and Ying-Chih Chuang , "Neighborhood Effects on an Individual's Health Using Neighborhood Measurements Developed by Factor Analysis and Cluster Analysis", J Urban Health ,January 2009, Pages: 5–18.

[8] Yongyang Huo, Francisco Azuaje, Paul Mc Cullagh, and Roy Harper "Semi-Supervised Clustering Models for Clinical Risk Assessment", UK, 2006.

[9] Risto Vaarandi," A Data Clustering Algorithm for Mining Patterns From Event Logs", Estonia, 2003.

[10] Shusaku Tsumoto, Shoji Hirano, Yuko Tsumoto and Hidenao Abe, "Temporal Data Mining in History Data of Hospital Information Systems", IEEE/ICME International Conference 2011, Japan, 2011. Pages. 2350-2356.

[11] Erich Teichmann, Eren Demir, Thierry Chaussalet, "Data preparation for clinical data mining to identify patients at risk of readmission", Health and Social Care Modelling Group, Department of Information Systems and Computing, School of Informatics, University of Westminster, London, W1W 6UW, UK, 2010. Pages.184-188

[12] Jinmei Feng, Zhimao Lu, Peng Yang and Xiaoli Xu, "A K-means clustering algorithm Based on the Maximum Triangle Rule", Proceedings of 2012 IEEE International Conference on Mechatronics and Automation August 5 - 8, Chengdu, China, 1989, Pages 1456-1461.

[13] Zhexue Huang , "A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining", Australia, 1997.

[14] A Tutorial on Clustering Algorithms [Online]. Available: http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html, Access 2013.

A Kinetic Energy-based Feature for Unsupervised Motion Clustering

Suthasinee Nopparit*, Natapon Pantuwong†
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
Email: *suthasinee.nop@gmail.com, †natapon@it.kmitl.ac.th

Masanori Sugimoto
Department of Computer Science
Hokkaido University
Hokkaido, Japan
Email: sugi@ist.hokudai.ac.jp

***Abstract*—Motion databases usually contain sequences of movements and searching these vast databases is not an easy task. Motion clustering can reduce this difficulty by grouping sample movements into various motion groups containing similar actions. The pose distance is often used as a feature during motion-clustering tasks. However, the main weakness of this strategy is its computational complexity. Query motions are also required to cluster motion sequences. To address these problems, we propose a motion-clustering algorithm based on the use of kinetic energy to cluster sample motions. Our method does not require query motions during the clustering process, so the clustering results can be generated without supervision. Our experimental results confirmed that our proposed method delivered comparable performance to pose distance-based methods, while its computational complexity was significantly lower than that of existing methods.**

I. Introduction

Recently, motion-capture data has been used by various applications that require human or animal animation. Several motion repositories are available on the web such as the Carnegie Mellon University (CMU) Graphics Lab motion-capture database [1]. The emergence of consumer-level devices that capture the 3D positions and geometry of physical objects, such as Microsoft Kinect™, means that motion-capture systems are now affordable. Users can easily access motion-capture systems to obtain their desired motion data. However, motion databases are usually very large and the number of available motion sequences may be in the several hundreds or thousands. It is not easy to search or retrieve data from these large databases. Thus, motion clustering is one of the methods used to overcome the difficulties of motion retrieval. This method can group example movements into different motion groups containing similar actions, which allows users to browse and search for their desired motion more easily.

Motion clustering generally requires supervision. Thus, existing algorithms [2], [3] require a reference motion for each cluster. Frame-by-frame, these algorithms calculate the distances between poses for each motion in the database relative to reference motions. This strategy has several limitations. The first is that users need to provide reference motions for each cluster. It may be difficult for users to find a good reference motion in the database. If reference motions are not provided, however, we have to calculate the distances for every combination of motions, which requires considerable computation time. The second limitation is the computational complexity of the distance calculations. The distances are calculated in frame-by-frame comparisons so the computational complexity of this strategy is quite high. The third limitation is that the two motions generally need to be exactly the same length. It may be difficult to find two motions with the same number of frames. Dynamic Time Warping (DTW) algorithms [4] can be used to align two motions in temporal space (if they are not the same length), but this approach increases the computational complexity of the motion-clustering process.

Given the limitations of the existing methods discussed above, there is a need for a new method for calculating the distance between two motions. This method should satisfy the requirements of simplicity, speed, and performance. The calculation should be completed within a short processing time with no requirement for any information from users, but the accuracy should still be comparable to existing algorithms.

In this paper, we propose the use of kinetic energy as a clustering feature. The kinetic energy represents the amount of movement in the motion data during each segment. Each motion type has a unique kinetic energy characteristic in each segment, so this feature can be used to describe the motion. For example, the kinetic energy of arm segments is relatively high during a punching motion compared with a kicking motion. Thus, we need to calculate the kinetic energy for each degree of freedom (DOF) in each frame of the motion data, before the feature vector can be calculated. Each element in the feature vector is an average value of the energy calculated for each joint. We define each motion as a single feature vector, so the distance between motions can be computed using a single calculation. However, each motion contains many DOFs so we have to apply a Principal Components Analysis (PCA) algorithm [5] to reduce the dimensionality of the feature vector. Finally, the feature vector can be used for clustering. In our implementation, we cluster all of the motions using a clustering algorithm based on the Gaussian Mixture Model (GMM) [6].

The contributions of this paper are shown below.

1) The feature vector proposed in this paper is suitable for unsupervised motion clustering. Users are not required to provide any reference motions, which is the case with previous methods.
2) The method used to obtain the feature vector has a low computational complexity. Therefore, the clustering results can be obtained in a short processing time.

This paper is organized as follows. Section 2 reviews the details and limitations of related studies. Our proposed kinetic energy-based feature is described in Section 3. Section 4 presents and discusses the experimental results, while our conclusions are provided at the end of this paper.

II. Related Work

This section describes previous studies that address the features used to calculate the distance between two motions, which can be used for motion clustering.

Typically, the pose distance is used to calculate the distance between two motion sequences. In [2], an algorithm was introduced for calculating the distance between two frames. However, this algorithm requires that two motions are exactly the same length. This method measures the distance between two frames using a point cloud driven by a skeleton. Ideally, this point cloud is a downsampling of a mesh that represents a character. The distance between frames is calculated by computing the weighted sum of the squared distances between the corresponding points in the two point clouds. This pose distance was also used in [3]. Pose distance calculations are used to cluster movements into motion groups, which are used for parametric motion synthesis. The clustering results are obtained by determining the DTW result for each motion in the database relative to user-defined reference motions.

A method known as "uniform scaling" was proposed in [7], which can handles the problem of two motions with a different number of frames. Similar regions in two motions can be detected using this method. This strategy can increase the flexibility of the method discussed in [3]. However, user-defined reference motions are still required.

A method for representing the actions found in motion data was proposed in [8] pro. The basic concept used by this method is that motion data contains several actions, which can be arranged into sets of actions. This method aims to segment each motion into a set of actions (*n*-grams). Each motion can then be represented as a histogram of overlapping event n-grams. The distance between two motions can be calculated based on the difference between the histograms representing each motion. This method was proposed originally for computer vision applications, but it can also be used for motion clustering. However, reference motions (in this case, video) are also required to define the template for each action.

The above examples can be used to cluster motion sequences, but all of these methods require supervision. Users need to provide suitable reference motions before performing the motion-clustering process. Finding appropriate reference motions may be too difficult for some users, so several researchers have proposed features for calculating the distance between two motions, which can be used to perform motion-clustering tasks without supervision.

A pose distance function was proposed in [9]. This function can also be used for motion clustering. The pose distance is calculated based on the difference between the orientations of each pair of corresponding joints in two motions. This algorithm produces several motion clusters and each cluster contains several motion fragments with similar poses. DTW calculation and reference motions are not required by this algorithm. Frame-by-frame comparisons are required where all combinations of motion sequences are determined to complete a motion-clustering task.

The algorithm presented in [10] divides a human character model into three partitions, which reduces the spatial complexity, and the temporal similarity of each partition is measured using a self-organizing map and the Smith–Waterman algorithm. The overall similarity of two motion clips can be determined by integrating the similarities of the separate body partitions. The hierarchical clustering method is used to produce the clustering result. This method is unsupervised but it cannot distinguish two motions if their poses are not significantly different. This method also requires a very long processing time to produce the self-organizing map.

III. The Proposed Method

All of the existing algorithms mentioned in the previous section fail to satisfy the requirements stated in Section 1. Some of these methods are unsupervised but they require a long processing time and their accuracy is not high. In this section, we present our proposed feature vector for motion clustering, which is based on an assumption that each motion sequence contains a unique amount of movement in each joint. The amount of movement can be determined based on the kinetic energy in each joint. Thus, our proposed feature vector contains the kinetic energy information for each joint in the motion data.

A. A Kinetic Energy-based Feature

The kinetic energy can be calculated by determining the velocity of each DOF in the motion data. Let N_f be the number of total frames in the motion data that is being evaluated. Each frame in the motion data contains information related to the joint angles, root orientation, and root coordinates. Thus, each motion can be represented as a set, F, where $F = \{f_i | i = 1, 2, 3, ..., N_f\}$. We assume that the motion data has D DOFs. Therefore, each frame, f_i, can be represented as $[f_i^1, f_i^2, f_i^3, ..., f_i^D]^T$, where each element in this vector represents the value of each DOF.

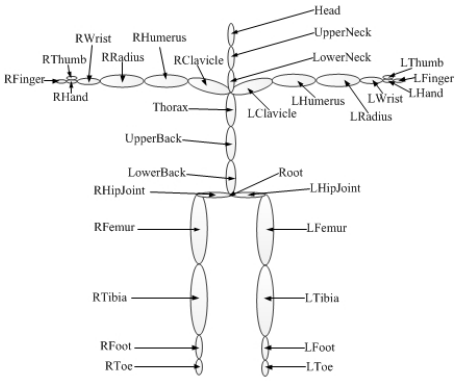
The proposed kinetic energy-based feature is calculated from the average kinetic energy of each DOF in all frames of the motion data. Thus, the proposed method generates D dimensional feature vectors, where each element of the feature vector is the average kinetic energy value. The kinetic energy is related to the velocity, so the first step is to calculate the velocity of each DOF in each frame of the motion data. The velocity of the d^{th} DOF in the i^{th} frame is calculated as follows:

$$v_{i,d} = \frac{f_{i+1}^d - f_i^d}{t_f} \tag{1}$$

where t_f is the duration between frames.

After the velocity of each DOF has been calculated in each frame, we can calculate the kinetic energy of each DOF in such frame. The kinetic energy of the d^{th} DOF in the i^{th} frame, $E_{i,d}$, is defined as follows:

$$E_{i,d} = w_d \times v_{i,d}^2 \tag{2}$$



Bone	DOF	w_d
Root (position)	3	1
Root (rotation)	3	5
(L,R)HipJoint	0	0
(L,R)Femur	3	4
(L,R)Tibia	1	4
(L,R)Foot	2	1
(L,R)Toe	1	1
LowerBack	3	3
UpperBack	3	3
Thorax	3	3
LowerNeck	3	2
UpperNeck	3	2
Head	3	1
(L,R)Clavicle	2	4
(L,R)Humerus	3	4
(L,R)Radius	1	3
(L,R)Wrist	1	2
(L,R)Hand	2	1
(L,R)Finger	1	1
(L,R)Thumb	2	1

Fig. 1: Left: The human skeleton model used to capture motion data from the CMU motion database. Right: The DOFs and manually defined weighting values for each joint, which were used to calculate the kinetic energy-based feature vectors.

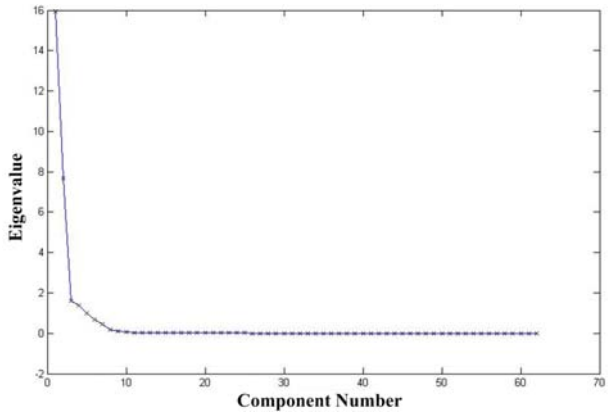


Fig. 2: Eigenvalues for the covariance matrix of the calculated kinetic energy-based feature vector.

where w_d is the weighting value for each DOF. The value of w_d varies for each joint. If a joint's movement affects the overall motion (e.g., a knee or elbow joint), it should have a large weighting value. In our implementation, we used example motions from the CMU motion database. Figure 1 shows the name, DOF, and w_d for each joint, which were defined manually based on the conditions described previously. These values yielded reliable results in our study. However, it might not be convenient for users to define the weighting values manually. Therefore, we propose to use the length of a bone as a criterion for defining the weighting value automatically. The weighting values of a root joint are fixed at 0.2 and 1.0 for the position and rotation information, respectively. The weighting values of other joints can be defined using the following equation:

$$w_d = \frac{l_B^d}{max(l_B)} \tag{3}$$

where l_B^d is the length of the bone for the d^{th} DOF and $max(l_B)$ is the maximum length of all bones in the skeleton. Therefore, the longest bone is given higher priority than shorter bones. This is reasonable because a change in the pose of the longest bone will have a bigger effect on the pose of the whole skeleton.

Finally, the kinetic energy-based feature vector, $E = [e_1, e_2, e_3, ..., e_d]$, can be calculated as the average energy of each DOF. Each element in the feature vector, e_d, is the average value of the energy calculated for the d^{th} DOF in all frames of the motion data.

$$e_d = log((\frac{1}{N_f}\sum_{i=1}^{N_f} E_{i,d}) + 1) \tag{4}$$

We apply a logarithm operator to the average energy, which emphasizes the difference in the calculated energy of different motions. The logarithm operator $log(x+1)$ is used to handle joints with zero energy.

B. Motion Clustering Using the Kinetic Energy-based Feature

In general, motion data contains a large number of DOFs (e.g., 62 DOFs in the CMU motion database) and it is difficult to cluster such high-dimensionality feature vectors. Therefore, we reduce the dimensionality of the feature vectors using a dimensionality reduction method based on the PCA algorithm. PCA is a method used to identify patterns in data, which represents data in a manner that highlights similarities and differences. Initially, we reduce the dimensionality by data normalization, which involves the subtraction of the mean (the average in each dimension) from each of the data dimensions. Next, we calculate the covariance matrix of the data to represent the relationships between each of the data dimensions. We then compute the eigenvectors and eigenvalues of the covariance matrix. The eigenvector with the highest eigenvalue is the principal component of the dataset. We can select the number of reduced dimensions by determining the eigenvalues. All eigenvectors with associated eigenvalues that are relatively high compared with other values can be used to derive a new dataset. Figure 2 shows the eigenvalues for all of the dimensions of the calculated kinetic energy-based feature vector. This figure can be used to guide the selection of a suitable number of dimensions for the new dataset. The figure shows that the eigenvalues of the first three principal components are reduced dramatically, so we can select the eigenvectors associated with the first three eigenvalues for dimensionality reduction. The final step is to project the initial feature vector onto the selected principal axes. Figure 3 shows the visualization of all the example motions in 3D space. We calculate the feature vector for each motion using the kinetic energy and we reduce the dimensionality of the feature vectors to three using the methods described above

After we have reduced the dimensionality of each of the feature vectors, we can use this new dataset to calculate the distance between two motions and to perform motion clustering. The difference between motions can be determined by calculating any type of distance metric, such as the Euclidean distance. Any type of clustering algorithm can be used to generate the motion groups during motion clustering. In our implementation, we used a motion-clustering method based on GMM.

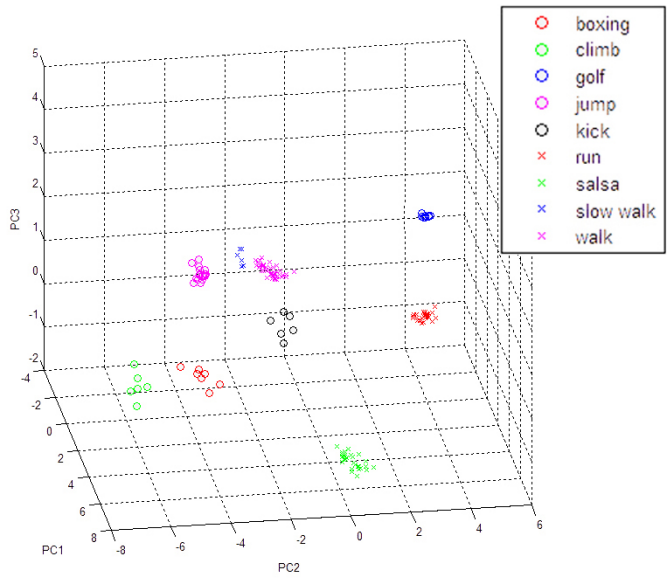


Fig. 3: Visualization of all the example motions in 3D space. The dimensionality of the calculated feature vectors is reduced to three using the PCA method.

TABLE I: Confusion matrix of the clustering results using the proposed method where the weighting value of each DOF was defined manually.

	M1	M2	M3	M4	M5	M6	M7	M8	M9
M1	7	0	0	0	0	0	0	0	0
M2	0	7	0	0	0	0	0	0	0
M3	0	0	11	0	0	0	0	0	0
M4	0	0	0	17	0	0	0	0	0
M5	0	0	0	0	6	0	0	0	0
M6	0	0	0	0	0	28	0	0	0
M7	0	0	0	0	0	0	30	0	0
M8	0	0	0	0	0	0	0	6	1
M9	0	0	0	0	0	0	0	0	46

M1: punching M2: climbing M3: golf M4: jumping M5: kicking
M6: running M7: salsa M8: slow walking M9: walking

IV. Experimental Results

We conducted an experiment to evaluate the accuracy of our proposed clustering method. We used 159 motions from the CMU motion database in this experiment, which were categorized into nine actions. Table I shows the confusion matrix for the clustering results when the w_d value of each DOF was defined manually. The columns show the predicted labels, whereas the rows contain the actual labels. In this experiment, we reduced the dimensionality of the feature vectors to three using PCA, as described earlier. The result is a near diagonal matrix, which is required for a good classifier. One of the slow walking motions was clustered in the walking motion group.

We conducted a similar experiment with the feature vectors generated using the weighting values calculated with Equation 3. The accuracy rate and confusion matrix were similar to the results produced when the weighting values were defined manually. This confirmed that the algorithm used to define the weighting values automatically for the proposed feature vector delivered acceptable results, and that the proposed method could be applied to other motion-capture datasets without defining the weighting values manually.

We conducted an experiment to compare the accuracy of the clustering results using the dimension-reduced feature vectors and the feature vectors with all dimensions. The accuracy of clustering process was 99.37% when the dimensionality of the feature vectors was not reduced, which was similar to the results obtained with the dimension-reduced feature vectors. This confirmed that PCA can be used as a feature vector dimensionality reduction technique. This technique can reduce the computational time required for the clustering process and it allows us to visualize feature vectors in simple 2D or 3D spaces.

Our proposed method required no supervision. In general, previous methods have used pose distance as the feature for motion clustering and they require supervision. We conducted an experiment to compare the accuracy of our proposed method with supervised methods. We used the same test dataset as that used in the previous experiments. We found that pose distance methods could deliver 100% correctness. The accuracy was slightly lower with our method compared with pose distance methods, but it was comparable (99.38 %). However, users need to provide query motions to produce 100% correctness with pose distance methods, which may be too difficult for them. The computational complexity of pose distance methods is also quite high compared with our proposed method. The main computation conducted by pose distance methods is DTW calculation. Thus, the computational complexity of existing methods is $O(D \times N_f^2)$, where D is the number of DOFs and N_f is the number of frames in the motion data. In our method, we represented the motion data as a feature vector with the size D. Each element of the feature vector was calculated based on the kinetic energy of each DOF in every frame. Therefore, the computational complexity for creating each feature vector is $O(D \times N_f)$. In the PCA, the main computation is the calculation of the covariance matrix, which has a computational complexity of $O(D^2)$. Any vector-distance calculation can be used, such as the Euclidean distance, and this step is not time-consuming. Therefore, the overall computational complexity of our method is $O(D^2+(D\times N_f))$. The number of frames in the motion data is usually significantly higher than the number of DOFs, so the computational complexity of existing methods is $O(N_f^2)$, whereas that of our method is $O(N_f)$.

We compared our method with the unsupervised approach reported in [10]. This method confused slow walking, walking, and running motions, whereas our proposed method did not experience these problems.

The experimental results confirmed that our method satisfies all of the three requirements. This was mainly because our method was unsupervised. The computational complexity of our proposed method was quite low compared with previous supervised methods, but it can provide a comparable performance in term of the accuracy of motion clustering.

V. Conclusion

In this paper, we introduced a kinetic energy-based feature for unsupervised motion clustering. This feature was based on an assumption that each motion has a unique pattern in

terms of the amount of movement in each joint. The amount of movement can be determined from the kinetic energy, so we calculated the velocity of each DOF in the motion data. Each element of the proposed feature vector is the average kinetic energy of each joint. The proposed method is simple to use because our proposed feature vector can be used to cluster all of the motion sequences without supervision. The performance of our proposed method was comparable to supervised methods in term of its clustering accuracy, while the computational complexity of the proposed method was lower. In future work, we would like to include information about changes in the kinetic energy to increase the reliability of our proposed method.

ACKNOWLEDGMENT

This research was supported by Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, and also was supported by AUN/SEED-NET.

REFERENCES

[1] CMU-Graphics-Lab, "Cmu graphics lab motion capture database," http://mocap.cs.cmu.edu/, 2012.

[2] L. Kovar, M. Gleicher, and F. Pighin, "Motion graphs," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 473–482, Jul. 2002.

[3] L. Kovar and M. Gleicher, "Automated extraction and parameterization of motions in large data sets," in *ACM SIGGRAPH 2004 Papers*, ser. SIGGRAPH '04, 2004, pp. 559–568.

[4] A. Witkin and Z. Popovic, "Motion warping," in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, ser. SIGGRAPH '95, 1995, pp. 105–108.

[5] I. Jolliffe, *Principal Component Analysis*. Springer Verlag, 1986.

[6] G. McLachlan and D. Peel, *Finite Mixture Models*, ser. Wiley series in probability and statistics: Applied probability and statistics. John Wiley & Sons, 2004. [Online]. Available: http://books.google.co.jp/books?id=7M5vK8OpXZ4C

[7] E. Keogh, T. Palpanas, V. B. Zordan, D. Gunopulos, and M. Cardle, "Indexing large human-motion databases," in *Proceedings of the Thirtieth international conference on Very large data bases - Volume 30*, ser. VLDB '04, 2004, pp. 780–791.

[8] R. Hamid, A. Johnson, S. Batta, A. Bobick, C. Isbell, and G. Coleman, "Detection and explanation of anomalous activities: Representing activities as bags of event n-grams," in *Proceedings of the CVPR '05*, 2005, pp. 1031–1038.

[9] J. Lee, J. Chai, P. S. A. Reitsma, J. K. Hodgins, and N. S. Pollard, "Interactive control of avatars animated with human motion data," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 491–500, Jul. 2002.

[10] S. Wu, Z. Wang, and S. Xia, "Indexing and retrieval of human motion data by a hierarchical tree," in *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '09, 2009, pp. 207–214.

A Parametric Motion Concatenation Method Using Cubic Bézier Interpolation

Suthasinee Nopparit*, Natapon Pantuwong†
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
Email: *suthasinee.nop@gmail.com, †natapon@it.kmitl.ac.th

Masanori Sugimoto
Department of Computer Science
Hokkaido University
Hokkaido, Japan
Email: sugi@ist.hokudai.ac.jp

***Abstract*—This paper presents a novel motion concatenation method for parametric motion synthesis techniques. First, motion groups are created based on the actions in each motion. We then extract all of the parameters that control the synthesized motions. To connect the motion groups, we propose a motion concatenation algorithm based on cubic Bézier interpolation that can be used to connect any pair of motions. All of the poses are pre-calculated before interpolation, so that the concatenated motions can be synthesized rapidly during the concatenation phase. Although there is no intersection region between the parameter spaces, the proposed method guarantees that transitions between motions can be generated for any consecutive motions, which is a problem found in existing methods.**

I. Introduction

Motion concatenation is performed by connecting pairs of motion sequences by determining the transition point based on pose similarity. Motion graphs are a commonly used method for motion concatenation. Motion graph-based methods can produce an infinite number of motions by determining all of the transition points in the graph. These transition points are based on pose similarity. If the poses are too different in the motions being concatenated, it is necessary to trace the graph using other motions. The main limitation of this method is that it does not allow control over the resulting motions. Thus, the characteristics of each motion segment are similar to the example motions in a database. This limitation was addressed by the parametric motion synthesis (PMS) technique [1]. Instead of determining the pose similarity, this method calculates the transitions between motions by blending the parameters of each motion pair. Each motion can be modified by altering the parameter values. For example, we can generate a new kicking motion by specifying the impact position of the foot. However, this technique also has limitations. One of the most important limitations is that it cannot generate a transition for any pair of motions if there is no intersection region between their parameter spaces.

We propose a novel motion concatenation method based on the PMS technique [1]. Figure 1 shows an overview of the proposed method, which creates a transition for any pair of motion groups. The motion groups shown in this figure have similar actions. Thus, we can apply the PMS algorithm to each motion group to generate the desired motion. After each motion has been synthesized, the resulting motion can be generated by concatenating all of the synthesized motions sequentially. To overcome the problems of existing methods,

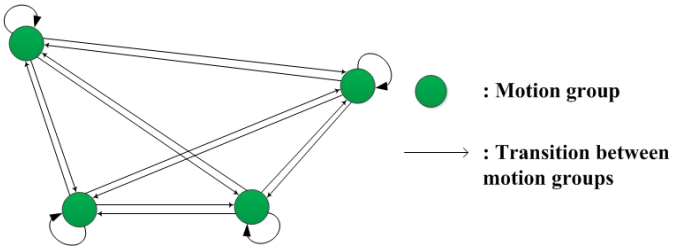


Fig. 1: The proposed method synthesizes motions for each required motion group and concatenates them to generate the resulting motion. A transition can be created for any pair of motions.

our method connects two motions directly based on their pose similarity, but without determining the parameter spaces of the two motions. The resulting motion can then be controlled by users, as in existing methods based on PMS techniques, but all of the synthesized motions can be connected, even though there is no intersection region in their parameter spaces. Our proposed technique uses cubic Bézier interpolation to concatenate each pair of motions. Cubic Bézier interpolation is used instead of the usual linear interpolation, which ensures a smooth transition between each pair of motions. The transition point of the first motion is the frame next to the end of the motion sequence, while the entry point of the second motion is the frame next to the first frame. We decided to create transitions using this strategy to ensure that the synthesized motions satisfied the user requirements. The motion graph technique can generate good quality transitions between two motions because it uses similar poses in the calculation, but it has a limitation in terms of the processing time, as mentioned previously. Cubic Bézier interpolation allows us to use all similar poses to create a transition in a similar way to the motion graph technique, except graph searching is not required because all poses are used in a single calculation. We precalculate all of the poses required by cubic Bézier interpolation, which ensures that the resulting motion can be generated rapidly.

The contributions of this paper are listed below.

1) The motion concatenation algorithm gives users control over the resulting motion sequence and it can guarantee that transitions can be generated between

motions.

2) The proposed method creates transitions between motions using cubic Bézier interpolation. This interpolation technique uses all of the similar poses in various motions to generate transitions in a similar way to the motion graph technique, except it uses all of the poses in a single calculation. Therefore, transitions can be created rapidly.

This paper is organized as follows. Section 2 discusses the details and limitations of related methods. Section 3 provides an overview of the proposed method. Section 4 explains the details of the proposed motion concatenation algorithm. The experimental results are presented and discussed in Section 5. Finally, our conclusions are given at the end of this paper.

II. Related Work

Motion concatenation is used for the automatic generation of longer motion sequences by connecting individual motion clips. In this section, we discuss two main approaches to motion concatenation, i.e., motion graph and parametric motion transitions.

A. Motion graph

MoveTree [2] was the first method to be introduced by the video games industry. This method uses a graph-like structure to describe the transition points between motion clips. MoveTree is a time-consuming method because the transition points between motions have to be specified manually.

To overcome the problems of MoveTree, many researchers have tried to use graph-like structures to represent the transition points in MoveTree with automatic algorithms. These algorithms [3], [4] organize motion clips into graphs for efficient concatenation. However, more recent work has been based on the "motion graph" method [5].

The main problem of these methods is that they usually require a graph-searching method. The most similar poses may still be too different when connecting two motion sequences, so we have to connect them with other poses in other motions. Motion graph is highly complex so a very long processing period is required to determine a suitable path in the graph. Another problem of motion graph is that users cannot control the resulting motion. Thus, each concatenated motion segment must be similar to one of those stored in a database. This means that users cannot determine the resulting motion with high precision using the motion graph method, such as a punching motion with a specific fist position. This method lacks the level of control provided by a continuous motion.

B. Parametric motion transitions

To overcome the limitations of motion graph-based methods, several researchers have tried to develop algorithms that concatenate several motion sequences, which can be controlled using a parameter set. Motion blending or the PMS technique may be combined with typical motion graph techniques to generate transitions between synthesized motions.

Methods that create transitions between parametric motions (synthesized motions) consider parameter values when creating

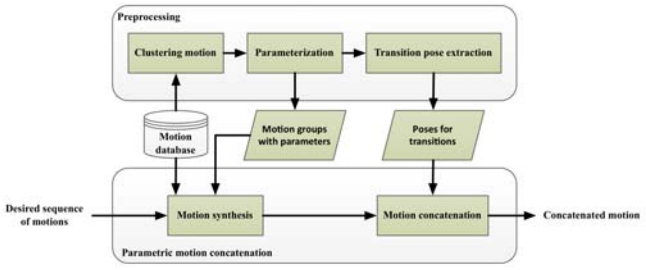


Fig. 2: Overview of the proposed parametric motion concatenation method.

the transitions between motion sequences. Based on the graph structure presented in [6], the fat graph method [7] groups edges together that represent similar motion segments and it represents each group using parameters. Fat graphs can be used to generate continuous streams of parametric motions, where the parametric transitions are executed in the same manner as motion blending in [8]. A major drawback of the fat graph method is that it constrains all motions to a discrete set of common poses. Thus, a character must return to a common transition pose before changing to the second motion at the transition point between two motions. This produces artifacts in the resulting motions.

The method known as parametric motion graph [1] was proposed to eliminate the problem of the fat graph method. First, this method builds parametric motions using the automated methods presented in [9], before organizing these motions into a parametric motion graph. The nodes of a parametric motion graph correspond to parametric motions while the edges represent the transitions between them. To concatenate two parametric motions, a parametric edge is generated by sampling the source and target motion spaces, before computing the region (sample subset) of the target motion that can be transitioned from each source sample. Runtime parametric transitions are executed using the k nearest neighbors method to find a region in the target motion space to which a transition can be made and the correct target time. Any graph-searching algorithm can be used to generate the resulting motion from the parametric motion graph. Parametric motion graph facilitates control over the resulting motion, but it has a few additional drawbacks. The most important drawback is that an edge can be generated between two parametric motions only if every motion in the source motion space can transition to a subspace in the target motion space.

III. Overview

The proposed method can be divided into two main steps, i.e., preprocessing and parametric motion concatenation. Figure 2 provides an overview of the proposed method.

During the preprocessing step, we cluster all of the example motions into motion groups. Next, we parameterize each motion group to identify all of the parameters in each motion group and we store all of this information for later use in the concatenation step. We also precompute all of the poses that are needed to generate transitions between any pairs of motion

groups (transition poses). This information is also used for motion concatenation.

To concatenate motions, users need to provide the desired sequence of motions. Each motion is synthesized from each motion group based on the parameter values specified by users. The concatenated motion can be generated with the proposed algorithm based on cubic Bézier interpolation using all of the precalculated transition poses

IV. The Proposed Method

A. Preprocessing

This section explains the preprocessing stage, which precalculates all of the information needed to concatenate the motion sequences. This stage has three steps: motion clustering, parameterization, and transition pose extraction.

First, all of the example motions are clustered into motion groups. Each motion group contains all motions with a similar semantic action description. In our implementation, we assumed that all of the example motions were created for a single action so motion segmentation was not necessary. Any motion clustering algorithm can be used to generate a set of motion groups, such as [9].

We can extract all possible parameter values from each motion group after the motion groups have been created, such as the location of a fist during a punching motion or the position of feet on the ground during a walking motion. These parameters can be used to control the precision of motions, which allows them to react to user requests. We perform parameterization using the method presented by [9]. This method requires all motions in each motion group to be registered with the reference motions in each group in time and space. In our implementation, the reference motions were selected manually.

The final preprocessing step is the extraction of all the poses for transitions. Cubic Bézier interpolation is used to connect two motions directly, regardless of the intersection region between the two parameter spaces. This interpolation method allows us to use a range of similar poses during transition generation, which is similar to the motion graph technique. However, our proposed method has less computational complexity because it uses all of the similar poses in a single calculation without any requirement to search for a path in a graph structure, which is a requirement of the motion graph technique. Thus, we can connect any pair of motions rapidly using this interpolation technique. Four poses are required for cubic Bézier interpolation: two poses (start–end poses) are the endpoints of the interpolation, while the others (control poses) are used to control the interpolation results. Extracting the poses required for interpolation during concatenation runtime is not a good option in interactive applications. Therefore, we extract all of the poses in advance that will be required for transitions. This information is used to concatenate motions during concatenation runtime.

We synthesize one motion for each motion group. The parameter values used to generate each motion are the averages of the sampled points in the parameter spaces. We do not modify the parameters that control the temporal information associated with the synthesized motions, so they are generated with a similar number of frames to the reference motion in each motion group.

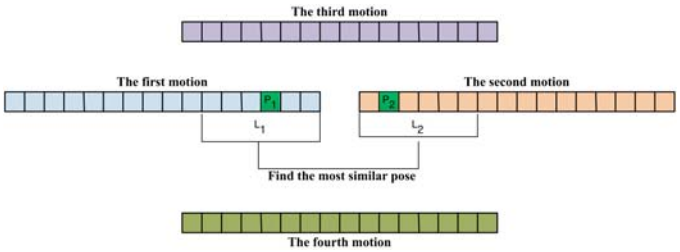


Fig. 3: Illustration of the method used to extract the start–end poses required for interpolation.

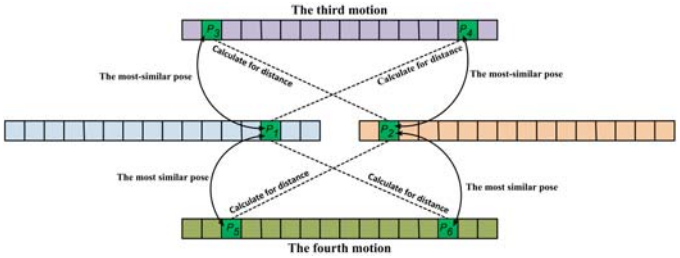


Fig. 4: Illustration of the method to extract control poses for interpolation.

All of the transition poses are extracted by determining the similarity between poses. Figures 3 and 4 show the method used to extract the start–end poses and the control poses, respectively. In this example, we concatenate the first motion with the second motion. The first, second, third, and fourth motions are synthesized motions in each motion group. The start–end poses used for interpolation are the poses near the end of the first motion and the pose near the first frame of the second motion. Figure 3 illustrates this process. We extract the most similar poses in these two motions (P_1 and P_2 in Fig 3). They are extracted by determining a distance matrix with $L_1 \times L_2$ dimensions, where L_1 and L_2 are the search window sizes of the first and the second motions, respectively. The two poses with the highest pose similarity are extracted. The pose similarity is calculated using the following equation:

$$d_{i,j} = \sum_{k=1}^{m} \left\| q_{j,k}^{-1} q_{i,k} \right\|^2, \tag{1}$$

where $q_{i,k}$ is the quaternion that expresses the orientation of joint k with respect to its parent in frame i. The value of $\|q_{j,k}^{-1} q_{i,k}\|$ expresses the difference between the orientation $q_{i,k}$ and the orientation $q_{j,k}$.

We require two other poses to control cubic Bézier interpolation. These poses are derived from the pose similarity between each extracted start–end pose and each frame in every motion, as shown in Fig 4. In this figure, for example, we calculate the distance between P_1 and all frames in the third and the fourth motions to extract the most similar poses from each motion. All of the poses extracted in this step (P_3, P_4, P_5, and P_6 in Fig 4) are used as control poses during interpolation.

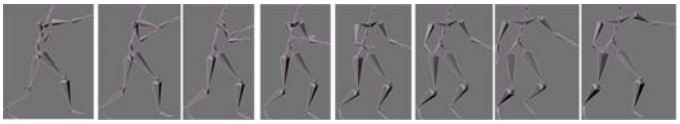
All of the poses extracted for transitions are calculated using synthesized motions with a fixed number of frames (similar to the reference motions in each motion group). However, users can alter the parameters during concatenation, which controls the temporal information associated with the synthesized motion. If this is the case, the frame numbers of the extracted transition poses need to be changed according to the parameter values specified by users.

B. Parametric motion concatenation

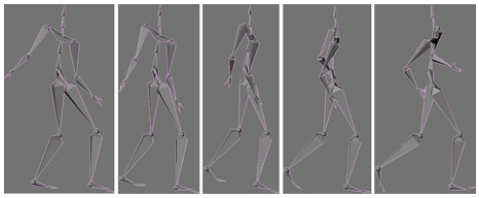
To generate the concatenated motion, users have to provide the necessary sequence of motions and all of the requisite parameter values to synthesize the motions in each motion group. After all of the requisite motions have been synthesized, they can be concatenated to generate a single resulting motion. However, each motion may be generated in a different position in the horizontal plane, i.e., the root positions of each synthesized motion may be different in the X–Z plane. Before performing concatenation, therefore, the two motions that need to be concatenated must be aligned in space. Next, we can generate the transition between two motions using cubic Bézier interpolation. Let us assume that we need to generate the transition between P_1 and P_2 shown in Fig 3. We perform linear interpolation of the root position between the start–end poses. Cubic Bézier interpolation is used to synthesize the rotation angles of the other joints. The start–end poses that concatenate these two motions are calculated during the preprocessing phase, so we can use them directly as the endpoints of interpolation. We need to select two poses from a set of control poses. Let us assume that the set of control poses is $c = \{c_1, c_2, c_3, ..\}$. The set of candidate control poses that can be used for interpolation is c_c ($c_c \subseteq c$). Each of the members of c_c, c_i is a candidate pose such that $d_{P_1,c_i} + d_{c_i,P_2} \leq d_{P_1,P_2}$. If c_c is an $\emptyset$ or there is only one member in c_c, we recalculate c_c where each element of the set c_i is a candidate pose such that $d_{P_1,c_i} \leq d_{P_1,P_2}$ or $d_{P_2,c_i} \leq d_{P_1,P_2}$. After c_c has been constructed, we can synthesize the transition using the following conditions.

- If there are two members in c_c, we can perform cubic Bézier interpolation using both poses. The pose that is closest to the start pose (based on the pose distance) is used to control the interpolation that moves from the start pose, while the other is used to control the interpolation that moves to the end pose.
- Otherwise, we divide all members of c_c into two groups based on their pose distance where the first group is closest to the start pose and the other is closest to the end pose. We average the joint angles (except the root) of all poses in each group. The average poses are then used for interpolation, as in the previous case.

In the current implementation, the transition length is set to $L_1 + L_2$. However, this length can be calculated automatically and adaptively using the algorithm presented in [10]. This calculation is performed iteratively for each consecutive motion according to the order specified by users and the resulting motion is obtained.



(a) Transition from punching to kicking



(b) Transition from kicking to running

Fig. 5: Examples of transitions between motions generated using the proposed method.

V. Experimental Results

We used 159 motions from the CMU motion database [11] in this experiment. Figure 5 shows two examples of transition poses generated using the proposed system. The results show clearly that our proposed method produced seamless transitions, which are required for the motion concatenation process. Figure 5(a) shows that a transition could be synthesized from a punching to kicking motion, although the position of the right arm during the kicking motion was not located in the impact position parameter space of the punching motion. This result confirmed that the proposed method can overcome the problems of a previous method [1], i.e., transitions can be created for any consecutive motions without considering their parameter spaces. Transitions could not be created using this previous method because there was no intersection region between the parameter spaces of two motions. Previous methods were incapable of creating transitions for some pairs of motions, so graph searching was required to find a path connecting the two motions. The computational complexity was increased because of the graph-searching process. The proposed method does not require a graph-searching method because it can create transitions for any pair of motions.

The proposed method uses cubic Bézier interpolation during the transition generation process and it can create good quality transitions because it employs various similar poses in the calculation. Linear interpolation can be used to generate the transition between two motions directly (without using several similar poses from other motions), but it cannot guarantee a seamless transition. If linear interpolation is used and the start–end poses are very different, the resulting concatenated motion may contain artifacts. We conducted a pilot study to compare the smoothness of the transitions generated with cubic Bézier interpolation and linear interpolation. We used linear interpolation to interpolate between the start–end poses and we asked 20 participants to observe the synthesized concatenated motion and provide feedback on its smoothness. The smoothness scale ranged from one (poor) to five (excellent). The average smoothness score with our cubic Bézier interpolation-based method was 4.1, whereas the average score with linear interpolation was 2.3 ($p < 0.001$ according to Wilcoxon's signed rank test).

VI. CONCLUSION

This paper presents the motion concatenation method based on cubic Bézier interpolation. First, all motions are grouped into motion groups with similar actions. Next, we extract the motion parameters and synthesize new motions for each motion group. We also calculate the transition poses required for cubic Bézier interpolation. The two motions are then concatenated. This strategy can generate concatenated parametric motions, although there is no intersection region between the parameter spaces of the two motions. Therefore, the proposed method allows users to control the resulting motion and it guarantees that transitions can be created between motions. In future work, we would like to improve the resulting motion by adaptively calculating the number of frames for each transition.

REFERENCES

[1] R. Heck and M. Gleicher, "Parametric motion graphs," in *Proceedings of the 2007 symposium on Interactive 3D graphics and games*, 2007, pp. 129–136.

[2] B. J. MIZUGUCHI, M. and T. CALVERT, "Data driven motion transitions for interactive games," in *Eurographics 2001 Short Presentations*, 2001.

[3] O. Arikan and D. A. Forsyth, "Interactive motion generation from examples," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 483–490, Jul. 2002.

[4] J. Lee, J. Chai, P. S. A. Reitsma, J. K. Hodgins, and N. S. Pollard, "Interactive control of avatars animated with human motion data," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 491–500, Jul. 2002.

[5] L. Kovar, M. Gleicher, and F. Pighin, "Motion graphs," in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, 2002, pp. 473–482.

[6] M. Gleicher, H. J. Shin, L. Kovar, and A. Jepsen, "Snap-together motion: assembling run-time animations," in *Proceedings of the 2003 symposium on Interactive 3D graphics*, 2003, pp. 181–188.

[7] H. J. Shin and H. S. Oh, "Fat graphs: constructing an interactive character with continuous controls," in *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 2006, pp. 291–298.

[8] L. Kovar and M. Gleicher, "Flexible automatic motion blending with registration curves," in *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 2003, pp. 214–224.

[9] ——, "Automated extraction and parameterization of motions in large data sets," in *ACM SIGGRAPH 2004 Papers*, 2004, pp. 559–568.

[10] J. Wang and B. Bodenheimer, "Synthesis and evaluation of linear motion transitions," *ACM Trans. Graph.*, vol. 27, no. 1, pp. 1:1–1:15, Mar. 2008.

[11] CMU-Graphics-Lab, "Cmu graphics lab motion capture database," http://mocap.cs.cmu.edu/, 2012.

Analysis of Factors Influencing the Mobile Technology Acceptance for Library Information Services: Conceptual Model

Singha Chaveesuk
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang,
Bangkok, Thailand
Singha@it.kmitl.ac.th

Sununthar Vongjaturapat
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang,
Bangkok, Thailand
Sununthar@kmitl.ac.th

Nopporn Chotikakamthorn
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
Nopporn@it.kmitl.ac.th

Abstract--Increasing powerful mobile technology, for example tablet devices, empower learners to seek information not only at home but also in mobile learning scenarios, virtually anywhere and anytime. However, the variety of library material types such as e-books, e-journals, images, audio and movies, require different levels of human- computer interaction. Moreover, the different library material types may be subject to different usability constraints. Thus, it is important for the academic library to look at how student use their mobile technology for library information services. This paper aims to propose a theoretical model for academic institutes and telecommunication service providers by addressing the following question: What are the key determinant factors for the mobile technology acceptance in using library information services? How the patron' s perspective and technical perspective are integrated in the process of mobile technology acceptance for library information services? We extend the extending Unified Theory of Acceptance and Use of Technology model (UTAUT2) to the context of library information services by examining the moderating role of library material type. Adapting the UTAUT2 model requires re-specification to suit a specific information system as well as including integration variables namely technology characteristics, task characteristics and attitude. The contribution of the model is a design guideline for the mobile technology function that is consistent with the library information services. Especially, for more better service management and delivery.

Keywords-information technology acceptance; mobile technology; Extending Unified theory of acceptance and use of technology medel (UTAUT2); task technology fit model; library service; information service

I. INTRDUCTION

It is expected by Gartner that smartphone and tablet usage will increase to 90% by 2015 [1]. Currently, Smartphone and tablet have the same capabilities [2], and have become an emerging phenomenon for personal voice, data, modern operating systems, internet access, and productivity-enhancing application. Consumers can use mobile technology for varied purposes. For instance, in the case of the Mobile Data Service: MDS, consumers can access MDS in various contexts-whenever they want and wherever they are-for different purposed, such as communication with their friends, retrieving online information, and conducting online transactions [3]. This means that the classifications of mobile technology have been based on their main objective, defining features, main capabilities, primary input mechanism or price [4]. For example, Mohageg and Bergman [5] listed the three main objective domains of the user: entertainment, information access and communication and assistant devices. They differ in user interaction and user expectation. For instant, in the entertainment domain the users experience is characterized by long periods of interaction, tasks are less structured and user attention and concentration may vary. Considering information access and communication devices, interaction periods are shorter, tasks are structured and users are more goal-oriented and focused. Assistant devices have a more specific user population and task set.

However, this is different from the library context where a patron's main objective of using IT is as an assistant device and information access. They interact with IT for long periods of interaction for gaining utilitarian benefits, efficiency, effectiveness and self-improvement [6]. Their tasks are structured and usually involved in the process of evaluating and perceiving value of the utility or the relevance of information [7], [8]. In addition to the acceptance of IT for library information services, it is more important to accept the content and services provided by a digital library system. While, library material type is the object that appears on the user's screen, there are many possibilities represented in e-books, e-journals (novel, journal and magazine), images, audio and movie forms [9]. Beside there is evidence that the type of technology or its unique characteristic could potentially make a difference [10] [11]. Likewise, the type of library material will may play a differential role, i.e., in some cases, they will be particularly pronounced, while in other cases, they may be weak or even non significant. Related to this, they require

different levels of human- computer interactions [12], and may be subject to different usability constraints [13], and lead to a reduction of a patron's perception of effectiveness, ease of use and facilitation among patrons where they perceived that the system is not useful and accomplish their task for library information services [14]. As a result, the acceptances of mobile technology for library information services is not widely and frequently used [2] Thus, with regard to the original relationship in UTAUT2, we also retain the moderation effect of demographic variables, such as age, gender and experience as control variables.

Such changeability depending on library material type will require expanding the theoretical domain related to UTAUT2 in the context of library services and an increase in the contextualization of UTAUT2. This is a key next step in leveraging UTAUT2 to understand a patron's use of mobile technology for library information services. Thus, compared to performance expectancy and effort expectancy and other factors that can influence user acceptance of information technology (IT), as described in UATUT2 [15], it is likely to play a critical role in mobile technology adoption for library information services. Therefore, to move from a stationary based system to a mobile technology based system, we need a greater effort to fully understand the nature of the patron's task and the needs for mobile technology support as well as to link the patron's perspective, mobile technology characteristics and the patron's task characteristics. This research aims to investigate and to represent what influences the factors on the mobile technology acceptance for library information services.

II. LITERATURE REVIEW

A. *Mobile technology in library context*

Mobile technologies include laptops, netbooks, e-readers, tablets, mobile phones, smartphone MP3/ MP4 players and internet capable handheld devices. In addition, wireless networks, such as 3G, are providing the network infrastructure for users, allowing tablets and smartphones to be widely used. Tablets are also available for browsing online, downloading, streaming and lending digital content from libraries. This means that searching a database, downloading articles, and seeking catalogs can be conducted via tablets. In addition, the benefits of mobile technology can support users to access documents, including e-books, audio-visual objects and websites. There are several libraries that have launched pilot programs to show that tablets provide convenience in accessing the library's resources, and tablets are new innovations that attract students. Specifically, the design of tablets with lighter weights, slimmer shapes and multi-touch screen interfaces can eliminate the need for peripheral devices [2]. This means that patrons will likely use mobile technology because of the ease of surfing the internet, listening to music, reading documents, and so on.

B. *UTAUT2*

In 2003, Venkatesh et al. compared eight outstanding models for IT acceptance research and formulated a unified model called the UTAUT in which explained about 70% of the variance in intention to use a technology among employees in a workplace [16]. The UTAUT was developed in the organizational use setting which was an important difference from the consumer uses setting [16]. In 2012, theUTAUT2 produced three constructs into UTAUT: hedonic motivation, price value, and habit. The UTAUT2 has become the latest and most powerful theory with 73 percent of the variance in intention (adjusted R2), to predict and explain information systems usage intention among consumer. This theoretical model comprises seven main determinants of intention and use: (1) performance expectancy (PE), (2) effort expectancy (EE), (3) social influence (SI), (4) facilitating conditions (FC), (5) hedonic motivation (HM), (6) price value (PV) and (7) habit (HA) as well as three moderating variables: (1) gender, (2) age and (3) experience [15]. While the Venketesh, UTAUT2 model helps in understanding and predicting the drivers of acceptance of new information technologies by consumers, the Goodhue [17], task-technology fit model helps explaining how technology leads to the impact of user involvement on performance when technology provides features and support that fit the requirement of a task. However, the shortcoming of the UTAUT2 model was that is did not clearly identify task characteristics and technology characteristics that were explicitly contained in the task-technology fit (TTF) model. The TTF model is considered inadequate in explaining the individual differences –namely, age, gender, and experience, social influence, [18] and behavior intention. There are limitations in explaining the acceptance of mobile technology for library information services.

C. *Overview of library meterial type*

Library materials are digital data services that can be accessed through mobile technology. Library materials have many types that allow patrons to conduct accessing online information. As a result, library materials have the potential to be the greatest evolution in the history of books after the birth of printing [19]. Library material type and their use by patrons have some interesting characteristics.

First, because a digital library might include objects that manifest themselves in different forms, they have different methods for access and retrieval [9]. For example, compared to the Web site, the documents are treated as structured objects internally which enhances the prospects for providing comprehensive searching and browsing facilities [7]. This document type have many possibilities to be structured text documents, page images, or page images along with the accompanying text. The interface allows patrons to locate words and phrases in the_document collection. Searching is useful for metadata-such as finding words in title and author fields. There are many multimedia document type: audio recording and photographic images; video, which includes both images and audio components; and musical objects that can be presented in several different formats such as WAV, AIFF, AU, MPEC and MIDI for audio or AVI and ASF from Microsoft, FLV from Adobe and Real Media from Real Networks for video [20]. The interface allow patron to locate

words in the multimedia collection. Searching is useful for metadata such as for title, composer and lyric fields [9].

Second, regarding the objects themselves, they appear on the user's screen in differentials. For example with, sound, patrons can listen to a particular recording using a standard software audio-player that has regular functions (pause, fast forward, and so on) on the small control panel. Perhaps the digital library designer considers providing an icon that represents the sheet music and the player separately [21]. For video, it combines time-based information with a spatial image component. As with audio, a time-based document can be made more convenient by being able to segment them, and videos can be automatically converted into sequences of thumbnails that correspond to scene changes. The player can be seen on the small control panel [22]. Thus, the varieties of library material type provide different types of value to patrons and require different levels of human- computer interactions [12]. Moreover, different library material types may be subject to different usability constraints [13].

III. PROPOSED RESEARCH MODEL

With regard to the original relationship in UTAUT2, we believe that the direct effecting factors of PE, EE, SI, FC, HM, PV and HA will hold in the mobile technology acceptance of information services in a library context. We also retain the moderating effects of age, gender, and experience as controlled variables in the model. However, the voluntariness of moderating variables was not applicable in the model due to the fact that acceptance mobile technology in a library context is voluntary. The three primary attributes were defined as the independent variables and assumed that they affect to BI of the mobile technology acceptance for library information services through PE and EE as shown in Figure1. Based on the proposed research model sixteen hypotheses were developed to be next verified its value.

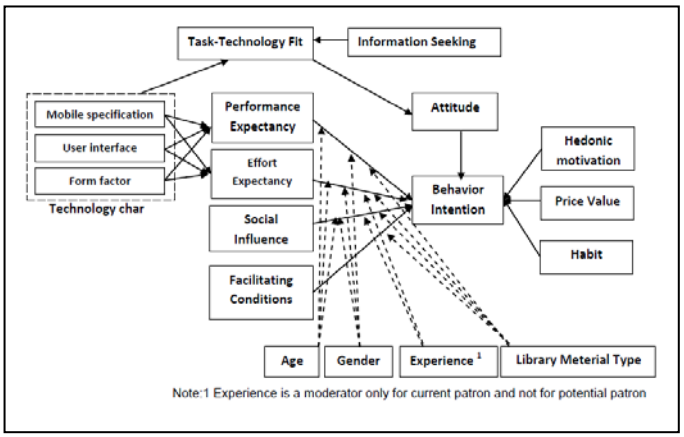


Fig1. Proposed research model

First, one further moderators was identified that may influence the patrons' use of mobile technology of information services in a library context. The library material type will be added to examine whether library material type will influence the use of mobile technology for library information services. They are e-book, e-journal (novel, journal and magazine), image, audio and movies. The variety of information services type provides different types of value to patrons and requires different levels of human- computer interactions [12]. Moreover, different information types may be subject to different usability constraints [13]. Therefore, this study establishes hypotheses for library information services types as follows:

A. *Moderation of PE-BI Relationship*

In their review, Venkatesh et al. [15] found that PE[a] has been consistently strong in explaining BI across many studies and technologies and the effect has been found to be stable over time, even as user experience with a technology increase. Regarding the UTAUT2, the user experience with the target technology is not theorized to be a moderator of the PE-BI relationship. Thus we expect the effect of PE on BI would be similar for both potential and current patrons. For mobile technology acceptance in a library context, where different types of library material may increase the barrier to e-reader adoption, it is hypothesized that PE would negatively influence BI. For example, document: Patrons used e-reader primarily for leisure reading and continue to rely on print for much of their reading. Moreover, patrons had little interest in borrowing e-reader compatible e-books from the library [23].This indicates that potential and current patrons do not believe that using the mobile technology for library material will help them to attain gains in assignments/tasks. Thus, we hypothesize as follows.

[Hypothesis 1A] The effect of PE on BI will be moderated by age, gender and library material type, such that the effect will be strongest for younger men and particularly for documents.

B. *Moderation of EE-BI Relationship*

We argue for differential moderating effects of the EE relationship with BI across the library material type among potential patrons (i.e. no experience with mobile technology acceptance for library information services) and current patron (i.e., low to high experience with mobile technology acceptance for library information services) Prior research has argued that gender × age × experience moderates the EE-BI relationship [15] [16]. Among patrons with experience, using a small virtual key to browse information or download multimedia document services via a small number of steps using a menu [24] will be less burdensome than the effort needed in terms of concentration to evaluate and perceive the value of document. Thus, the multimedia type would positively influence BI and we hypothesize as follows.

[a] Despite the different labeling of PE in different model (e.g., perceived usefulness: Technology Acceptance Model, extrinsic motivation: Motivational model, and relative advantage: Innovation diffusion Theory), the conceptual similarities among them have been acknowledged in may prior studies [15]

[Hypothesis 1B-1] For current patrons, the effect of EE on BI will be moderated by age, gender, experience and library material type, such that the effect will be strongest for older women in the early stages of experience and particularly for multimedia types.

When it comes to potential patrons, we expect the differences between the library material types to be the opposite of what was predicted for current consumers. We expect the interaction term to be more emphatic in the context of document type. We believe our prediction to be concerned about eye strain [23], i.e. patron interact with IT for long periods of interaction for gaining utilitarian benefits, efficiency, effectiveness and self-improvement [6]. Their tasks are usually involved in the process of evaluating and perceiving value of the utility or the relevance of information [7], [8]. Concentrations and focus will be a constant challenge in order to effectively use document type [2] [25] [26]. Older women groups, who are concerned about EE will be affected by these problems. As EE often operates as a hurdle, it can increase in importance when the use of a system is perceived to be more effortful [27]. Thus, we hypothesize as follows.

[Hypothesis 1B-2] For potential patrons, the effect of EE on BI will be moderated by age, gender, and library material type, such that the effect will be strongest for older women and particularly for document type.

C. Moderation of SI-BI Relationship

SI is the normative pressure based on social values the defined normal behavior for a group to which an individual belongs [15], [16]. According to prior research, SI has been an important determinant of individuals' BI to use ICT [28] and consumer acceptance for personal information and communication technology services [24]. In the library material service acceptance context, we believe that the influence of peers and lecturer would affect BI because of individual and group assignment provided by the lecturer [29], library document type is not the first target for the needs of a task. Moreover, a potential patron perceives that others believe they should not use the new facilities due to of the fact that the library document material is not useful and will not accomplish their task [14]. Intercommunication reflects that when studying something in depth, hard copy is better [30]. Thus, we hypothesize as follows.

[Hypothesis 1C-1] For potential patrons, the effect of SI on BI will be moderated by age, gender, and library information services type, such that the effect will be strongest for older women and particularly for document type.

Among those currently experiencing the service, we expect that the effect of SI would be more important for document rather than multimedia document type. The main disadvantage of the digital document type is the difficultly in reading from the screen and it not being recognized [19], [25], [32]. However, there is evidence that in the early stages of experience, patrons (student) carried out the tasks with enthusiasm and eagerness to learn using the new facilities of tablets as well as collaboration on problem solving such as they were also helping each other in finding the right tools on the interface [29]. Thus peers might influence their views if their experience was successful about using library material that was consistent with existing patron's devices. Thus we hypothesize as follows.

[Hypothesis 1C-2] For current patrons, the effect of SI on BI will be moderated by age, gender, experience and library information services type, such that the effect will be strongest for older women in early stages of experience and particularly for document type.

D. Moderation of FC-BI Relationship

According to theory, a consumer who has access to a favorable set of facilitating conditions is more likely to have a higher intention to use technology [15]. Whether or not the patron can access library material type seamlessly and in a timely manner can be viewed as an important environmental factor that may impede or facilitate the library material type. Such access issues will be relevant for both types of material. However, the relationship is expected to be stronger in the multimedia type rather than the document type. Because of portability, interoperability (i.e different type of mobile technology that support different type of multimedia) [21], [23], is related to the feasibility of downloading video over the Internet depends on technical factors, for instance the bandwidth of the connection as well as storage space will be required for a large collection [31]. Moreover, it is difficult for patrons to adapt to downloading video over the internet which depends on technical factors, such as unresolved technical issues, affecting smooth data transfer or mobility constraints [32]. We believe that the lack of stability in multimedia service, i.e., accessibility, responsiveness, and availability of multimedia documents have become an obstacle [33]. Overall, compared to library material types to which patrons have already adapted, multimedia is the material that is the dynamic media for receiving the value and the mobile context amplify the importance of FC to patron. Thus, we hypothesize as follows.

[Hypothesis 1D] The effect of FC on BI will be stronger for multimedia type and document type.

Second, a hypothesis regarding the *mobile specification* refers to the aspect of any component, specification, capability, service or function that offers users more flexibility and solves technical limitations [34] [35], such as the small screen size and keypad. There is evidence that students prefer doing assignments on mobile tablet technology due to the fact that allows users to directly and naturally write on the document with their fingers [29]. Moreover, the mobile specifications, such as reader modes, hand-writing, highlighting and adding annotations to the text [30], [36], may influence the behavior intention to use mobile technology in library information services. Therefore, this study establishes hypotheses for mobile specification as follows:

[Hypothesis2-1] Mobile specification influences on PE of the mobile technology acceptance for library information services are positive.

[Hypothesis2-2] Mobile specification influences on EE of the mobile technology acceptance for library information services are positive.

Third, a hypothesis for *user interface* was established. The user interface has two aspects: (1) the UI hardware interface includes the output devices such as displays and as well as the input devices such as keypads [37]. And (2) the UI software interface has different software levels, e.g. operating systems, UI software and application software. According to the research of Yung Fu Chang, Chen and Zhou [34], User interface is a very critical factors for the success of a particular brand of smartphone in the market place. Therefore, this study establishes hypotheses for user interface as follows:

[Hypothesis 3-1] Mobile user interface influences on PE of the mobile technology acceptance for library information services are positive.

[Hypothesis 3-2] Mobile user interface influences on EE of the mobile technology acceptance for library information services are positive.

Fourth, a hypothesis for *form factor* was established. Form factors of mobile technology come in two basic types: calm-shell and bar. (1) Calm-shell is a form with the ability to open up in the same way. (2) Bar is a form with no moving parts aside from the buttons [38]. Devices such as phones are designed to be used with one hand only, whereas other devices, like tablets require the user to hold the device in one hand while using the other hand to provide input through typing or with a stylus [39]. According to the research of Yung Fu Chang, Chen and Zhou [34], form factor is one of the critical factors for the success of a particular brand of smartphone in the market place. Therefore, if a patron perceives that the system is effective and easy to use, they will keep using IT. Therefore, we hypotheses for form factor as follows:

[Hypothesis 4-1] Form factor influences on PE of the mobile technology acceptance for library information services are positive.

[Hypothesis 4-2] Form factor influences on EE of the mobile technology acceptance for library information services are positive.

Fifth, technology characteristics are viewed as tools used by individuals in carrying out their task [24], and the independent variables, may influence the patrons' perception of the effort expectancy for accepting mobile technology in library information services. The technology characteristics were indicated by form factor, user interface, and weight, such as lighter weights, slimmer shapes, and multi-touch screen interfaces. Once these characteristics are compatible with library information services, it may lead to feelings of ease and therefore lead to increased use of the system [2]. Therefore, this study establishes a hypothesis for technology characteristics as follows:

[Hypothesis 5] technology characteristics influences on TTF of the mobile technology acceptance for library information services are positive.

Six, Information Seeking is defined as a task characteristic/assignment carried out by individuals in turning input to output [40]. This paper defines the information seeking task of a patron's queries in IR systems as two processes: (1) Patrons have to express what they want from the system (e.g. in the form of a query), (2) then patrons are usually involved in the process of evaluating and perceiving value of the utility or the relevance of information that the IR system retrieves [7], [8]. Patrons performed this task via user interfaces that facilitates a user's control and interaction with information technology [41]. These may influence the patron's perception of the degree of fit between information retrieval tasks and technology characteristics of mobile technology. So we assume that the Information seeking task is frequently dealt with by searching literature. Therefore, this study establishes hypothesis for Information Seeking as follows:

[Hypothesis 6] Information Seeking influences on TTF of the mobile technology acceptance for library information services are positive.

Seven, Task-Technology fit is the degree to which a technology assists an individual in performing their task. This model depicts the way in which TTF leads to attitude or belief toward expected consequences of utilization that patrons measure the different characteristics about the system. TTF is an important determinant of whether systems are believed to be useful, important, or give relative advantage [17], [40]. According to the definition of task-technology fit, if technology characteristics fit well to a patron's task, patrons should have a positive attitude toward technology which in turn influences behaviour intention. Therefore, this study establishes a hypothesis for TTF as follows:

[Hypothesis 7] TTF influences on Attitude of the mobile technology acceptance for library information services are positive.

Eight, Attitude refers to the degrees of a person's favourable or unfavourable evaluation or appraisal about performing target behaviour [42]. According to the Theory of Plan Behaviour (TPB) and TAM, the attitude impacts the user's behaviour intention, which in turn influences actual behaviour [43], [44]. When patrons form positive attitudes toward the ideal fit between task characteristics and technology characteristics, they may have behaviour intention to accept it. Therefore, this study establishes a hypothesis for TTF as follows:

[Hypothesis 8] Attitude influences on behaviour intention of the mobile technology acceptance for library information services are positive.

IV. CONCLUSION

This paper builds a user acceptance structural model of mobile technology acceptance for library information service based on UTAUT2 from the view of user cognition. In addition to the acceptance of mobile technology for library information services, it is more important to accept the content and services provided by digital library systems. So this paper builds influence for user acceptance model of mobile technology for library information services by extending the UTAUT2 to the context of library information services and examining the moderating role of library material types. This is only a conceptual model by applying two stages in the preliminary study: 1) Reviewing and investigating the potential factors from previous research.2) Using semi-structured interviews by applying check list methods to collect data. What we should do next is going to design a questionnaire and collect data with potentials and current patron. Final step will be evaluated its value to design guidelines of the mobile technology function that consist of the library context and enables users to use them efficiently.

REFERENCES

[1] Gartner, Top Predictions for IT Organizations 2012 and Beyond, WWW page http:// gartner.com/it/page.jsp?id=13281132010, Access 01/09/2012.

[2] K. Kosturki, and F. Skornia, "Handheld Libraries 101Using Mobile Technologies in the Academic Library," WWW page http:// infotoday.com/cilmag/jul11/Kosturski_Skornia.shtml2010, Access 01/06/2012.

[3] S. Lee, B. Shin, and H. G. Lee, "Uuderstanding post-adoption Usage of mobile Data Service: The Role of Supplier-side Variables," Journal of the Association for Information Systems, vol. 10, no.12, pp. 860-888, 2009.

[4] T. Young, Software Interface Design for Small Device: Softeare Practices Lab, University of British Columbia, 2003.

[5] M. Mohageg, E. Bergman, Paper presented at the CHI'00 extended abstracts on Human factors in computing system, The Haque, The Netherlands: 63-64, ACM Press, 2000.

[6] W. Hong, J.Y.L.Tong, W.M. Wong, and K.Y. Tam, "Determinants of user acceptance of digital libraries: An empircal examination of individual characteristics and system characteristics," J. Manage. Inf. Syst.,vol.18, pp. 97-124, 2002.

[7] A. Tombros, I Ruthven, and M. J Jose, "How Users Access Web Pages for Information Seeking." Journal of the American society for information science and technology, vol. 56, no.4, pp. 327-344, 2005.

[8] M. A., Gonvalve, et al. "What is a Good Digital Library?" – A Quality Model for Digital Libraries. Virginia Tech CS Department, Federal University of Minas Gerais, Horizonte MG Brazil, Preprint submited to Elsevier Science, pp. 1-34, 2006.

[9] I.H. Witten, D. Bainbridge and D.M. Nichols. How to build a digital library. Second Edition, Imprint; Morgan Kaufman. 2009, pp.1-629.

[10] J. Y. L. Thong, S. J. Hong, and K. Y. Tam, "The effects of post-adoption beliefs on the expectation-confirmation model for information technology continuance," Int. J. Hum.-Comput. Stud., vol. 64, pp. 799–810, 2006.

[11] H. Van Der Heijden, "User acceptance of hedonic information systems," MIS Q., vol. 28, pp. 695–704, 2004.

[12] W.J. Orlikowshi and C.S.Iacono. "Desperately seeding the 'IT' in IT research: A call to theorizing the IT artifact," Inf. Syst. Res., vol.12, pp.121-134, 2001.

[13] V. Venkatesh and V. Ramesh, "Web and wireless site usability: Understanding differences and modeling use," MIS Q., vol 30, pp.181-205, 2006.

[14] A.L. Rahman, A. Jamaludin, Z. Mahmud, "Intention to use digital library based on modified UTUTA Model: Perspectives of Malaysian postgraduate Students," World academy of science, Engineering and technology, vol.75, pp.116-122, 2011.

[15] V.Venketesh, J.Y.L. Thong. and X. Xu, "Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology," Forthcoming MIS Quarterly, vol. 36, no.1, pp.157-178, 2012.

[16] V. Venketesh, et al. "User acceptance of information technology: toward a unified view," MIS Quarterly, vol. 27, no. 3, pp. 245-478, 2003.

[17] D.L.Goodhue, R.L.Thompson "Task-technology fit and individual performance," MIS Quarterly, vol.19, no.2, pp.213-236, 1995.

[18] J. Zhang, J.Huang, and J. Chen, "Empirical research on use acceptance of mobile search," Tsinghua science and technology, vol.15, no.2, pp.235-245, 2010.

[19] Y. Fu, X. Zhang, X. Hu, "Realization of Electronic Paper-Based Mobile Digital Library," IEEE, pp. 6719-6722. 2011.

[20] N. Chapman and J.Chapman, Digital multimedia, 3rd Edition, Wilay, 2009.

[21] P.S. Lampropoulou, A. S. Lampropoulou, and G.A.Tsihrintzis, "Evaluation of Middleware System for Accessing Digital Music Libraries in Mobile Services," Fitth International Conference on Information Technology: New generations, IEEE, pp. 1206-1207, 2008.

[22] W. Hürst and K. Meier, "Interfaces for timeline-based mobile video browsing" Proc. ACM International Conference on Multimedia (MULTIMEDIA 08), ACM Press, 2008, pp. 469–478.

[23] N. M. Foasberg, "Adoption of E-Book Readers among College Students: A survey," Information technology and libraries, pp. 108-127, 2011.

[24] J.Y.L.Thong et al. "Consumer acceptance for personal information and communication technology services," IEEE transaction on engineering management, vol.58, no.4, pp.613-625, 2011.

[25] Zhao Liping, "Mobile Digital Library Driven by Electronic Paper,"Library and Information Service, pp.25-27, 2010.

[26] Jiang Xuane. "Electronic paper reader and the influence of digital library Information Research", pp.33-35, 2007.

[27] V. Venkatesh, "Creation of favorable user perceptions: Exploring the role of intrinsic motivation," MIS Q., vol. 23, pp. 239–260, 1999.

[28] V. Venkatesh and M. G. Morris, "Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior," MIS Q., vol. 24, pp. 115–139, 2000.a

[29] G.M.N, Isabwe "Invertigating the usability of iPad mobile tablet in formative assessment of a mathematics course," International Conference on Information Society (i-Society 2012), IEEE, 2010.

[30] W.A. Shelburne, "E-bok usage in an academic library: User attitudes and behaviors," Library Collection, Acquisitions,& Techninal Services, vol. 33, pp. 59-72, 2009.

[31] W. Hong, J. Y. L. Thong, W. M. Wong, and K. Y. Tam, "Determinants of user acceptance of digital libraries:An empirical examination of individual characteristics and system characteristics," J. Manage. Inf. Syst., vol. 18, pp. 97–124, 2002.

[32] I. Lee, J. Kim, and J. Kim, "Use contexts for the mobile internet: A longitudianl study monitoring actual use of mobile internet services," Int. J. Hum.-Computer. Interact., vol. 18, pp. 269-292, 2005.

[33] H. Jeong, "An investigation of user perceptions and behavioural intentions towards the e-library." Library Collection, Acquisitions, & Technical Services, vol. 35, pp. 45-60, 2011.

[34] Yung Fu Chang, C.S.Chen, Hao Zhou , "Smartphone for mobile commerce," Computer Standards & Interfaces, vol. 31, pp. 740-747, 2009 .

[35] S.H.,Han et al. "Identifying Mobile Phone Design Feature Cirtical to user Satisfaction." Human factors and Ergonomics in Manufacturing, vol. 14, no.1, pp.15-29, 2004.

[36] Samsung (2013) WWW page http://www.samsung.com/th/consumer/ mobile-devices/galaxy-note., accessed 18.02.2013.

[37] P. Ketola, Intergrating usability concurrent engineering in mobile phone development , Docteral Dessertation, University of Tampeere, Tampere, 2002.

[38] J.A.V. Biljon, A model for representing the motivational and cultural factors that influence mobile phone usage variety. Doctor of philosophy, Computer Science, University of South Africa, 2006.

[39] D.L. Goodhue, "Understanding user evaluations of information system." Management Science, vol 41, no2, pp. 1827-1844, 1995.

[40] V. Cho, T. C. E. Cheng, W.M. Lai, "The role of perceived user-interface design in continued usage intention of self-paced e-learning tools." Computers& Education, vol. 53, pp.216-227. 2009.

[41] M. Fishbein, I. Ajzen, Belief, Attitude, Intention and Behaviour: An Introduction to Theory and Research, Addison-Wesley, Reading, MA, 1975.

[42] I. Ajzen, "The Theory of Planned behaviour." Organizational behaviour and Human Decision Processes,vol. 50, no. 2, pp.179-211, 1991.

[43] F.D. Davis, "Perceived usefulness, perceived ease of use and user acceptance of information technology." MIS Quarterly, vol. 13, no 3, pp. 319-340, 1989.

ASSESSMENT OF SERVICE MATURITY OF ”KARTU JAKARTA SEHAT” APPLICATION SYSTEM

Muhammad Fathir Gumilang, Kho I Eng, Maulahikmah Galinium
Faculty of Engineering and Information Technology, Swiss German University
{muhammad.gumilang@student.,ie.kho@,maulahikmah.galinium@}sgu.ac.id

***Abstract*— Kartu Jakarta Sehat (KJS) or “Jakarta Health Card” is a health insurance program proposed by current local government of DKI Jakarta through Jamkesda Dinas Kesehatan Province DKI Jakarta to the community in the form of medical treatment. KJS uses EDC (Electronic Data Capture) to integrate with clinic and hospital in Jakarta. This card is used to save medical data of every patient who hold the card. The research is conducted to obtain information about the usage of KJS. The method to collect the information is by interviewing the respondents and spread out questionnaire to clinics between sub-district health center and urban health center in certain area of Central Jakarta. The purpose of this research is to measure maturity level of KJS application system using COBIT Framework. As a result, it is found that the system is good enough with maturity level as follow: result DS1 (3.25), DS2 (1.73), DS5 (3.33), DS7 (1.87).**

Kartu Jakarta Sehat, EDC (Electronic Data Capture), Maturity Level, COBIT

I. INTRODUCTION

DKI Jakarta is the capital of Indonesia which has 9.6 millions of population. Among all provinces and cities in Java, the populations are concentrated in DKI Jakarta. According to ratio of the size of Jakarta and its population, Jakarta can be referred as the most densely populated city in Indonesia [1].

The level of welfare of its inhabitants is variant which the roughly be is from high class to low class communities. It can be assumed that from middle-high class to low class communities have slightly difficulties to obtain health services from health services agent like hospitals and clinics. However, most of middle class, middle-low class and low class communities have significant obstacles to obtain those kinds of service. That particular condition affects the overall welfare of population in Jakarta. That is why program Kartu Jakarta Sehat (KJS) or “Jakarta Health Card” was established.

A similar KJS program had been established in the period of previous DKI Jakarta Governor. Even though it was distributed several years ago, that program was not properly managed. Afterwards, when the period of new governor, they proposed a plan for re-managing KJS program which has been distributed and published.

The purpose of this research is to measure service effectiveness and calculate maturity level of KJS system in order to achieve the level of IT management of its current (as-is) and the level of expected information technology management (to-be) using COBIT framework. Reference [2] used this framework to evaluate integrated toll collection system while Sultani [3] used COBIT framework to build audit application system in a hospital. COBIT can also be used to analyze gap in IT governance process for data management [4].

This research uses Domain Delivery and Service (DS1, DS2, DS5, DS7) framework COBIT in clinics between urban health center and sub-district health center Tanah Tinggi (Central of Jakarta). The result of this research and effectiveness evaluation’s of KJS System may contribute for improving health service in Indonesia.

II. TECHNOLOGY AND INFRASTRUCTURE

Technology requirements of KJS system for each clinic and hospital are EDC tools [5] to track the “id” patients who are examined in hospital and clinic. The process of KJS system is explained as follow:

1) PT. AdMedika has an application to operate the information from EDC card and connecting using internet to its data center from PT. Telkom.
2) For every clinics and hospital, PT. ASKES [6] will set the EDC machine without the server. The required aspect is connection to read the EDC card.
3) The duty of the clinics for patients who register without KJS is to input their data from their ID (*KTP*) and their family card (*Kartu Keluarga (KK)*) . The purpose of PT.ASKES for having the patients data is to create KJS based on data which clinics application has.
4) The EDC machine will send the patient information to PT. ASKES server by reading the barcode in the KJS card.
5) The patients will get the medical record number which has ID (*KTP*) number in KJS.
6) PT.ASKES and PT. Telkom have the patient’s ID with certain diagnosed disease obtained by using the running clinics application system.
7) PT. ASKES and PT. Telkom implement the clinics information system and the duty of PT. ASKES is to give limit to data patient by using DRG (Diagnosis relation group).

Fig. 1 illustrates a flowchart about clinic and hospital treatment process. It starts from people come to clinics untuk they finish to be examined by doctors either in Clinic or Hospital.

III. COBIT

COBIT Framework is used as one of the supporting tools. COBIT Framework is a measurement tool for identifying the maturity level of a system developed by ITGI [7]. The IT Governance Institute (ITGI) was established in 1998 to advance international thinking and standards for directing and controlling an enterprise information technology. ITGI develops COBIT Framework to ensure IT support business goals, business investment in IT and defining the guidelines to manages related IT risks. There are some parts of standard tools/framework as :

1) COBIT (Control Objectives for Information and Related Technology)
2) COSO (Committee of Sponsoring Organizations) Internal Control-Integrated Framework
3) ISO/IEC Code of Practice for Information Security Management
4) ITIL (Information Technology Infrastructure Library)

In this research, Domain Delivery and Support of COBIT's modules that are used are defined as follows

1) DS1. Define and Manage Service Levels
2) DS2. Manage Third-Party Services
3) DS5. Ensure System Security
4) DS7. Educate and Train User

Define and Manage Service Levels (DS1) covers Service Level Agreement (SLA) between Health Department of DKI Jakarta, PT. Askes Indonesia, PT. AdMedika and clinics or hospitals. Manage Third-Party Services (DS2) covers the relationship between each party. Ensure System Security (DS5) evaluates the existence of supporting tools application. Educate and Train User (DS7) evaluates the training and education of users for each party.

COBIT Framework is created by ISACA for IT management and IT governance. It is a support tools for that allows managers to bridge gap between, control requirement, technical issues and business risks. COBIT is designed as a tool for the management of IT governance in understanding and managing the risks and benefits associated with information and related technology [8]. COBIT is a comprehensive set of resources that contains all the information that organizations need to adopt a framework of IT governance and control framework. And COBIT accommodate all types of IT related risks and offer solutions for all types of business organizations and all types of all platforms. COBIT has purpose for control, audit guidance, performance and outcome metrics, critical success factors and maturity model.

A. COBIT Framework

COBIT provides a parameter for the assessment of how well high and IT management in an organization by using maturity models that can be used for the assessment of management awareness (awareness management) and level of maturity (maturity level). According to reference [9], COBIT has a maturity model to achieve minimum standard maturity level define in level 3, at which the risk would be minimum. Table 1 defines the controlling of IT processes by using the method of assessment (scoring) so that an organization can assess IT processes owned from nonexistent to optimized scale (from 0 to 5).

TABLE I
MATURITY MODEL [7]

Level	Maturity Model
0 Non-existent	Complete lack of any recognizable processes. The Enterprise has not even recognized that there is an issue to be addressed
1 Initial/Ad Hoc	There is evidence that the enterprise has recognized that the issues exist and need to be addressed. There are, however, no standardized processes; instead, there are ad hoc approaches that tend to be applied on an individual or case-by-case basis. The overall approach to management is disorganized
2 Repeatable but intuitive	Process have developed to the stage where similar procedures are followed by different people undertaking the same task. There is no formal training or communication of standard procedures, and responsibility is left to the individual. There is a high degree of reliance on the knowledge of individuals and, therefore, errors are likely.
3 Defined Process	Procedures have been standardized and documented, and communicated through training. It is mandated that these processes should be followed; however, it is unlikely that deviations will be detected. The procedures themselves are not sophisticated but are the formalization of existing practices.
4 Managed and Measurable	Management monitors and measure compliance with procedures and takes action where processes appear not to be working effectively. Processes are under constant improvement and provide good practice. Automaton and tools are used in a limited or fragmented way.
5 Optimized	Processes have been refined to a level of good practice, based on the result of continuous improvement and maturity modeling with other enterprises. IT is used in an integrated way to automate the workflow, providing tools to improve quality and effectiveness, making enterprise quick to adapt

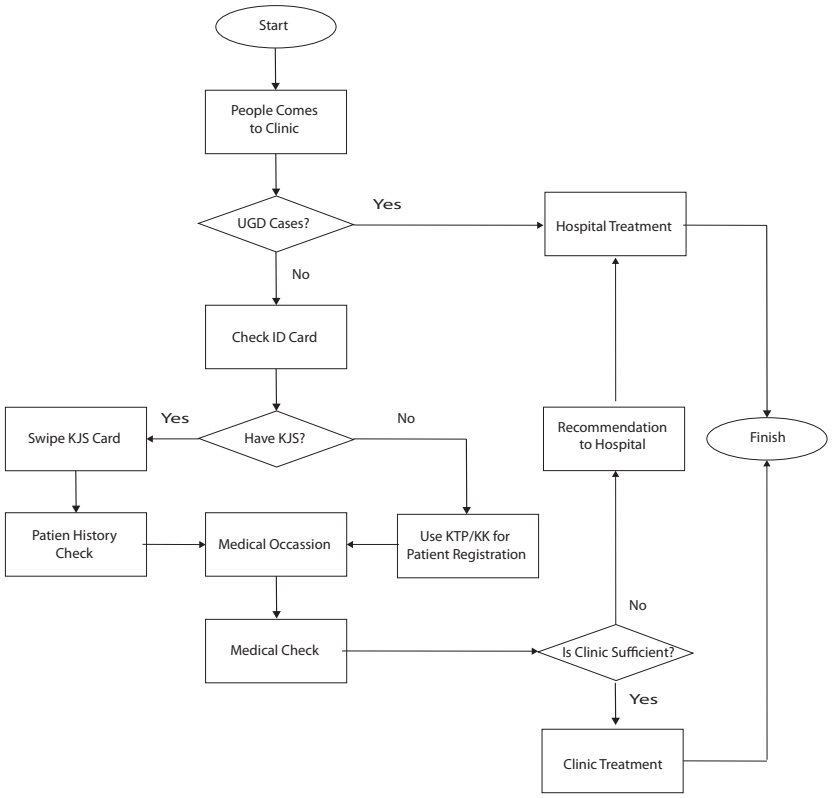


Fig. 1. Flowchart Treatment Clinic and Hospital

IV. METHODOLOGY

A. Maturity Index Calculations attributes

Maturity index of every attribute domain modules obtained from summing the number of respondents who answered for each attitude scale. Each module DS is multiplied by the weight scale and then divided by the number of respondents, such as the following :

$$Index\,Maturity\,Attribute = \frac{\sum(total\,Answer * Weight)}{Respondents\,Number} \quad (1)$$

B. Maturity index calculations

Assessment is done by considering the level of maturity index value at COBIT maturity attributes that includes :

1) Awareness and Communication (AC),
2) Policies, Standrads and Procedures (PSP),
3) Tools and Automation (TA),
4) Skill and Expertise (SE),
5) Responsibilities and Accountabilities (RA),
6) Goal Setting and Measurement (GSM).

Maturity index formula :

$$Maturity\,Level = \frac{\sum(Index\,Maturity\,Attribute)}{6} \quad (2)$$

TABLE II
REPRESENTATION OF INDEX MATURITY COBIT

Index Maturity	Maturity Level
0 0,50	0 Non Existent
0,51 1,50	1 Initial / Ad hoc
1,51 2,50	2 - Repeatable But Intuitive
2,51 3,50	3 Defined Process
3,51 4,50	4 Managed and Measurable
4,51 5,00	5 - Optimized

TABLE III
MATURITY LEVELS IN SUB-DISTRICT HEALTH CENTER

Domain	Process	Current Maturity
DS1	Define and Manage Service Levels	3.25
DS2	Manage Third-Party Services	1.73
DS5	Ensure System Security	3.33
DS7	Educate and Train Users	1.87
Result	2.54	

C. Representation of Index Maturity

After calculating Maturity Index, we will obtain maturity level for each sub-domain in Delivery and Support domain that is shown in Table II

V. RESULTS

The result of Maturity Level DS1, DS2, DS5, DS7 and sub-district health center can be read in table III.

TABLE IV
CURRENT MATURITY LEVEL OR AS-IS (CML) COMPARE TO EXPECTED MATURITYLEVEL OR TO-BE (EML)

Domain	Process	as-is	to-be
DS1	Define and Manage Service Levels	3.25	4
DS2	Manage Third-Party Services	1.73	4
DS5	Ensure System Security	3.33	4
DS7	Educate and Train Users	1.87	4

From the Table IV, it can be seen that the DS1 - Define and manage service levels are at maturity level 3.25, and DS2- Manage Third-party services are at maturity level : 1.73, and D5 5Ensure Securitysystem is the level of maturity :3.33, and DS7 - Educate and Train Users are at maturity level : 1.87 Thematurity level of Delivery and Support domain is illustrated in Fig. 2 that consists of current maturity level "as-is" and expected maturity level "to-be".

A. Analysis from the point of view society

Ten people is interviewed in around urban health center and sub-district health center in Tanah Tinggi (Central of Jakarta). Their age is around 19-44 years old. The questions are about how their experience use KJS. Furthermore we use cross case analysis using word table [10] to compare their answers. All respondents answered that the process of getting KJS is very easy because it only requires Jakarta's KTP and KK. Therefore all Jakarta community can obtain KJS. Eventhough many people do not have KJS, they can still use KTP and KK for treatment in clinics and hospital. They have same functions like KJS. The summary of the interview process is described as follow:

1) The respondents around "Tanah Tinggi" thought that program KJS is very pleased because it is very useful for treatment, but many people may misuse KJS like people rich treatment in clinics using KJS. Sometimes many people bring cars as their vehicles to clinic for treatment.
2) The respondents of "Tanah Tinggi" replied that KJS is very helpful because it is easy to referral from clinics for treatment to the hospital but they have several requirements that must be filled since they have rules from governor of DKI Jakarta. He stated that if residents of DKI Jakarta get sick, they can be treated in clinics for each region. However, if the clinics lack medical equipments for treatment, they can get a referral to the hospital.
3) The author interviewed to some respondents. They said that not all patient KJS get the same treatments like other patients. It all depends on the hospital rules because not all of hospital agree to the KJS. The establishedment of KJS makes the patient's turnover increases two times than usual.
4) The respondents said that the clinic's location and hospital for treatment that uses KJS is very easy to discovered because all clinics can treat patients that use KJS, plus the hospital. However, nowadays some hospital rejects the existence of KJS because there are some loss payments in several hospitals. KJS has a requirement if people wants to be treated. The patients can go through to clinics first. However, if the clinic lack of some medical equipments, the patients requires to get a referral to RSUD and if the RSUD also lack of medical equipments for treatment, people are required go to hospitals.
5) The author asked to respondents how they know the information about KJS and almost respondents answered they know KJS from social media like a newspaper and television or mouth to mouth.
6) Nowadays, KJS needs handling to be tightened because some patients that only get stiff sore, catch a cold, minor injuries and etc. go to be treated to hospital. For this matter, KJS needs extra improvement to avoid patients spoiled.
7) The author asked to respondents like doctors and nurses in clinics or hospitals. All of the respondents were quite satisfied with KJS program since it increases patients in hospital or clinics.
8) KJS can be used for long-term utilization because by using KJS, patients can be treated in the hospital room class III because the patients of KJS can only provided in class III.

After analysing the result, there are some suggestions for DKI Jakarta Government that are mentioned as follow:

1) If the patient is firstly examined as what is defined in previous program (re: GAKIN), there should be no overloaded patients in clinics and hospitals.
2) The rule should be changed from People of Jakarta is able to use KJS intolow middle class people of Jakarta is able to use KJS.
3) The suggestion for Governor of Jakarta is to add more RSUDs that can treat patients with KJS service in Jakarta since at this moment, there are only three RSUDs for KJS while there are five RSUDs in Jakarta.
4) Another suggestion for the Governor of Jakarta is to make more detail information about KJS,especially for several hospital providing KJS or service since there are still many people do not know about it.

B. Interview and questionnaire Framework COBIT

The analysis result of DS1 is described as follow:

1) PCARE Application provider has given SLA formal process to the health center for PCARE equipment and application installation for KJS, its cooperation agreement in which it meticulously describes about SLA scope definition, organizational structure, purpose, responsibility, in order to run SLA, which is saved in a document which explains organizational structure along with its role, purpose, and responsibility.
2) PCARE Application provider has covered tools and application service of clinic service about SLA saved in catalog.

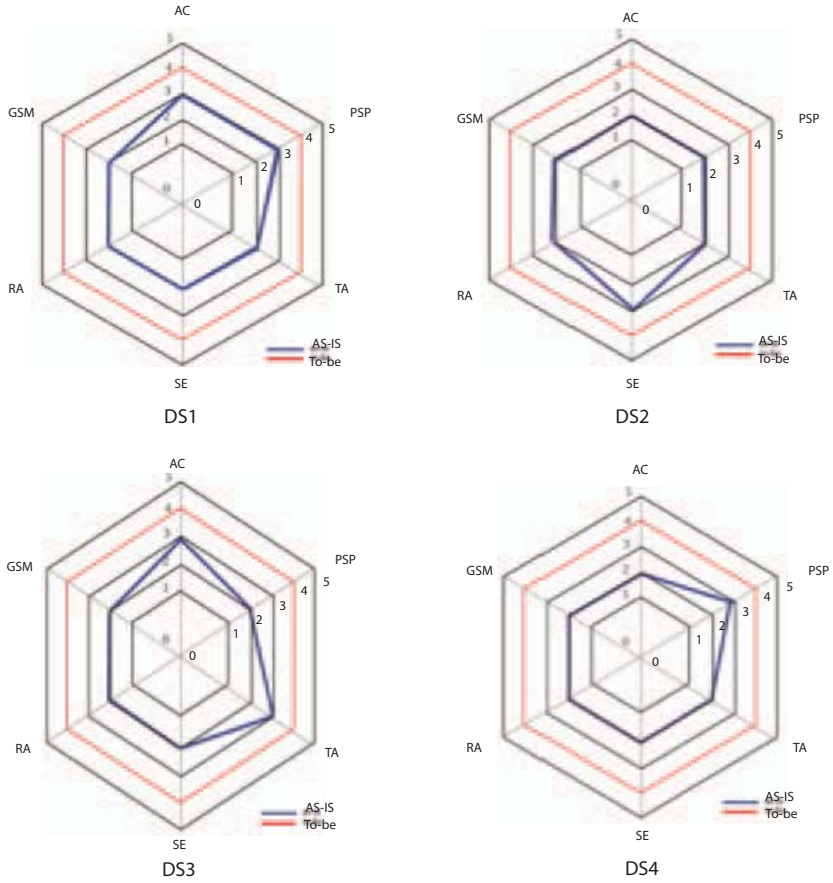


Fig. 2. Graphic Analysis "as-is" and "to-be" COBIT Framework

3) SLA (Service Level Agreement) agreement between PCARE application provider and health center has been approved by the head of it but it does not have the capability of a health center, and therefore it needs more human resource to manage it.
4) PCARE Application provider has given OLA (Operating Level Agreement) to the health center although only a few staff know whether they have been equipped with OLA which supports SLA or not.
5) Health center has given report to PCARE Application provider using standard format about the previous agreement, in case the tools and application system are damaged.
6) PCARE application provider discuss the execution of this agreement with the health center to renew the progress of the agreement such as meeting outside of the office once a month or two, but not in a from of SLA.

The analysis result of DS2 is described as follow:

1) PCARE Application provider has been able to identify supplier category which becomes income of PCARE tools and application, and up to now the relationship with the supplier is in a good condition, shown by sudden maintenance or replace when there is damage of EDC tools which can reduce the health center performance.
2) PCARE Application provider guarantee good quality of connection between PCARE application provider and supplier, so if the health center needs EDC spare parts then health center can get it as soon as possible.
3) PCARE Application provider can give efficient and effective service, from maintenance, tools supplier, and problem solving. This is shown by the repair of the damage of EDC tools when there is no cost receipt.

The analysis result of DS5 is described as follow:

1) PCARE application provider has framework which control the IT security within some aspects. Some of them are in the form of identification, authentication, and the establishment of ones right of an application.
2) PCARE application provider has arranged an IT security plan by setting id and password for every user to minimize risk in order to keep safe its secrecy. Even though the username and password are not unique, the password can be implemented periodically by using a tool/ module in the application.
3) PCARE application provider can modify a user if there is a change of personal data/referral letter of a tool/ module in the application.
4) Health center staff will not know whether PCARE application provider check the security periodically or

not.

5) Health center can communicate with PCARE tools and application provicer if there is a system damage.
6) Health center does not know whether PCARE application has advanced a technology related to security so that its application is immune to threat or not.
7) PCARE application provider has not used cryptography to secure the data.
8) PCARE application provider has not used antivirus to protect PCARE application.
9) PCARE application provider has not used firewall, security application, and network to detect and control the flow of information into and out of network.
10) PCARE application provider has been able to give service which process that data transaction can be done in good control, so it can guarantee the authenticity of the data. Within the transaction process, PCARE application user does not intervene it, and all of the data transaction can be done through computer.

The analysis result of DS7 is described as follow:

1) PCARE Application provider has compiled and renew the curriculum to identify the education and set the need of training in using KJS application, and it has been proved by the module of the instruction of using PCARE application
2) PCARE Application provider has given good education material, so that the training material can be distributed and understood, because every month there will be training and updates about PCARE Application.
3) Health center doesn't evaluate the previous training a long side PCARE application provider.

The result of this research is expected to be a reference. Based on the analyis of DS1, DS2, DS5, and DS7, we suggest that local clinic should require PCARE application provider to do following several improvements :

1) Clinic has to ask PCARE application to do some deeper analysis to the quality of service provided by PCARE application provider.
2) Clinic has to ask PCARE application provider to make it be more secured system.
3) PCARE application provider must hold training for staff in each clinic not only in the form of use, but also in its feature. The conduct of this research was done manually, especially at the time of data collection and process. Thus, the researcher suggest the next researcher to make a certain supporting software which can be used for online questionnaire filling and data processing to consider clinic spread in all over Jakarta to make it more effective and efficient

VI. Conclusions

In conclusion, the KJS system based on COBIT assessment is mature. The research result of DS1, DS2, DS5, DS7 maturity level domain Delivery Service provided by PCARE Application provider to clinic is in the level of 3 (Defined Process). It is satisfied based on categorization of maturity level system as mentioned in reference [9]. To avoid abusement, this procedure must be followed well. Even though the procedure has not complete yet, it has been practically formalized and run well.

Furthermore, the KJS program is effective for society especially lower middle class society. However, there are some misuse cases in the usage of KJS. For example, rich people use KJS for free treatments in clinic. By implementing KJS program, increasing number of patients in hospital and clinic cannot be avoided. The middle-class people also takes advantages of this KJS program as well as people in unserious disease such as itchiness or caught in a cold.

Since this research only covers Clinics in Tanah Tinggi region, it would be better if this research covers other regions too. Examples of the regions are Tambora in Jakarta Barat, Pademangan Timur in Jakarta Utara, etc. This research used Delivery and Service domain Module (DS1, DS2, DS5, DS7) using COBIT Framework. In further research, other framework can also be used and then compare with the current result.

Acknowlegment

I would like to say thanks to all workers in clinic. They helped me a lot so I can collect the needed data.

VII. References

[1] Government Jakarta. (2013) Daerah khusus ibukota jakarta. [Online]. Available: http://id.wikipedia.org/wiki/Daerah Khusus Ibukota Jakarta
[2] B. Nurhidayat, "Evaluasi integrated toll collection system dengan menggunakan framework cobit," M.Sc. thesis, Institut Pertanian Bogor, Bogor, Indonesia, 2011.
[3] Sultani, "Pengembangan applikasi audit sistem informasi berdasarkan cobit framework di rumah sakit xxx," in *Proceedings of Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2012 (Semantik 2012)*, 2012.
[4] W. Hartono, "Analisa kesenjangan tata kelola teknologi informasi untuk proses pengelolaan data menggunakan cobit," M.Sc. thesis, Institut Teknologi Sepuluh November, Surabaya, Indonesia, 2008.
[5] P. AdMedika, "Edc tools."
[6] Warta DKI. (2013) Pt askes mengelola dana kartu jakarta sehat. [Online]. Available: http://wartadki.com/berita-587-ptaskes- mengelola-dana-kartu-jakarta-sehat-.html
[7] ISACA, "Cobit 4.1," 2007.
[8] M. Salee, "It service management and it governance : review, comparative analysis and their impact on utility computing," Hewlett-Packard Company, 2004.
[9] J. Ivanyos. (2009) Implementing cobit based process assessment model for evaluation it controls. [Online]. Available: http://training.ia-manager.org/file.php/1/Summary_of_ITCA_Principles_V1.pdf
[10] M. Galinium and N. Shahbaz, "Factors affecting success in migration of legacy systems to service-oriented architecture (soa)," M.Sc. thesis, Lund University, Lund, Sweden, Jun. 2009.

Automatic Exudate Extraction for Early Detection of Diabetic Retinopathy

Syna Sreng
International College
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
srengsina@yahoo.com
s4601152@kmitl.ac.th

Noppadol Maneerat, Don Isarakorn,
Bundit Pasaya
Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
kmnoppad@kmitl.ac.th

Jun-ichi Takada
Graduate School of Science and Engineering
Tokyo Institute of Technology
Tokyo, Japan
takada@ide.titech.ac.jp

Ronakorn Panjaphongse M.D
Ophthalology Department
Bhumibol Adulyadej Hospital
Royal Thai Air Force
Bangkok, Thailand
ronakorn_p@rtaf.mi.th

Ruttikorn Varakulsiripunth
Faculty of Information Technology
Thai-Nichi Institute of Technology
Bangkok, Thailand
kvruttik@hotmail.com

***Abstract*—Diabetic Retinopathy (DR) is the most common cause of blindness in diabetic patients, but early detection and timely treatment can prevent this problem. Exudates have been found to be one of the signs and serious DR anomalies so the proper detection of these lesions and the treatment should be done immediately to prevent loss of vision. The aim of this study is to automatically detect these lesions in fundus images. To achieve this goal, the proposed method first preprocesses to improve the quality of fundus image, and then Optic Disc (OD) is detected and eliminated to prevent the interference to the result of exudate detection by combination of 3 methods; image binarization, Region Of Interest (ROI) based segmentation and Morphological Reconstruction (MR). Next, exudates are detected by applying the maximum entropy thresholding to filter out the bright pixels from the result of OD region eliminated. Since the result contains some noises which appear as bright light at the edge of fundus area in some images, that affect is considered and eliminated to improve the result of false positive. Finally, exudates are extracted by using MR. The proposed technique has been tested on 100 fundus images from hospital. Experimental results show that 91 % of exudate is extracted correctly with the average process of 3.92 second per image.**

Keywords— optic disc; exudate; diabetic retinopathy; maximum entropy thresholding; morphological reconstruction

I. INTRODUCTION

Diabetes mellitus is a group of diseases characterized as the chronic condition in which there is an excess of glucose circulating in the bloodstream [1]. The effect of diabetic mellitus on the eye is called DR. It is a visual complication and becomes the most common cause of visual loss and blindness. Early detection and timely treatment can prevent the problem. For this reason, all diabetic patients should get their both eyes examined quit often (at least once every year). Ophthalmologists play a most important role to help diabetic patients for an eye examination. However, the number of ophthalmologists is limited in Thailand. The study of Thailand diabetic project shows that about one thousand ophthalmologists have to work with six million diabetic patients. Among all of them about 30 percent of patients are effected by DR [2]. Therefore, automatic retinal image analysis is very useful to help ophthalmologists in detection and diagnosis of DR. DR clinical signs includes microaneurysms, microvascular abnormalities, hemorrhages, cotton-wool spots and exudate. Exudate is among the signs of DR so the proper detection of these lesions is an essential task. It appears as bright lesions with the random yellowish deposits of varying size, shapes and location in retinal images. On the research work leading to automatic analysis of exudate detection, the knowledge of OD location is very useful. OD is known as the bright circular region which appears with similar features as exudates. So, in order to prevent the result of exudate detection from the interference of OD, it should be removed out first.

In the past researches, exudate detection was presented in different categories. Thresholding methods base on intensity levels and color image [3-5]. A combination of region growing and edge detection was used in [6, 7]. The classification methods build a feature vector for each pixel classification by

This work is financial supported by AUN SEED-Net program under the Research collaboration scheme, Coordinating Center for Thai Government Science and Technology Scholarship students (CSTS) and National Science and Technology Development Agency (NSTDA).

using a naive Bayes classifier [8] and color images clustering [9], as well as a neural network classifier was used in [10].

The aim of this study is to automatically detect exudates in fundus images. To achieve the goal, this paper proposes a new algorithm to extract exudates based on maximum entropy thresholding and morphological reconstruction, which falls into the category of thresholding method. This proposed method does not require any training set which can make common issue to be processed. In addition, training sets method requires large amounts of data therefore it takes long time for processing and can be difficult to obtain in the real application. Thresholding method is used in this work because it does not require high computing power and takes short time for processing. Computing power and processing time are very important for application in rural area in developing country where both expert ophthalmologists and high performance computers are rarely available.

In the previous work, we proposed the method for detection of exudate by calculating the pixel count for estimating the histogram of the gray level image and blob boundary detection as well as MR for elimination of OD first. Then maximum entropy thresholding was applied for exudate detection [11]. On verification the result from the ophthalmologists, the accuracy of 89% was obtained. Since this result is limited by variable intensity in retinal images, we perform a new method to measure the maximization of information by using maximum entropy threshold. Next ROI based segmentation and MR are used for the step of OD detection and elimination. Then maximum entropy thresholding is applied again on result of OD region eliminated. After that the bright light which appears at the edges of retinal area in some images is also considered and eliminated to improve the result of false positive. Finally, exudates are extracted by using MR. Overall accuracy of 91% is successful.

The rest of paper is organized as follows: section II briefly summarizes the different stages of the proposed algorithm. Section III describes the results and discussions, and finally section IV is conclusion.

II. METHODOLOGY

The proposed method for exudate detection is composed of three main stages: *A.* preprocessing, *B.* OD elimination and *C.* exudate extraction.

A. Preprocessing

Abnormality detection in fundus image is very complicated with the differences in luminosity, brightness and contrast inside fundus images. Preprocessing step is very important to solve these problems. The fundus images database from hospital are stored in RGB color space which consists of three components (red, green and blue). Since the process of this proposed method has two main steps (OD and Exudate detections) so the preprocessing step has to consider in both cases. From the analysis of RGB component of the images, we found that the exudate is more easily discriminated in green component because it has higher contrast and more energy [12-13], but the OD is more easily discriminated in red component [14-17] because it is brighter in intensity. Fig. 1 shows the original image with its red, green and blue components. This figure illustrates that in green component, the OD is separated by blood vessel. Therefore it is difficult to define the shape of OD when the image is converted to binary image, but in the red component, the OD is easily defined with the characteristic of the most circular region after binary segmentation.

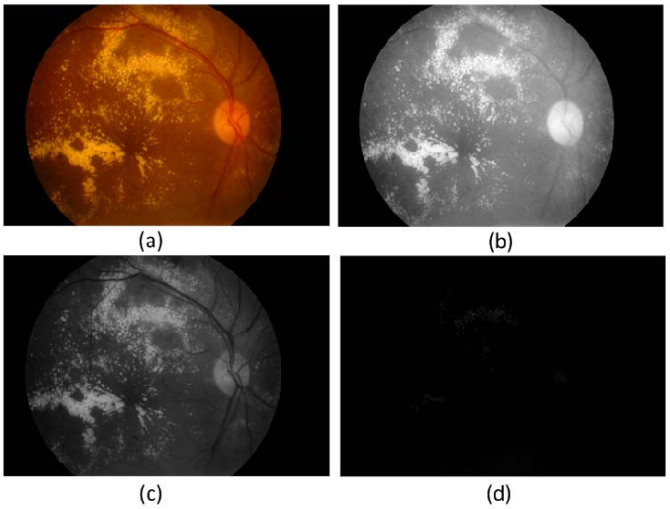


Figure 1. (a) Original Image (b) Red Component (c) Green Component (d) Blue Component

For analysis using both green and red components, first the original image is converted into the red component (for OD detection) then into the green component (for exudate detection), and finally median filtering is applied to reduce noise on the green component.

B. OD Elimination

The OD is referred to the optic nerve inside our eye. It appears as a bright circular region which is similar characteristic to the exudate. In order to prevent the result of exudate detection from the interference of OD, it should be removed out first [5, 8, 18]. In this paper, OD elimination detection involves three steps as shown in following:

- Image binarization
- ROI based segmentation
- Morphological reconstruction

1) Image binarization

In clinical environment, the variable intensity in fundus images is considered as one of the main obstacles for exudate detection as well as OD detection because the range of intensity value varies from one image to another. To take these differences into account, we measure the maximization of the information between object and background by using maximum entropy thresholding described in [19, 20]. In this research the entropy function is studied on 1D histogram of the

gray level image. Suppose i is the gray level value of an image so $i = [0, 1 \dots t-1, t, t+1 \dots 255]$, and the probability of the pixel being i can be defined as:

$$P_i = \frac{n_i}{n} \quad (1)$$

where n is the total number of pixels in the image and n_i is the number of pixels that have gray level i.

Let o and b represent the object and background respectively thus the probabilities of o and b can be calculated as:

$$P_{\mathrm{o}} = \sum_{i=0}^{t-1} P_i \quad (2)$$

$$P_{\mathrm{b}} = \sum_{i=t}^{255} P_i \quad (3)$$

So the entropies of object and background's probability distributions are given as:

$$H_{\mathrm{o}}(t) = -\sum_{i=0}^{t-1} \frac{P_i}{P_{\mathrm{o}}} \log_2 \frac{P_i}{P_{\mathrm{o}}} \quad (4)$$

$$H_{\mathrm{b}}(t) = -\sum_{i=t}^{255} \frac{P_i}{P_{\mathrm{b}}} \log_2 \frac{P_i}{P_{\mathrm{b}}} \quad (5)$$

Thus the entropy of the gray level image segmented by threshold t is:

$$H(t) = H_{\mathrm{o}}(t) + H_{\mathrm{b}}(t) \quad (6)$$

The theory of maximum entropy is used to select t which makes H becomes maximum value. So, the optimal threshold t can be selected by maximizing the entropy of $H(t)$ which can be defined as:

$$t = \mathrm{Arg\ Max}(H(t)) \quad (7)$$

where $0 \leq t \leq 255$

The object is detected as the bright pixels after apply the level of threshold t on the red component of fundus image.

2) ROI based segmentation

In this method the OD is extracted by finding the region of interest based on two processes; neighborhood connecting pixels and compactness measurement, respectively. After binarization, there are more than one candidate regions identified. So, first we applied neighborhood connecting pixels to label all group pixels then all small holes of the group pixels is filled to make them smooth. Finally, compactness measurement in [21] is used to find the most circular region which is considered as OD. The maximum value of compactness will be 1 if the region is a circle. This value will decrease depend on its shape. For example elliptical-shaped regions, and irregular-shaped regions, as well as the other complicated boundaries regions. The equation of compactness measurement is defined as:

$$C(\mathcal{R}) = \frac{4\pi A(\mathcal{R})}{l_{\mathrm{p}}^2(\mathcal{R})} \quad (8)$$

where $\mathcal{R}$ is the connected region with the pixels more than twenty thousands, as all connected region with the pixels less than twenty thousands are considered as noise and are eliminated. $A(\mathcal{R})$ is the number of pixels in region $\mathcal{R}$ and $l_p(\mathcal{R})$ is the length of boundary region $\mathcal{R}$.

3) Morphological reconstruction

The process of MR is based on dilation on two images, a marker and a mask. The resulting image from ROI based segmentation is used as a marker and the result of green component from the preprocessing is used as a mask. All the pixels in the marker image are overlaid on the mask image by the dilation of the marker image repeatedly until the contour of the marker image fits under the mask image. The expression is defined as:

$$\mathrm{OD_g} = R_B(\mathrm{F}) \quad (9)$$

where B is the mask image, F is the marker image, and $\mathrm{OD_g}$ is the reconstruction of B from F, which is the result of OD region eliminated. The process and result of OD elimination is shown in Fig. 2.

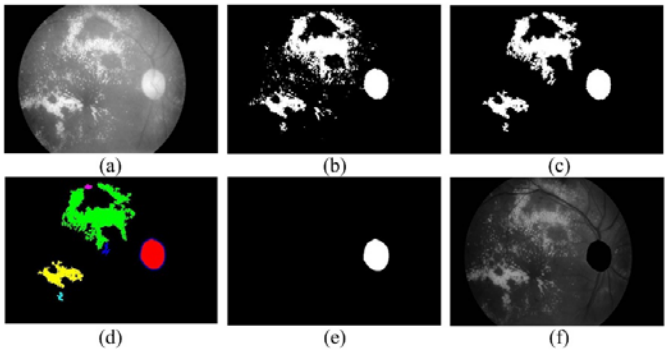


Figure 2. OD Elimination. (a) Red Component of Retinal Images (b) Result of Binarization (c) Removing Small Pixels (d) OD Detection (e) OD Mask (d) Result of OD Region Eliminated

C. Exudate Extraction

In some cases the interference of the exudates detection is not only the OD, but also the bright light which appears at the edge of fundus area depending on the quality of the images. Fig. 3 shows these kinds of poor quality images which are also included in this database.

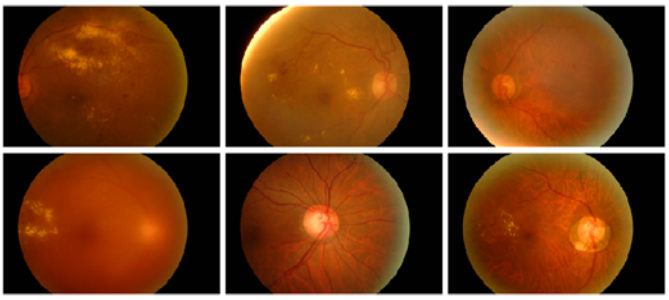


Figure 3. Fundus Images which Bright Light Appears at the Edge of Its Fundus Area

The fundus images obtained from a hospital consist of a circular fundus and a dark background surrounding the fundus area so in order to detect the bright light affect first we have to make that affect becomes the edge of fundus image by cropping the result image after OD elimination. Then, maximum entropy thresholding apply in image binarization is used to filter out the left bright objects that are considered as the exudate mask. Since the result of this work not contains only exudates but also some noises from the effect of light at the edge of retinal area in some cases so clearing border objects [22] which is a part of morphological image processing is used. Its syntax is:

$$C_b = \text{imclearborder}(C_r, \text{conn}) \quad (10)$$

where C_r is the cropping result of OD region elimination and C_b is the result of exudates mask. The value of conn is 8 because 8-connected neighborhood is used.

Finally exudates mask which obtained is inverted before they are overlaid on the original retinal image to extract the lesions by utilizing Eq. (9), where the inverted exudate mask is used as the marker image, and the original image is used as the mask image. Fig. 4 illustrates the result of this process.

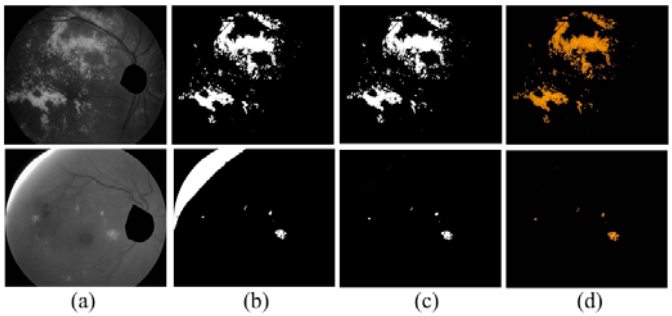


Figure 4. Exudates Extraction. (a) Cropping the Results of OD Elimination (b) Binarization after Cropping (c) Exudates Mask after Removing the Bordered Objects (d) Results of Exudates Extraction

III. RESULTS AND DISCUSSIONS

In this study, 100 fundus images with variable color, brightness and quality are used. The images were obtained from Bhumibol Adulyadej Hospital with the resolution of 3872x2592 pixels in 24 bit JPEG format. These images contain all lesions related to DR (microaneurysms, microvascular abnormalities, hemorrhages, cotton-wool spots, exudate, etc.) and healthy images. The accuracy is the overall correctness of the system and is calculated as the sum of correct classifications divided by the total number of classifications as define in Eq. (11).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FN+FP}} \quad (11)$$

From Eq. (11), TP, TN, FP and FN are numbers of classifications. TP is when all exudates are correctly identified as exudate. FN is when some exudates are incorrectly identified as non-exudate. TN is when all non-exudates are correctly identified as non-exudate and FP is when some non-exudates are incorrectly identified as exudate. Performance evaluation has been done over all the images by the verification of ophthalmologists. The result shows that 91 % of exudates detection was successful. The reason that the proposed method cannot achieve 100% results because of presence of noise and artifacts in some images as well as the quality of retinal images. This result can be improved by choosing a better algorithm to discriminate the result of exudate detection with the seldom noises which appear as false positive. However, the proposed method does not require any training set which can make common issues with human segmentation inconsistencies. The training sets method requires large amounts of data and training process. Therefore it can be difficult to obtain in the real application. Although this result is not achieve 100%, the average process of this research takes only 3.92 second per image so it can be used to help the Ophthalmologists for the real application if a better algorithm is investigated to improve the false positive.

Because of many methods have been published throughout the years using a variety of datasets, and the lack of a database with their ground truth for retinal structures, so it is difficult to compare our algorithm with the other reported work in the literature. This is the reason that many researchers only show their methods and results, but do not compare with the past researches [3, 4, 8, 11, 17, 18, 23, 24]. However some methods of [3, 5, 8, 10, 13, 16, 17, 23, 24] do not report the average processing time per image and the other methods of [9, 25, 26] report the lower performances to our approach with the same average processing time of 3 minutes.

IV. CONCLUSION

This paper proposed the method to extract exudates for early detection of DR. First, preprocessing is performed to improve the quality of fundus image, and then OD is detected and eliminated to prevent the interference to the result of exudate detection by combination of 3 methods; image binarization, ROI based segmentation and MR. Next the maximum entropy threshold is applied on result of OD region eliminated to filter out the exudate mask. Since the result contains some noises which appear as bright light at the edge of retinal area in some images, that affect is considered and eliminated to improve the result of false positive. Finally, exudates are extracted by using MR. Overall accuracy of 91%

was successful with the average process of 3.92 second per image.

For the future work, the reason of the failure in some images should be more carefully examined, so that a new algorithm can be investigated to improve the present work as real application software for early detection of diabetic retinopathy screening system.

ACKNOWLEDGMENT

This research is supported by Coordinating Center for Thai Government Science and Technology Scholarship students (CSTS) National Science and Technology Development Agency (NSTDA). The authors also would like to thank AUN/SEED-Net for financial support of one of the authors (Mr. Syna Sreng) for his study at King Mongkut's Institute of Technology Ladkrabang (KMITL), THAILAND.

REFERENCES

[1] P. H. Scanlon, C. P. Wilkinson, S. J. Aldington and D. R. Matthews, A Practical Manual of Diabetic Retinopathy Management. Wiley-Blackwell, Oxford, UK, 2009.

[2] T. Chetthakul, C. Deerochanawong, S. Suwanwalaikorn, N. Kosachunhanun, C. Ngarmukos, P. Rawdaree, et al., "Thailand diabetes registry project: prevalence of diabetic retinopathy and associated factors in type 2 diabetes mellitus," J Med Assoc Thai 89, no. Suppl 1, S27-36, 2006.

[3] C. I. Sánchez, M. García, A. Mayo, M. I. López, and R. Hornero, "Retinal image analysis based on mixture models to detect hard exudates," Medical Image Analysis 13, vol. 4, pp. 650-568, 2009.

[4] L. Giancardo, F. Meriaudeau, T. P. Karnowski, Y. Li, K. W. Tobin, and E. Chaum, "Automatic retina exudates segmentation without a manually labelled training set," In Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium, pp. 1396-1400, IEEE, 2011.

[5] B. Venkatalakshmi, V. Saravanan, and G. Niveditha, "Graphical user interface for enhanced retinal image analysis for diagnosing diabetic retinopathy," Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference, pp. 610-613, IEEE, 2011.

[6] H. Li, and O. Chutatape, "Automated feature extraction in color retinal images by a model based approach," Biomedical Engineering, IEEE Transactions, vol. 51, no. 2, pp. 246-254, 2004.

[7] C. Sinthanayothin, J. F. Boyce, T. H. Williamson, H. L. Cook, E. Mensah, S. Lal, and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images," Diabetic Medicine, vol. 19, no. 2, pp. 105-112, 2002.

[8] A. Sopharak, K. T. Nwe, Y. A. Moe, M. N. Dailey, and B. Uyyanonvara, "Automatic exudate detection with a naive Bayes classifier," Proceedings of the 2008 International Conference on Embedded Systems and Intelligent Technology, pp. 139-142, 2008.

[9] A. Osareh, B. Shadgar, and R. Markham, "A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images," Information Technology in Biomedicine, IEEE Transactions on, vol. 13, no. 4, pp. 535-545, 2009.

[10] M. Garciaa, C. I. Sáncheza, M. I. Lópezb, D. Abásoloa, and R. Horneroa, "Neural network based detection of hard exudates in retinal images," Computer methods and programs in biomedicine, vol. 93, no. 1, pp. 9-19, 2009.

[11] S. Sreng, N. Maneerat, D. Isarakorn, R. Varakulsiripunth, B. Pasaya, J. Takada, and R. Panjaphongse, "Feature extraction from retinal fundus image for early detection of diabetic retinopathy", unpublished.

[12] J. Nayak, R. Acharya, P. S. Bhat, N. Shetty, and T. C. Lim, "Automated diagnosis of glaucoma using digital fundus images," Journal of medical systems, vol. 33, no. 5, pp 337-346, 2009.

[13] S. Kavitha, S. Karthikeyan, and K. Duraiswamy, "Neuroretinal rim Quantification in Fundus Images to Detect Glaucoma,". IJCSNS, vol. 10, no. 6, pp 134, 2010.

[14] H. F. Jaafar, A. K. Nandi, and W. Al-Nuaimy, "Detection of exudates from digital fundus images using a region-based segmentation technique," In 19th European Signal Processing Conference EUSIPCO, 2011.

[15] M. Esnaashari, S. A. Monadjemi, and G. Naderian, "A content-based retinal image retrieval method for diabetes-related eye diseases diagnosis," International Journal of Research and Reviews in Computer Science (IJRRCS), vol. 2, no. 6, 2012.

[16] K. Saranya, B. Ramasubramanian, and S. Kaja Mohideen, "A novel approach for the detection of new vessels in the retinal images for screening Diabetic Retinopathy," Communications and Signal Processing (ICCSP), 2012 International Conference, pp. 57-61, IEEE, 2012.

[17] A. Feroui, M. Messadi, I. Hadjidj, and A. Bessaid, "New segmentation methodology for exudate detection in color fundus images," Journal of Mechanics in Medicine and Biology, vol. 13, no. 01, 2013.

[18] K. Wisaeng, N. Hiransakolwong, and E. Pothiruk, "Automatic Detection of Exudates in Diabetic Retinopathy Images," Journal of Computer Science, vol. 8, no. 8, pp 1304-1313, 2012.

[19] M. Azarbad, A. Ebrahimzade, and V. Izadian, "Segmentation of Infrared Images and Objectives Detection Using Maximum Entropy Method Based on the Bee Algorithm," International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM), vol.3, pp. 026-033, 2011

[20] Shannon, C.E. and W. Weaver (1949), "The Mathematical Theory of Communication (Urbana, lL." University of Illinois Press 19, no. 7, 1949.

[21] N. Jamil, and Z. A. Bakar, "Shape-Based Image Retrieval of Songket Motifs," In 19th Annual Conference of the NACCQ, pp. 213-219, 2006.

[22] http://www.mathworks.com.

[23] R. SriRanjini, and M. Devaki, "Detection of Exudates in Retinal Images Based on Computational Intelligence Approach," International Journal of Computer Science and Network Security, vol.13, no. 3, pp. 86, march 2013.

[24] A. Sopharak, B. Uyyanonvara, and S. Barman, "Automatic exudate detection from non-dilated diabetic retinopathy retinal images using fuzzy C-means clustering," Sensors 9, no. 3, pp. 2148-2161, 2009.

[25] A. Sopharak, B. Uyyanonvara, S. Barman, and T. H. Williamson, "Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods," Computerized Medical Imaging and Graphics 32, no. 8, pp. 720-727, 2008.

[26] A. Sopharak, B. Uyyanonvara, and S. Barman, S. Vongkittirux, and N. Wongkamchang, "Fine exudate detection using morphological reconstruction enhancement," International Journal of Applied Biomedical Engineering vol.1, no.1, pp. 45-50, 2010.

Automatic Microaneurysms Detection Through Retinal Color Image Analysis

Preeyaporn Yunuch, Noppadol Maneerat,
Don Isarakorn, Bundit Pasaya
Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
mind_preeya@hotmail.com

Ronakorn Panjaphongse M.D
Ophthalology Department
Bhumibol Adulyadej Hospital
Royal Thai Air Force
Bangkok, Thailand
ronakorn_p@rtaf.mi.th

Ruttikorn Varakulsiripunth
Faculty of Information Technology
Thai-Nichi Institute of Technology
Bangkok, Thailand
kvruttik@hotmail.com

***Abstract*— This paper proposes an automatic system to diagnose the diabetic retinopathy symptom, which can cause a loss of vision by analysis the abnormality in retinal image. Digital image processing system is developed for the retinal image analysis which helps ophthalmologists to identify diabetic patients. The retinal images derived from ophthalmologists are used to analysis by using HSV, area identification and eccentricity techniques to distinguish diabetic retinopathy symptoms from normal diabetic patients. First color bar is evaluated by using HSV method and then using the eccentricity technique with area of pixel to find out the abnormality of Microaneurysms (MAs). The accuracy result of experiment is around 93% when compares to the analysis of ophthalmologists.**

Keywords—microaneurysms; diabatic retinopathy; ophthalmologists ;

I. INTRODUCTION

Diabetic retinopathy is a complication which causes blindness in diabetic patients. High level of glucose in diabetic patient blood circulation may affect blood vessels in retina which will make the blood vessel loss its elasticity. In the prodromal stage, this disease shows no other symptoms to be normally detected. Only the retinal image analysis should be used to diagnose the disease. Diagnosis of the symptom in prodromal stage allows curing to be done in order to reduce risk that will cause blindness in diabetic patients. In this paper an algorithm to diagnose diabetic retinopathy is proposed in order to assist ophthalmologists identifying suspected patient which will most likely reduce the workload of ophthalmologists.

This method is used for the abnormalities identification analysis and future model development. The retinal image is converted into the HSV color model and changed to a binary image to define the target area. The eccentricity technique is used to distinguish noise and veins from MAs area by considering its area and range. These techniques increase the accuracy of MAs detection.

For the initial diagnosis, MAs detection technique is implemented. There are several published papers proposed methods to diagnose MAs which are the earliest signs of diabetic retinopathy. Mendenca et al. [1] proposed a technique to detect MAs by using Gaussian Filter. Sophark et al. [2] proposed a technique to find the proper value by adjusting morphological operators to detect MAs in case of MAs shape and blood vessels are not clear. Shanin et al. [3] proposed a system to separate the abnormal retinal image by using the morphological and thresholding techniques in order to identify the edge and increase the image contrast. Yen et al. [4] proposed a method by using the histogram equalization and morphological operation techniques. Saleh et al. [5] proposed the grading of severity level of DR based on detecting and analyzing the early clinical signs associated with the disease, such as MAs.

Most of techniques used for MAs detection are color detection or defining other components, but in the real diagnosis, the colors and components in retinal images also MAs are different according to personal heredity. In this paper, HSV color space is used with the eccentricity technique to find out the abnormality of MAs.

The rest of this paper is organized as follows: methodology is presented in section II. Section III is result which is come to evaluate this work. Finally conclusion is shown in section IV.

This work is financial supported by Coordinating Center for Thai Government Science and Technology Scholarship students (CSTS) and National Science and Technology Development Agency (NSTDA).

II. METHODOLOGY

This paper proposes method to identify MAs by using the combination of these techniques: HSV color model, area identification and eccentricity techniques as shown in Fig. 1 The fundus image has been used for analysis diabetic retinopathy symptom first and then the red color pixels are filtered by using HSV method to find the threshold level. After that the filtered image is converted to a binary image which is easier to analyze the symptom. Finally, the binary image is filtered by the area filter to find eccentricity.

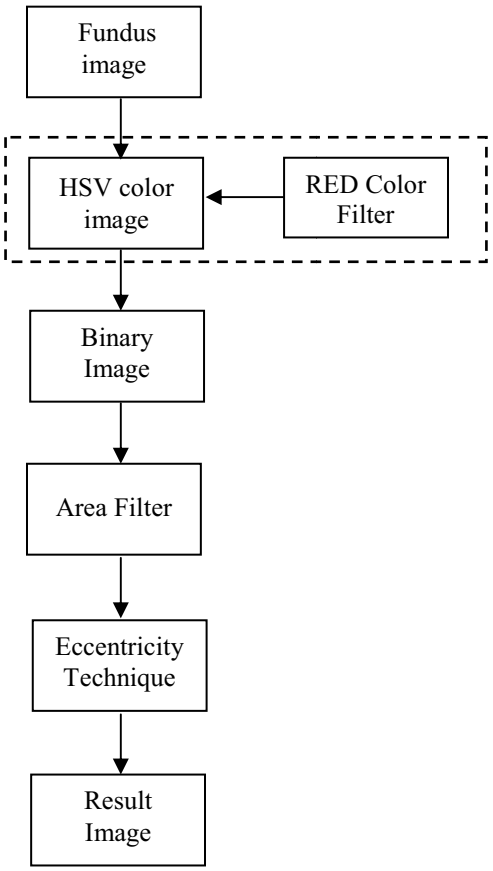


Fig. 1. Flow Chart of This Work

A. HSV Color Model

HSV color model is the method considering hue, saturation and value of image. H is the value of main color (green blue) which is in range of 0–255. If the hue value is 0, it will be red color and if the hue value is increased, the color will be changed as spectrum color until the value is 255 and the color will be red again. S means the saturation of color and V value means the brightness. In the lowest V value, it is shown as black color and the maximum V value will be white color. By this reason, this method is proper to be used for the retinal disease analysis. HSV color model is also the most similar color to the human color perception. HSV color model is a cone shape as show in Fig.2.

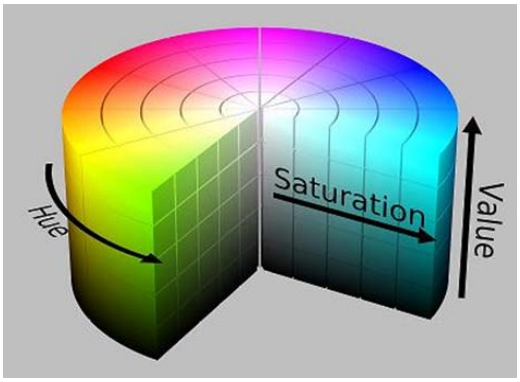


Fig. 2. HSV Color Model Cylindrical Coordinate [6]

To convert RGB color to HSV color, the following equation are used.

$$Hue = 60^{\circ} \times H' \quad (1)$$

$$Satuation = \frac{\max(\mathrm{R},\mathrm{G},\mathrm{B}) - min(R,G,B)}{max\ (R,G,B)} \quad (2)$$

$$Value = \max(R,G,B) \quad (3)$$

$$\mathrm{H'} = \begin{cases} \frac{\mathrm{G}-\mathrm{B}}{\max(\mathrm{R},\mathrm{G},\mathrm{B}) - \min(\mathrm{R},\mathrm{G},\mathrm{B})} & if\ \max(\mathrm{R},\mathrm{G},\mathrm{B}) = R \\ 2 + \frac{\mathrm{G}-\mathrm{B}}{\max(\mathrm{R},\mathrm{G},\mathrm{B}) - \min(\mathrm{R},\mathrm{G},\mathrm{B})} & if\ \max(R,G,B) = G \\ 4 + \frac{\mathrm{G}-\mathrm{B}}{\max(\mathrm{R},\mathrm{G},\mathrm{B}) - \min(\mathrm{R},\mathrm{G},\mathrm{B})} & if\ \max(R,G,B) = B \end{cases}$$

The advantage of HSV method is the flexibility. In case that the new color tones of retina images are derived in the system, the new target color may have to be analyzed and adjusted to find a new suitable color for this system. This reason may be able to support the diversity of human eye component and also the tone of image received.

B. Microaneurysm Detections

The samples of MAs colors in each color bar are analyzed to define the average of H_{min}, H_{max}, S_{min}, S_{max}, V_{min} and V_{max} in order to identify the range of target color.

To find the suitable target color that has most correctness to identify MAs in retina images, many color bar are arranged to be compared. Each color bar is composed of 10 random MAs captured area.

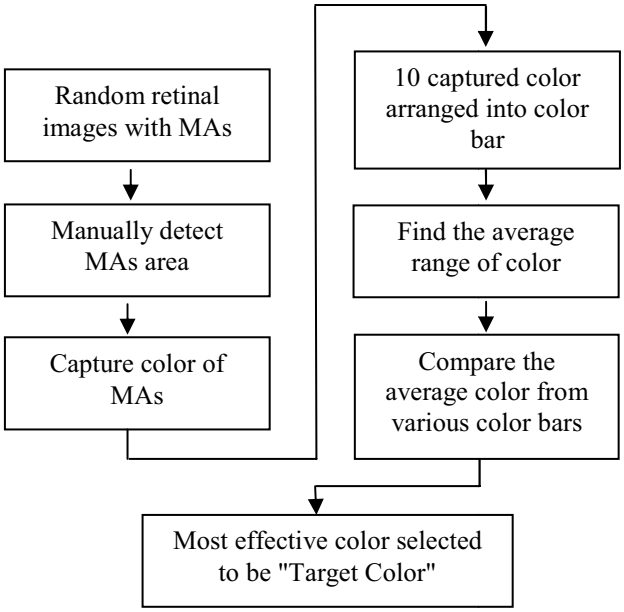


Fig. 3. Target Color Finding Diagram

The average ranges of color found from each color bar were used for MAs detection testing. A set of retinal image which represents MAs disease was brought to use in testing. The statistical results were corrected to find the average color that has most correctness. The selected color was used to be the base color for the retinal image diagnosis as shown in Fig. 3.

For MAs detection, the suspected retinal images were applied with algorithm to detect the target color in the retinal image. The dot in image which was in the range of target color was detected to define MAs position.

Fig. 4. Target Color Calculation

Fig. 4 represents a part of original retinal images which the MAs is occurred. Samples of retinal images are captured and arranged to color bar. For this picture, the analysis result shows $H_{average}$ is 0.019, $S_{average}$ is 0.682, $V_{average}$ is 0.647. This color is the most suitable color to be used for MAs detection of this image color tone. These HSV values are added to algorithm for the retinal image detection.

The HSV values used for analysis in the algorithm have to be in range of [0,255] but the results from color bar are in range of [0,1]. So the result value has to be re-scaled by using the equation (4).

$$Analysis\ value = Result\ value \times 255 \quad (4)$$

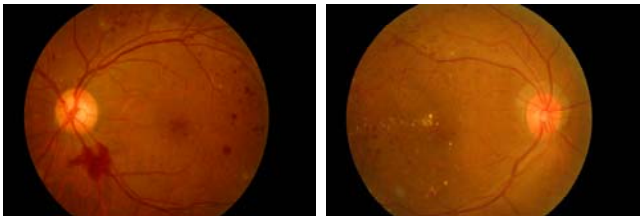


Fig. 5. Diabetic Retinopathy

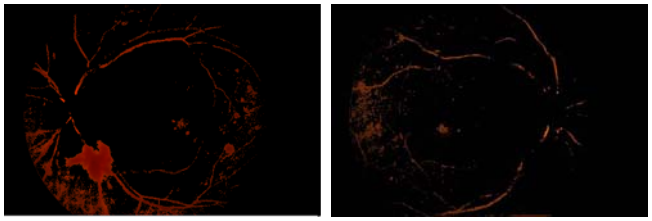


Fig. 6. Microaneurysms Detection Result

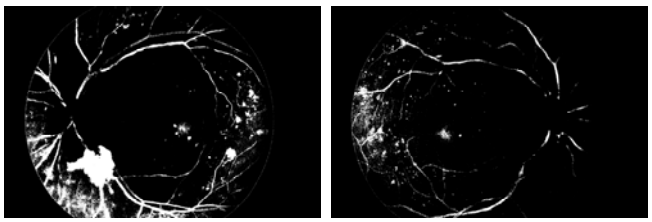


Fig. 7. Binary Image

Various types of samples are chosen randomly in order to increase the precision of MAs detection in retinal images. Fig. 5 represents two fundus images captured for diabetic retinopathy analysis. Both fundus images have abnormalities in retinal area which are both small and large red pixel areas that spreads around the images. Then the target areas are separated from theirs background as shown in Fig. 6 by defining threshold value. Threshold value is derived by identifying most suitable gray scale in experiments which in this example are target color values. Then the images in Fig. 6 are converted to binary as show in Fig. 7. Theirs pixels will be shown as 2 values which are black and white that will be easier to analyze. The black color is value '0' and white color is '1'. The binary image is used in the eccentricity method for identifying MAs that can be assumed as MAs.

C. Area and Eccentricity

Area and Eccentricity is a method to identify MAs and eccentricity border of MAs

a) Area of Microaneurysms

After the fundus image converted into HSV color model image, it will be converted again into a binary image. The pixel area of MAs will be white color ranged from 200-2200 pixels in binary image. The pixel which has its value less than 200,

it will be defined as noise. The pixel having value more than 2200 will be defined as the error from the capturing process such as the reflection from camera lens or the border color of image. In Fig. 8 shows the number of each range area pixel of white color in binary image.

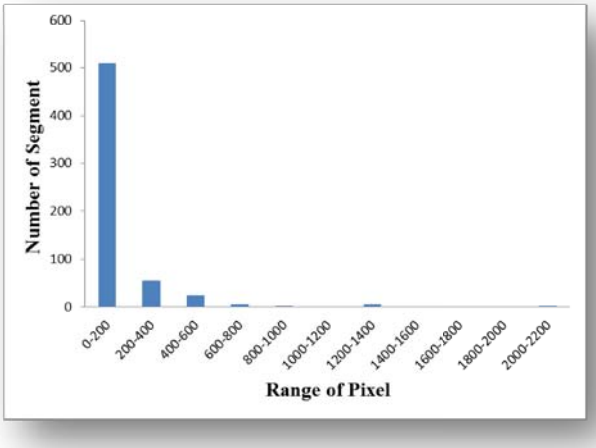


Fig. 8. Show the number of each range area pixel of white color in binary image.

b) Eccentricity of Micaroaneurysms

The shape of pixel groups can be divided by eccentricity which ranges between 0-1. A group of pixels is dot if eccentricity is 0 and if eccentricity is 1, it means that the group of pixels is line. Fig.8 shows the distribution of eccentricity in abnormality of binary image.

TABLE I. Equation for Eccentricity

E	interval
0	e = 0
$e = \sqrt{1 - \frac{b^2}{a^2}}$	0< e <1
1	e = 1

This paper uses the eccentricity that range between 0.3-0.89 and defines the eccentricity that less than 0.3 as noise. If the eccentricity is more than 0.89 then it is defined as vein. Fig. 9 shows the range of eccentricity from the group of pixels in binary image.

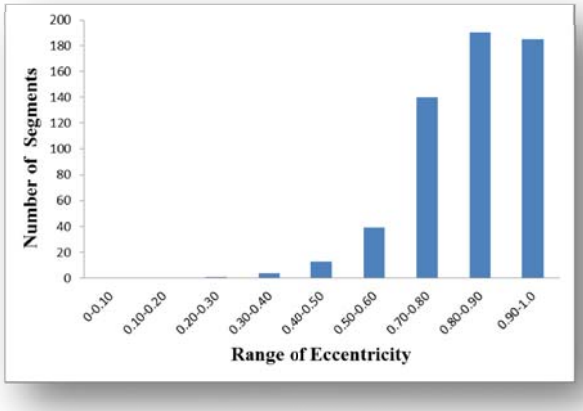


Fig. 9. Range of Eccentricity in binary image

III. RESULTS

In order to identify the MAs of diabetic retinopathy symptom, the process to diagnose expansion of blood vessel in fundus image is used. The captured fundus image will go through HSV method which will select the value of H, S and V from color bar for identifying threshold value. The target color value will be selected and then converted to binary image for easier analysis. After that the area and eccentricity methods are used for filtering the abnormalities.

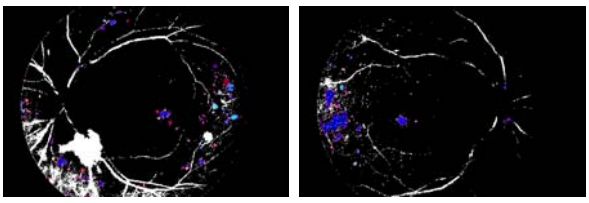


Fig. 10. Microaneurysms Symptom Area

In Fig. 10 shows the binary images which are converted from fundus images via the area and eccentricity methods for distinguishing the abnormalities area. The abnormality area that has 200-2200 pixels and eccentricity value is within 0.3-0.89 are defined as MAs identification area and it shows as color dot in the picture. The accuracy result of experiment is around 93% compared with the analysis of ophthalmologists.

IV. CONCLUSION

In this paper presents an idea to design the structure of automatic system for diagnose diabetes eyes of the patients by using the characteristic of diabetic retinopathy symptom which is MAs to find abnormalities in retinal of patients. The appropriate result of HSV is H_{averge} which is 0.019, $S_{average}$ which is 0.682 and $V_{average}$ which is 0.647. The appropriate area of MAs is 200-2000 pixels. The appropriate Eccentricity is between 0.3-0.89 and the result has an accuracy of 93% when compared to the analysis of expertise ophthalmologist which can be used to reduce diagnose time for ophthalmologist.

ACKNOWLEDGMENT

This research is supported by Coordinating Center for Thai Government Science and Technology Scholarship students (CSTS) National Science and Technology Development Agency (NSTDA).

REFERENCES

[1] A. M. Mendonça, A. J. Campilho, and J. M. Nunes, "Automatic segmentation of microaneurysms in retinal angiograms of diabetic patients," In Image Analysis and Processing, 1999. Proceedings. International Conference, pp. 728-733, 1999, IEEE.

[2] A. Sopharak, B. Uyyanonvara, S. Barman, and T.Williamson, " Automatic microaneurysm detection from non-dilated diabetic retinopathy retinal images," Lecture Notes in Engineering and Computer Science, 2191, 2011.

[3] E. M. IShahin, T. E. Taha, W. Al-Nuaimy, S. El Rabaie, O. F. Zahran, and F. E. A.El-Samie, "Automated detection of diabetic retinopathy in blurred digital fundus images," In Computer Engineering Conference (ICENCO), 2012 8th International, pp. 20-25, 2012, December, IEEE

[4] W. L. Yun, U. Rajendra Acharya, Y. V. Venkatesh, C. Chee, L. C. Min, and E. Y. K. Ng, "Identification of different stages of diabetic retinopathy using retinal optical images," Information sciences, Vol. 178, n. 1, pp. 106-121,2008.

[5] M. D. Saleh, and C. Eswaran, "An automated decision-support system for non-proliferative diabetic retinopathy disease based on MAs and HAs detection,". Computer methods and programs in biomedicine, Vol. 108, pp. 186-196,2012.

[6] http://commons.wikimedia.org/wiki/File:HSV_color_solid_cylinder.png

Automatic Mood Classification of Indonesian Tweets Using Linguistic Approach

Viktor Wijaya[1], Alva Erwin[1], Maulahikmah Galinium[1], Wahyu Muliady[2]
[1]Information and Communication Technology Faculty
Swiss German University, BSD, Indonesia
{viktor.wijaya@student. , alva.erwin@, maulahikmah.galinium@}sgu.ac.id
[2]Akon Teknologi, BSD, Indonesia
wahyu.muliady@akonprojects.org

***Abstract*—Research concerning Twitter mining becomes an interesting research topic recently. It is proven by numerous number of published paper related with this topic. This research is intended to develop a prototype system for classifying Indonesian language tweets. The prototype includes preprocessing step, main information retrieval and classification system. This research proposes a system that uses grammatical rule for retrieving main information from the tweet, and then classifies the information to the suitable mood space. The classification algorithm, which is used, is lexicon based classifier. The proposed classification system has 53.67% accuracy for classifying tweets into 12 mood spaces and 75% accuracy for classifying tweets into 4 mood spaces. As the comparison, the same dataset is also classified using SVM and Naïve Bayes.**

***Index Terms*—Text Mining, Lexical Approach, Grammatical Rule, Twitter**

I. Introduction

Twitter is a micro blogging service platform which gains its popularity since 2009. In Twitter, the relationship between users is defined as following and followed. As a micro blogging site, it restricts the users can only post 140 characters message. People usually use Twitter for sharing the actual news or their opinion about an event [1]. In Twitter, someone can "retweet" or post the same message from the other users. This makes twitter can spread news rapidly. "Retweet" also makes Twitter becomes a wide spreading news media [2]. People usually use Twitter for daily chatting, sharing information and reporting news [3]. People usually express their moods in their conversation and their post through social media.

There are many active researches about mining the data from social media. Most of them use lexicon approach for classifying social media data [4], [5], [6], [1], [7], [8]. Lexicon approach is a classification algorithm that uses list of word in dictionary to classify the data. Machine learning classification techniques rely on sufficient large data for having good classification result. Regarding to the previous fact, lexicon approach is more desirable for classifying twitter data.

The purpose of this research is developing a prototype of mood analysis system for Twitter. This research proposes lexicon scoring and main phrase detection based on POS-tag. The main phrase detection is able to improve the proposed classification accuracy. SVM and Naïve Bayes are used to compare the proposed system performance.

II. Related Works

There are some issues for involving natural language processing in text mining: stop word, stemming, noisy data, word sense disambiguation, tagging, collocations, grammar, tokenization, text representation and automated learning [9]. In order to overcome those issues, there are necessary steps to preprocess the data in order to prepare the data for information retrieval process. Qing Hu and Liu Bing proposed two sub processed that must be done in preprocessing in order to make the data is ready for the classification step [10]. The first sub process is removing links in the tweet, the username and special Twitter word, such as: RT. This sub process is intended to reduce the noise in the sentence in that tweet itself, because the information that is used is only the tweet sentence. The second sub process is correcting the spelling. In Twitter, people usually write the sentence using informal style, do some spelling mistake and abbreviate some words, so this sub process is needed to change that false word to the correct spelling and change the abbreviation word to the original form.

Liliana Ferreira et al. made a comparative study of feature extraction algorithms in customer reviews [11]. They compared the performance of two approaches in feature extraction in customer review: Likelihood Ratio Test Approach and Association Mining Approach. They conclude Association Mining algorithm is suitable to extract feature of on-topic customer review's content whereas Likelihood Ratio Test approach can be used to extract feature on mixture of on- and off- topic content. John Bollen et al. proposed a method for classifying mood in tweet to one of six tweet dimensions (tension, depression, anger, vigor, fatigue and confusion) using the extended POMS [6]. The classification result then compared with fluctuation in Stock market data, which is taken from DJIA, and also implemented it in event case study, such as: US presidential election and Thanksgiving [6]. This comparison is done for finding out the relation between collective moods in Twitter and the social-economic phenomena.

There are some related works those use Twitter data as the research data. Lansdall-Welfare et al. use Twitter data for reflecting the nation mood in UK [5]. In the other research, Landsall-Welfare et al. use Twitter data for investigating the effect of public mood in social and economic events [4]. In the opposite of that, Choudhury et al. uses social media data in exploring human emotional states [12]. A research by Roberts et al. proposed a linguistic approach for detecting and annotating tweet data into 7 categories: anger, disgust, fear, joy, love, sadness and surprise [13].

Yerva et al. implemented mood analysis on Twitter for creating travel recommendation system based on weather and location [8]. The salient summarization from this research is the usage of ANEW word list for classifying tweets into proposed 2-dimensionals mood space, which consists of 12 subspaces. Moreover, this mood space is suitable for building the proposed fusion framework for mixing the mood analysis with weather sensor data. In this research, they also extracted location from tweets based on its geo-tag metadata and the text contains location.

Finn Arup proposed a method for constructing a new word list for extracting sentiment from tweet [1]. Finn also constructed a synonym set for this word list. Since the proposed thesis topic also uses dependent language method, there is a necessity for building a synonym set for the word list.

WordNet is a lexical dictionary that has semantic relations between words: synonym, antonym, hyponym, meronym[14]. The usage and development of WordNet plays a vital role for NLP research, which is language dependent. One of those Indonesian WordNet research is done by Putra et al [15]. Putra et al proposed to use large number of human annotators for building Indonesian WordNet. However this research is still in early phase and the vocabulary in that WordNet is not complete.

This research follows research in [8] for classifying the twitter data. In advance, this research proposed an additional step for selecting the main phrase, which represents the mood of a tweet. Besides proposing main phrase detection, this research also proposed additional preprocessing steps for preparing the data: word shrinking and punctuation-word separation.

III. Data Collection and Analysis

The data for this research is limited into Indonesian tweets only. The data is crawled from sixth May 2013 until thirteenth May 2013. The data is taken from various search word and hash tag from various topics. A tweet usually contains a link, which contains no information regarding tweet's mood, within it. Therefore the link is removed from a tweet. Besides link removal, the similar tweet is also filtered by using Jaro Winkler formula with 0.8 as similarity threshold. Jaro Winkler formula is started with calculating jaro distance using formula (1) where m is the matching characters, t as the number of character transposition, $S_1$is the length of the first string, $S_2$is the length of second string. After the jaro distance is calculated, then winkler distance is calculated using formula (2) where $d_j$is jaro distance and l is length of common prefix from the start of string up to 4 characters, and p as the constant scaling factor with 0.1 as its value.

$$d_j = \begin{cases} 0 & if\, m = 0 \\ \frac{1}{3}\left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m}\right) & if\, m\,! = 0 \end{cases} \quad (1)$$

$$d_w = d_j + (lp\,(1 - d_j)) \quad (2)$$

After the data filtering, the tweet's mood is assessed using crowd source approach. Table I reflects the real number of assessed tweets based on mood spaces in Fig. 3. The assessed tweets are used as the dataset for this research.

TABLE I
NUMBER OF TWEET

Mood	Number of Tweets
nyaman (pleased)	138
senang (happy)	173
tertarik (excited)	162
terganggu (annoyed)	104
marah (angry)	79
tegang (nervous)	37
sedih (sad)	131
bosan (bored)	194
mengantuk (sleepy)	66
tenang (calm)	51
damai (peaceful)	127
rileks (relax)	126

IV. Proposed Classification System

A. Classification System Overview

This research is intended for developing the prototype of mood analyzer system that automatically determines the Indonesian tweet's mood. The classification system consists of text pre-processing process, text tokenization, POS-tagging process, main phrase detection process and lexicon based classification process. To illustrate the execution of each step in proposed classification system, a tweet is given as follow:

T = `Film kakak @radityadika luuuuucuuuuu dik. Abis nonton film :cinta ' RT@anitasaha mw nonton film @radityadika setelah dari rumah makan`

B. Text Pre-processing

Text pre-processing consists of three steps: punctuation mark removal, word shrinking and punctuation-word separation. Fig. 1 shows how text-processing works. Punctuation mark removal is a process for deleting unused punctuation mark, such as: " ",',\,:,-, ". Word Shrinking process detects and removes the consecutive same character in a word. Punctuation-word separation separates word and punctuation mark in a word becomes two or more strings. By applying this text processing for tweet T, punctuation mark " and : is removed from the tweet, word "`luuuuucuuuuu`." is shrunk into "lucu" and separates "dik." to "dik" and ".". The result that is shown by Tp:

Tp = `Film kakak @radityadika lucu dik . Abis nonton film cinta RT @anitasaha mw nonton film @radityadika setelah dari rumah makan`

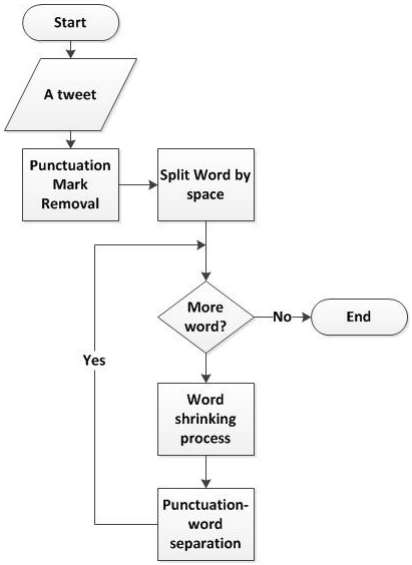


Fig. 1. Text Pre-processing

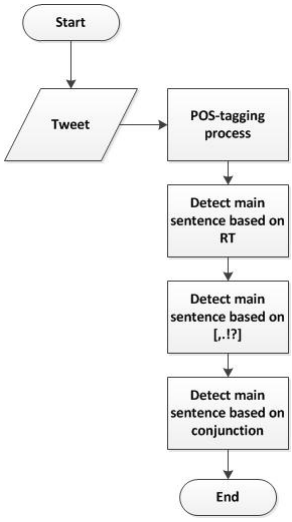


Fig. 2. Main Sentence Detection

C. Text Tokenization

Text tokenization process is a process of dividing a sentence into a group of words. This process starts by splitting a tweet based on space character. Then this process utilizes n-gram process for grouping n words in a text to forms a phrase. The n-gram process starts with n equals 1, means it selects a word and check it in WordNet. The process increases the number of n for selecting more words until the selected n words is not exist in WordNet. The iteration moves to the unselected words for the next step. This process iterates until the last word in sentence.

In this example, `rumah makan` is considered as a single phrase because the result of 2-gram process for word "rumah" and :makan" is exist in WordNet. In this The application of this tokenizing process for $T_P$is shown as follow:

T_T = {`Film , kakak , @radityadika , lucu , dik , . , Abis , nonton , film , cinta , RT , @anitasaha , mw , nonton , film , @radityadika , setelah , dari , rumah makan`}

D. POS Tagging Process

In order to be able to define the pos (part of speech) of a word in a sentence, the build-up WordNet is used. Each word in WordNet is divided into 8 category based on its pos: noun (n), adjective(a), verb(v), pronoun(pro), conjunction(co), adverb(d), preposition(p), username(u). The additional pos tags are RT, punctuation mark, and "-" for undefined word. RT and punctuation mark is kept as the pos tag because it plays a role for defining the main sentence of a tweet. The execution result of this process for T_T is shown as follow:

T_{PT} = { `n n u a n . - v n RT u - v n u kon p n .`}

E. Feature Extraction-Main Sentence Detection

For retrieving the main sentence of a tweet, there are 3 consecutive processes: main sentence retrieval based on "RT" phrase, main sentence retrieval based on punctuation mark, main sentence retrieval based on conjunction. In each step, the main sentence is retrieved. How this consecutive processes are shown in Fig. 2.

Based on Fig. 2, first T_{PT} is divided into sentences based on its "RT" character, The chosen sentence is the sentence that appears before "RT". The process continues by dividing the selected sentence into another group of sentences based on its punctuation mark. In T_{PT} , the punctuation mark is fullstop (.). The selected sentence is the one that appears before punctuation mark. Since the selected sentence does not contain any conjunction word, the application of this process forT_{PT} causes the first sentence is selected as the main sentence , which is shown below:

$$T_{MS}= \{\frac{Film}{n} , \frac{kakak}{n} , \frac{@radityadika}{u} , \frac{lucu}{a} , \frac{dik}{n}\}$$

F. Feature Extraction-Pos Tag Simplification

After the main sentence is detected, shrinking pos tag is applied. This step is used to prepare the sentence by grouping the words based on its sequential POS-tag. This process is applied with intention for keeping the sentence structure is simple enough. Applying this process to T_{MS} groups word "Film" and "kakak", as follows:

$$T_{PS}= \{\frac{Film\ kakak}{n} , \frac{@radityadika}{u} , \frac{lucu}{a} , \frac{dik}{n}\}$$

G. Feature Extraction-Main Phrase Detection

Main phrase detection is the algorithm for getting the main phrase of a main sentence, which contains the sentence's mood. It works by finding the most appropriate grammatical rule for the preprocessed tweet. In this context, a grammatical rule is a set of POS-tag, which matches text's POS-tag,

TABLE II
EXAMPLE OF MAIN PHRASE DETECTION ITERATION PROCESS OF T_{PS}

	POS-tag	Grammatic Rule	Appropriate Rule	Result	Main Phrase
1	{n,u,a,n}	{u,n,**v**,**a**}	0	{}	{}
2	{n,u,a,n}	{d,n,**v**,**a**}	0	{}	{}
3	{n,u,a,n}	{n,**v**,**a**}	1	{p_1, p_2}	{v,a}
4	{n,u,a,n}	{n,u,**a**}	3	{p_2}	{a}

with some of its members are tagged as main phrase. The main phrase for a grammatical rule is determined by human beforehand.

The iteration process is started with the grammatical rule with the most members. The iteration stops if the appropriate rule is found. If there is no appropriate rule, the system tags it as a new rule and add it to queue for human review. The appropriateness is measured by the number of matched elements between POS-tag and grammatical rule in order, starting from the first member. For example in table II, the iteration stopped at the fourth grammatical rule because it has found the appropriate rule and the main phrase is p_2 or the last adjective. Through this iteration process for T_{PS}, the main phrase a is selected. The word that are selected as the main phrase is lucu (funny) because its POS-tag is adjective, corresponds to the result of table II.

H. Lexicon based Classification System

The classification algorithm which is proposed is created based on classification system by Yerva et al [8]. This classification algorithm basically is lexicon-based classification algorithm. This classification algorithm is only implemented for Indonesian language only.

The lexicon structure consists of information: word, valence value and arousal value. Valence value is the value that determines the negative-positive value for the words whereas arousal value determines the passive-active value for the words. The lexicon is built from three resources: ANEW wordlist [1], ANEW wordlist synonyms, and dataset that is retrieved and scored manually by crowd source.

The algorithm for calculating overall valence and arousal value for a tweet is shown by formula (3). L_S is total selected words. S_V is average valence value of tweet. It is calculated by adding all selected word's valence value divided by total of selected words. $S_A$is total arousal value of a tweet. It is calculated by adding all selected word's arousal value divided by total of selected words.

$$S_A = \frac{\sum_{i=0}^{n} A_i}{L_S} \; ; \; S_V = \frac{\sum_{i=0}^{n} V_i}{L_S} \qquad (3)$$

2-dimensional mood space model is created with these following criteria: 1-9 as the range value, valence as x-axis and arousal as y-axis. The valence average value and the arousal average value forms a data point in mood space model. This 2-dimensional mood space model is created based on Walson and Telegen Model [16]. Mood space model is shown in Fig. 3.

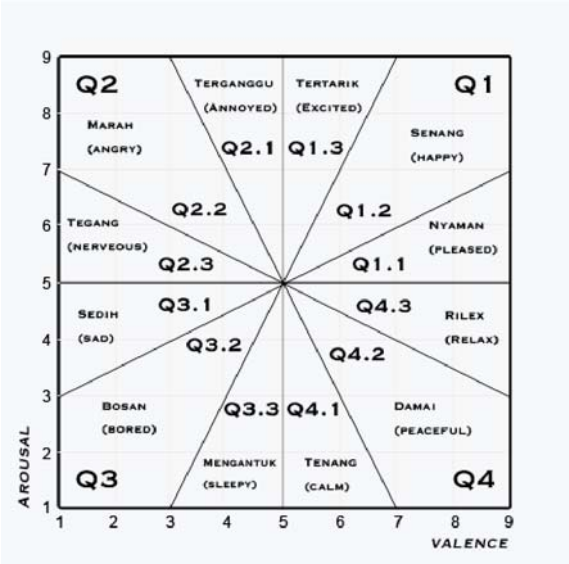


Fig. 3. 12 Mood Space Model

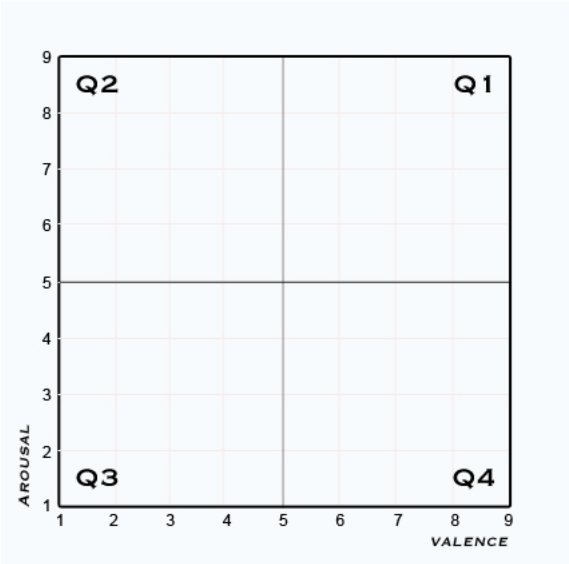


Fig. 4. 4 Mood Space Model

By using formula (3), the valence average value of T_{PS} is 7 and the arousal average value of T_{PS} is 8. Then this result forms a data point (7,8) and gets "`Senang`" mood label.

V. SUPERVISED CLASSIFICATION SYSTEM

The performance of the proposed algorithm is compared with supervised classification systems to find out which algorithm has better performance. In this research, the proposed algorithm is compared with Naïve Bayes and SVM. Naïve Bayes and SVM classification system is executed using Rapid Miner. Those two is selected because those algorithms are the most widely used for classifying text [17], [18]. The implementation of SVM and Naïve Bayes is done by using RapidMiner, supporting tools for data mining.

Implementation of SVM and Naïve Bayes is started by creating a model for preprocessing the tweets. The preprocessing consists of tokenization, lower case transformation, stop word filtering, and tokens filtering. Tokenization is process of dividing a sentence into a group of words. Lower case transformation process transforms each word into lower case. Stop word filtering uses a dictionary for filters Indonesian stopword from a sentence, whereas token filtering process removes any words which consist of less than 3 letters.

The vector for SVM and Naïve Bayes classifier is built from TF-IDF value of word. This vector is used as the attribute for determining the class of a tweet. The application of SVM and Naïve Bayes in RapidMiner is divided into 2 consecutive steps: training builder and testing model application. Training builder prepares the training set for the classifier, whereas testing model application is a process of applying the classifier for testing data.

SVM is basically two class classifier. For classifying text, SVM works by finding the optimal hyperplanes for dividing the training set and determining the class of test data based on data position against the hyperplanes [19]. In this research, the data is multiclass data. Therefore a complement algorithm for classifying multiclass data is implemented. The multiclass SVM model is built by combining binominal classifier and aggregates the result of these binominal classification models for classifying tweets. In this research, the multiclass SVM classifier follows one-against-one methods in research by Hsu et al [20] for classifying tweets into mood spaces. One-against-one method works by selecting the class with the most voting number between other classes through binominal classification.

Naïve Bayes classifier implementation determines the class of a tweet based on Bayes Theorem [17]. Naive Bayes classifier classifies a tweet into a mood space using Bayes' theorem. Naive Bayes classifier considers this entire vector contributes for determining a tweet's class independently.

VI. Result Analysis

For finding out the effect of main phrase detection process, the first experiment uses 12-class mood space model as the classifier model. 12-class mood space model is shown in Fig. 3. As shown in table III, the lexicon approach with main phrase detection leads to better accuracy than the lexicon based classifier without main phrase detection approach. Table III shows lexicon approach with main phrase detection classifies data 25.36 % more accurate than the lexicon based classifier without main phrase detection approach.

However, the performance of those approaches for classifying tweets into one of 12-class is below than 60%. The misclassification of lexicon with main phrase detection approach is caused by the close distances between moods in a quadrant, as shown in table IV. Table IV shows that Tweet number 1 and Tweet number 2 actually have same main phrase "keren", but it has different mood: tweet 1 is categorized as exciting and tweet 2 is categorized as happy based on crowd source. Since tweet 1 and tweet 2 has same main phrase, both of them is classified by system as exciting, which causes misclassification for tweet 2.

TABLE III
Standard Lexicon and Lexicon with Main Phrase Detection Approach Performance Comparison

Classification Approach	Accuracy Rate
Standard Lexicon Approach (without main phrase detection)	28.31%
Lexicon with Main Phrase Detection Approach	53.67 %

TABLE IV
Lexicon Approach Example

Number	Tweet	Main Phrase	Mood	Mood Pred
1	#ChampionsFCB keren ga min :.D RT@madritBUSUK: Lagu favorite kalian pagi ini ? #TanyaAja :D	keren	tertarik (exciting)	tertarik (exciting)
2	Pake gelas aja bisa jadi backsound lagu! keren :D	keren	senang (happy)	tertarik (exciting)

This finding is also similar with the results in [21]. Therefore second experiment, which classifies tweet into 4-mood space model, is conducted. 4-mood space model is shown by Fig. 4. The result of this second experiment shows the accuracy enhancement for both lexicon based classifier. As shown in table V, lexicon with main phrase detection approach shows a better accuracy than lexicon based classifier without main phrase detection approach with 75.14 % accuracy rate.

The performance of lexicon with main phrase detection approach is compared with SVM and Naïve Bayes for comparing the accuracy between lexicon approach and machine learning approach. With the data shown in table VI, the accuracy of lexicon with main phrase detection has better accuracy rate with 75.14% classifies data correctly compared with Naïve Bayes and SVM.

TABLE V
4-mood lexicon based accuracy comparison

Classification Approach	Accuracy Rate
Standard Lexicon Approach (without main phrase detection)	56.12%
Lexicon with Main Phrase Detection Approach	75.14%

TABLE VI
Accuracy Comparison

	Lexicon with Main Phrase Detection Approach	Naïve Bayes	Support Vector Machine
4-Mood	75.14 %	35.88 %	41.43 %
12-Mood	53.67 %	16.07%	19.52%

VII. Future Work

The accuracy for this research can be increased by improving 3 main parts in this research. Those three parts are rule for grammatical rule, Indonesian WordNet and Indonesian affective lexicon. By enriching the grammatical rule for retrieving the phrase that contains the tweet's mood, it will result a better accuracy for retrieving the main phrase for the tweet. In future works, the accuracy can be improved by enriching Indonesian WordNet with more words and more accurate pos tag for each word. More than that, an approach for enriching Indonesian WordNet and Indonesian affective lexicon with slang from Twitter is necessarily important for increasing the classification accuracy.

In this research the use of grammatical rule for defining the main phrase of a tweet has a main role for increasing the classification system accuracy. The process for defining grammatical rule in this research is done manually. In future work, there is necessity for having an algorithm that can generate grammatical rule automatically. This process can be used for enriching and standardizing the usage of grammatical rule for NLP research area in Indonesia.

The other process that can be done in future work is an automatic POS-tag identification process for slang word or slang abbreviation. This process is expected to be able indentify slang and abbreviation's POS-tag in a sentence. This process is also expected to increase the performance of main phrase detection process.

VIII. Conclusion

Based on the results, the proposed lexicon with main phrase detection approach is proven to be more accurate than mood classification using lexicon without main phrase detection approach, SVM and Naïve Bayes. Due to the proximity between mood spaces in same quadrant, misclassification happens between those mood spaces. In second experiment, the tweet is classified into 4 mood space and generates a better accuracy.

The proposed approach utilizes Indonesian WordNet, lexicon classification and Indonesian grammatical rule for building semantic classification approach. The contributions of this research are:

- Affective lexicon for Indonesian words, which contains valence and arousal value
- The usage of main phrase detection process for complementing lexicon approach.

Acknowledgment

The author would like to thank Swiss German University for its support. The authors are also grateful to Akon Teknologi, which helps this research by providing resource for the research crowd sourcing. This research is devoted to Indonesian researchers, which are involved in Indonesian NLP researchers.

References

[1] Finn Årup Nielsen. A new anew: Evaluation of a word list for sentiment analysis in microblogs. *CoRR*, abs/1103.2903, 2011.

[2] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 591–600, New York, NY, USA, 2010. ACM.

[3] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, WebKDD/SNA-KDD '07, pages 56–65, New York, NY, USA, 2007. ACM.

[4] Thomas Lansdall-Welfare, Vasileios Lampos, and Nello Cristianini. Effects of the recession on public mood in the uk. In *Proceedings of the 21st international conference companion on World Wide Web (WWW '12)*, pages 1221–1226, Lyon, France, 03/2012 2012. ACM, ACM.

[5] Thomas Lansdall-Welfare, Vasileios Lampos, and Nello Cristianini. Nowcasting the mood of the nation. *Significance*, 9:26–28, 2012.

[6] Johan Bollen, Alberto Pepe, and Huina Mao. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *CoRR*, abs/0911.1583, 2009.

[7] Johan Bollen, Huina Mao, and Xiao-Jun Zeng. Twitter mood predicts the stock market. *CoRR*, abs/1010.3003, 2010.

[8] Surender Reddy Yerva, Ho Young Jeung, and Karl Aberer. Cloud based Social and Sensor Data Fusion. In *FUSION*, 2012.

[9] Anna Stavrianou, Periklis Andritsos, and Nicolas Nicoloyannis. Overview and semantic issues of text mining. *SIGMOD Rec.*, 36(3):23–34, September 2007.

[10] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '04, pages 168–177, New York, NY, USA, 2004. ACM.

[11] Liliana Ferreira, Niklas Jakob, and Iryna Gurevych. A comparative study of feature extraction algorithms in customer reviews. In *Proceedings of the 2008 IEEE International Conference on Semantic Computing*, ICSC '08, pages 144–151, Washington, DC, USA, 2008. IEEE Computer Society.

[12] Munmun De Choudhury, Scott Counts, and Michael Gamon. Not all moods are created equal! exploring human emotional states in social media. In *ICWSM*, 2012.

[13] Kirk Roberts, Michael A. Roach, Joseph Johnson, Josh Guthrie, and Sanda M. Harabagiu. Empatweet: Annotating and detecting emotions on twitter. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Ugur Dogan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may 2012. European Language Resources Association (ELRA).

[14] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. Introduction to WordNet: an on-line lexical database. *International Journal of Lexicography*, 3(4):235–244, 1990.

[15] Desmond Darma Putra, Abdul Arfan, and Ruli Manurung. Building an indonesian wordnet.

[16] J. R. Shack G. J. Meyer. Structural convergence of mood and personality: Evidence for old and new directions. *Journal of Personality and Social Psychology*, 57:691–706, 1989.

[17] Vishal Gupta and Gurpreet Lehal. A survey of text mining techniques and applications. *Journal of Emerging Technologies in Web Intelligence*, 1(1), 2009.

[18] R. G. Ramani P. Nancy. A comparison on performance of data mining algorithms in classification of social network data. *International Journal of Computer Applications*, 32, 2011.

[19] Susan Dumais et al. Using svms for text categorization. *IEEE Intelligent Systems*, 13(4):21–23, 1998.

[20] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on*, 13(2):415–425, 2002.

[21] Munmun De Choudhury, Michael Gamon, and Scott Counts. Happy, nervous or surprised? classification of human affective states in social media. In John G. Breslin, Nicole B. Ellison, James G. Shanahan, and Zeynep Tufekci, editors, *ICWSM*. The AAAI Press, 2012.

Automatic Text Summarization Based on Semantic Analysis Approach for Documents in Indonesian Language

Pandu Prakoso Tardan[1], Alva Erwin[1], Kho I Eng[1], Wahyu Muliady[2]
[1]Information Technology Faculty
Swiss German University, BSD, Tangerang, Indonesia
[2]Akon Teknologi
BSD, Tangerang, Indonesia
[1]{pandu.tardan@student. , alva.erwin@, ie.kho@} sgu.ac.id
[2]wahyu.muliady@akonprojects.org

***Abstract*—Research about text summarization has been quite an interesting topic over the years, proven by numerous number of papers related with discussion of their studies such as approaches, challenges and trends. This paper's goal is to define a measurement for text summarization using Semantic Analysis Approach for Documents in Indonesian language. The applied measurement requires Indonesian version of WordNet which had been implemented roughly. The main idea of semantic analysis is to obtain the similarity between sentences by calculating the vector values of each sentence with the title. The need of WordNet is to define the depth of each word as being computed for word similarity. Combining all required formulas and calculations, a compact and precise summarization is produced without depriving the gist information of certain document.**

***Index Terms*—Text summarization, semantic analysis, sentence similarity, word similarity, Indonesian document**

I. Introduction

A quote from Oxford English Dictionary defined automatic text summarization as "The creation of a shortened version of a text by a computer program" [1]. The summary should represents the gist informations of a text. Instead of conducting manual self-summarization, automatic text summarization should take over that role. For that purpose, numerous number of researches had been done.

There are two classifications of summarization method; extractive and abstractive summarization. The extractive summarization performs the extraction of important and relevant sentences or other features in a more dense form. On the other hand, abstractive summarization tends to reveal the meaning of a text and paraphrase the summarization. Therefore, this proposed measurement is classified as extractive summarization as its result is a set of relevant sentences which are not paraphrased. For extractive summarization, the important features [2] are content word, title word, sentence location, sentence length, proper noun, etc. Regarding to the current facts, applying semantic analysis approach in text summarization may achieve certain level of accuracy.

The goal of this research is developing a measurement for automatic text summarization using semantic analysis approach for Indonesian language-based document. The problems defined in this research are limited researches of text summarization development for Indonesian documents and achieving an acceptable quality for the generated summary. Therefore, semantic analysis is chosen as the approach with high expectations to achieve the best summary result quality.

This paper contains six more chapters as follows respectively: Section 2 explains related works, significant features are presented in Section 3, word similarity measurement is explained in Section 4, Section 5 presents sentence similarity and ranking, the results and measurements comparison will be given in Section 6, Section 7 will gives the result and Section 8 explains the conclussion and future works.

II. Related Works

There are numerous researches in text mining area; specialized in the automatic text summarization. Years ago, pair of researchers [1] listed the trends of Automatic Text Summarization over the years such as statistical approach [3], [4], [5], [6], natural language processing (NLP) , semantic analysis approach [7], fuzzy logic [8] and swarm intelligence [9]. By exploring and analyzing the strengthss and limitations of previous methods, another measurement could be discovered as the purpose is to outrank the previous methods.

Inside the category of semantic analysis approach, the existence of ontology knowledge tends to be critical proven by related researches [10], [7]. The common ontology knowledge is WordNet [11], [12] as a lexical ontology word resource. Therefore, our research used WordNet for Indonesian language [13], [14] as the lexical word resources.

In extractive text summarization, sentence weighting plays an important role. Achieving the sentence weighting requires sentence similarity measurements. Through the list of proposed sentence similarity methods [15], [16], [17], the proposed method from Li and McLean should be the best applied method for the limitation of this research.

Despite of the various researches in this field, the publications of Indonesian text summarization [18], [19] are

proven to be infrequent. The only highly notable research of Indonesian text summarization was conducted using genetic algorithm [18]. Therefore, chances to research in Indonesian text summarization field are still wide-open.

III. System Overview

Figure 1 shows the main phases of generating summary based on this proposed measurement such as *Pre-Processing*, *Feature Computation*, *Feature Ranking* and *Generating Summary*. The document is pre-processed for eliminating the noises and wastes which exist in the original document so that the result from pre-processing phase can be processed for further phases. This phase consist of *Sentence Extraction*, *Tokenization*, *Stop Words Removal*, *N-Gram Detection* and *Stemming*. The further details will be described in Section 4.

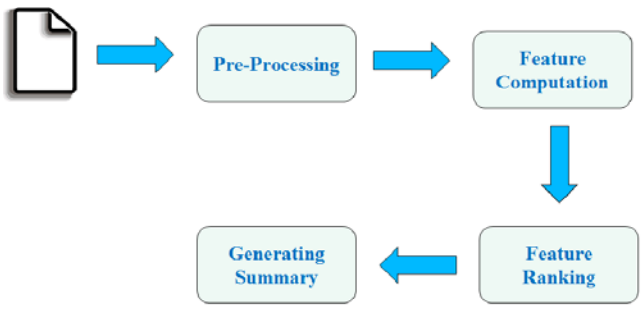


Figure 1. The Proposed Measurement Architecture

The purpose of *Feature Computation* phase is to extract and calculate the sentence similarity value between each sentence and the title. The basic idea of using that computation is to rank sentences based on the relevancy with the whole document which the main meaning is implied in the title. This phase consist of two sub-measurement; *Word Similarity Measurement* and *Sentence Similarity Measurement*. The details of those sub-measurement will be described in Section 5 and 6.

In the *Feature Ranking* phase, the sentence similarity values will be ranked ascendingly; from the highest to the lowest value. Afterwards, sentences that are considerably irrelevant should be eliminated since the general summary just extracts the gist information in the document.

For altering the summary result to become more readable, another process should be conducted. In the *Generating Summary* phase, the ranked sentences are sorted based on the position in the document. This phase is required to generate a readable summary that has a good flow of information. In conclusion, the result of the overall phases in this proposed measurement generates the relevant sentences in the readable summary format.

IV. Text Summarization Pre-Processing Features

Regarding the scope, topic and proposed method of this research, the proposed text summarization pre-processing features are defined as follows: WordNet, N-Gram, Stemming and Stop Words.

A. WordNet

WordNet is a lexical word database with features of mapping similar words into *synonym sets* or *synsets* to denote semantic relationship between sets. [12] Over the years, Princeton University had managed to develop WordNet for English language containing large sets of words and classifications such as noun, verbs, adjective and adverb. In 2008, researchers from University of Indonesia constructed Indonesian version of WordNet [13] then another research for building WordNet with Monolingual Lexical Resources were conducted too in 2010 [14]. Proven the implementations of WordNet for Indonesian are exist and published, the access for sensitive resources of Indonesian WordNet is somehow confidential. Therefore, a raw self-implementation of Indonesian Wordnet is required.

B. N-gram

N-gram is an adjoining array of *n* items in a text or speech. A n-gram could form a word or phrase. The range of n in n-gram is from 1 referred as unigram, 2 referred as bigram and 3 referred as trigram. Researchers named Brown and deSouza [20] discussed a classification of certain n-gram from the stream of text. They extract the valid n-gram based on the frequency of words or phrases in the document plus their relations with their adjacent words. However, in this paper's scope, a n-gram is considered valid if the respective n-gram is found in the WordNet.

C. Stemming

Stemming is a process of compressing affixed words using certain rules into its root form or so called stem. In this paper, stemming algorithm is used in word similarity measurement process for normalizing affixed words which have similar stem in order to calculate their base form similaritynot based on the actual form. Many approaches of Indonesian stemming have been proposed [21], [22]. A survey paper for Indonesian stemming composed the statistics of performance and stemming correction percentage comparison. Its result stated that stemming algorithm proposed by Nazief and Adriani (1996) were considered for being the most efficient since its correction percentage is 93%. Therefore, this paper used Nazief and Adriani [22] proposed algorithm for stemming.

D. Stop Words

Stop words is list of over-commonly words in corpora which are filtered for being not relevant in natural language data or text. The inclusion of stop words in text could bias the tangible value of sentences or corpus. Results of both ignoring and including stop words in pre-processing phases proved significantly different which ignorance of stop words give better subjectivity measurement result.

V. Word Similarity Measurement

Semantic similarity is an idea of assigning metric based values of documents or terms based on meaning or semantic similarity. To perceive the semantic value of documents or

terms, word similarity measurement process should be adopted first. There are two main features for word similarity such as depth of words and Wu & Palmer measurement.

A. Depth of Words

To obtain semantic similarity of two words using synthetical WordNet, defining depth of those words is prioritized before further calculation. Depth of words are acquired by traversing the shortest distance paths from one word to another word using *synonym sets* or *synsets* graph. Essentially, the more paths from one word to another word are traversed, the more depth's value are given. The calculation is defined in formula (1):

$$depth(w_1, w_2) = min(paths(w_1, w_2)) \quad (1)$$

where $w_1$and w_2 are the words, respectively; $paths(w_1, w_2)$ is collection of paths value from $w_1$and w_2; $min(paths(w_1, w_2))$ is the minimum value of $paths(w_1, w_2)$. For example, the paths from word *makan* and *minum* are four (4), three (3) and six (6). Since the depth of words extracts the shortest distance path between two words, the value of $depth(makan, minum)$ is three.

B. Wu & Palmer Measurement

Wu & Palmer calculates the similarity measurement of two concepts by enumerating the depths of those concepts in WordNet taxonomy. The measurement includes depth of Least Common Subsumer (LCS) and the depths of the respective words. The formula is defined as follows in formula (2):

$$sim_{wp}(w_1, w_2) = \frac{2 * depth(lcs(w_1, w_2)}{depth(w_1, w_2) + depth(w_2, w_1)} \quad (2)$$

Least Common Subsumer (LCS) is a shortest distance of two concept compared in lexical taxonomy. For example, *organisme* and *plantae* are the subsumers of *oryza sativa* and *zea mays L.*, but *plantae* is the less common subsumer than *organisme* for them. In this paper, we use our raw self-implementation of Indonesian WordNet which is considered as simple and incomplete. Thus, since the taxonomy relationship of our WordNet is still raw, we consider the depth of LCS is one as the depth of a root in taxonomy.

VI. Sentence Similarity Measurement and Ranking

Most measurements can be categorized into two groups: edge counting-based (or dictionary/thesaurus-based) and information theory-based (or corpus-based) [17]. After investigating those measurements, we propose the base of our sentence similarity measurement should be categorized in dictionary-based methods. Therefore, an appropriate method proposed by Li and McLean [17] should be the best reference for sentence similarity measurement since their research based on semantic approach. However, some modifications of their methods are necessary due to differences in required features.

Title : Prediksi Squad Liverpool Kontra Indonesia XI

Sentence	Sentence Similarity Value
Liverpool akan melawan Indonesia XI dalam laga persahabatan di Stadion Gelora Bung Karno	0.7364389408902138
Pelatih "The Reds" Brendan Rodgers kemungkinan besar akan melakukan banyak eksperimen pada pertandingan nanti	0.863982746507212
Bleacherreport memiliki prediksi skuad Liverpool yang akan diturunkan	0.9057639524245196
Rodgers sepertinya akan menggunakan formasi 4-2-2-1-1 untuk membungkam Boaz Solossa dan kawan-kawan	0.8022543177694139

Figure 2. Example Result of Sentence Similarity Measurement using Cosine Similairity method

A. Semantic Vector Derivatization

To illustrate the details of sentence similarity computation, we provide two sentences as follow:

- s_1= Spesies lemur tikus baru ditemukan
- s_2= Lemur adalah anggota keluarga primata seperti halnya monyet

From those sets, tokenization and stop words disposal process should be accomplished for both sets. To equate the dimension of both semantic vectors for further calculation, their dimension's lengths must follow the formula (3):

$$length(sv_{1,2}) = length(s_1) + length(s_2) \quad (3)$$

where $length(s_1)$ and $length(s)$ is the amount of words in $s_1$and s_2 respectively without any stop words included. Using that formula, we can form a matrix of word similarity values as illustrated in Table I which each cell consists of the similarity result of words from respective row and column. This, semantic vector values of $sv_{1,2}$is formed by extracting the highest values in each column of matrix below. The semantic vector of $sv_{1,2}$is:

$$sv_{1,2} = \{1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0.33 \quad 0.5 \quad 1 \quad 0.22 \quad 0.13\}$$

The same way applied for deriving $sv_{2,1}$ as below is the result:

$$sv_{2,1} = \{1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0.5 \quad 1 \quad 0.13 \quad 0.2\}$$

B. Semantic Vector Computation

After semantic vectors from two sentences have been formed, the next step is computing the similarity of those vectors using various possible methods such as Euclidean distance and Cosine similarity. Regardless of the current context, there should be an explanation of Jaccard similarity formula even though it is not included in semantic analysis approach. Figure 2 shows an example of semantic similarity measurement computation results for an article.

1) Euclidean Distance: The Euclidean distance is a measurement of distance that one would measure with a ruler. Basically, the distance between two points is the length of the line segment connecting them by using Euclidean distance. The formula of general Euclidean distance between two points (p,q) is given as follows:

Table I
SEMANTIC VECTOR DERIVATIZATION PROCESS

	Spesies	lemur	tikus	ditemukan	Lemur	anggota	keluarga	primata	halnya	monyet
Spesies	1	0.13	0.09	0.18	0.13	0.33	0.5	0.13	0.22	0.1
lemur	0.13	1	0.13	0.13	1	0.13	0.13	1	0.13	0.13
tikus	0.09	0.13	1	0.09	0.13	0.09	0.09	0.13	0.09	0.07
ditemukan	0.18	0.13	0.09	1	0.13	0.18	0.2	0.13	0.18	0.09
	↓	↓		↓	↓	↓	↓	↓	↓	↓
$sv_{1,2}$	1	1	1	1	1	0.33	0.5	1	0.22	0.13

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \tag{4}$$

where the result is the distance between two points or *vectors*. The conversion formula from distance to similarity is given as follows:

$$sim(p,q) = \frac{1}{1 + d(p,q)} \tag{5}$$

which the constant of one as denominator is used for eliminating the possible values of zero. The similarity value of s_1 and s_2 is 0.3721760683269597 using this euclidean similarity.

2) Cosine Similarity: Unlike euclidean distance, cosine similarity is a similarity measurement between two vectors by considering the cosine angle between them. The maximum value of this similarity is 1 since cosine value of the lowest angle which is 0^o is 1. The formula is given as follows:

$$sim(p,q) = cos(\theta) = \frac{p \cdot q}{\| p \| \cdot \| q \|} \tag{6}$$

By using this formula, the similarity value of $s_1$and s_2 is 0.7261975526051178.

3) Jaccard Similarity: Jaccard similarity is a similarity measurement for evaluating the diversity of sets. In general, this similarity computes the distance between two sets by counting both match and mismatch elements. This computation does not require WordNet as lexical word database since this method can be classified in statistical approach; not semantic analysis. The formula is given as follows:

$$J(p,q) = \frac{| p \cap q |}{| p \cup q |} \tag{7}$$

where can be derived into:

$$J(p,q) = \frac{M_{11}}{M_{10} + M_{01} + M_{11}} \tag{8}$$

The jaccard similarity value between $s_1$and s_2 is 0.2.

C. Sentence Position

The gist informations of a document is also determined by the position of sentence in documents. A good writing document should place the main idea in these particular location:

1) At the beginning of the paragraph (Deductive Paragraph)
2) In the middle of the paragraph

Title : Prediksi Squad Liverpool Kontra Indonesia XI

Sentence	Sentence Similarity Value
Bleacherreport memiliki prediksi skuad Liverpool yang akan diturunkan	0.9057639524245196
Pelatih "The Reds" Brendan Rodgers kemungkinan besar akan melakukan banyak eksperimen pada pertandingan nanti	0.863982746507212
Rodgers sepertinya akan menggunakan formasi 4-2-2-1-1 untuk membungkam Boaz Solossa dan kawan-kawan	0.8022543177694139
Liverpool akan melawan Indonesia XI dalam laga persahabatan di Stadion Gelora Bung Karno	0.7364389408902138

Figure 3. Example Result of Sentence Ranking

3) At the end of the paragraph (Inductive Paragraph)

Therefore, considering that deductive paragraph is the most common type among other two, inclusion of the first sentence in every paragraph or article into text summarization should be done. Thus, those main sentences are also used for helping the result to become more readable format.

D. Sentence Ranking

The basic idea of text summarization is extracting the relevant informations of a text or document. The extracted informations or sentences should be ranked based on their relevances with the essence of the text. To obtain the sentence's value of relevance, sentence similarity between each sentences in the content and title is calculated. From that list of similarities, we eliminate sentences that are considered irrelevant by calculating first the mean of them using this equation:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} a_i \tag{9}$$

where n is the size of the sentence list and a_i is the value in i^{th} sentence. Afterwards, sentences with the values below mean are considered irrelevant with the text and shouldn't be included in summarization. The relevant sentences are then sorted based on the position in text. Theoretically, these processes would produce text summarization with a human-friendly readable format. Figure 3 illustrates an example of sentence ranking result based on the sentence similarity value, whereas Figure 4 shows the list of sentences that are considered as relevant with the document.

VII. RESULT

Tests were performed by implementing text summarization based on the proposed measurements above and comparing each of the results. The comparison of result will be based

Title : Prediksi Squad Liverpool Kontra Indonesia XI

Sentence	Sentence Similarity Value
Bleacherreport memiliki prediksi skuad Liverpool yang akan diturunkan	0.9057639524245196
Pelatih "The Reds" Brendan Rodgers kemungkinan besar akan melakukan banyak eksperimen pada pertandingan nanti	0.863982746507212

Figure 4. Example Result of Sentence Ranking after eliminating Irrelevant Sentences

MK: DPD Berwenang Bahas UU Terkait Daerah

Mahkamah Konstitusi memutuskan bahwa Dewan Perwakilan Daerah (DPD) berwenang untuk ikut serta mengajukan dan membahas Rancangan Undang-Undang yang terkait daerah. Dalam putusan yang dibacakan Rabu (27/3/2013), MK mengabulkan sebagian permohonan uji materi atas UU 27/2009 dan UU 12/2011. "Penyusunan Program Legislasi Nasional dilaksanakan oleh DPR, DPD, dan Pemerintah," ungkap Ketua Majelis Hakim Konstitusi sekaligus Ketua MK, Mahfud MD, dalam putusannya. Menurut MK, DPD juga memiliki hak menyusun program legislasi nasional (Prolegnas) sebab kedudukan DPD setara dengan Presiden dan DPR. UU 27/2009 adalah tentang Majelis Permusyawaratan Rakyat, Dewan Perwakilan Rakyat, Dewan Perwakilan Daerah, dan Dewan Perwakilan Rakyat Daerah. Sedangkan UU 12/2011 merupakan UU tentang Pembentukan Peraturan Perundang-Undangan. Uji materi atas kedua UU tersebut diajukan oleh Ketua DPD Irman Gusman, Wakil Ketua DPD Laode Ida, dan Wakil Ketua DPD Gusti Kanjeng Ratu Hemas. Hakim konstitusi Akil Mochtar, saat membacakan pertimbangannya, menjelaskan DPD bisa mengajukan RUU dan tidak boleh dibedakan dengan wewenang presiden dan DPR. "Namun demikian, DPD hanya memiliki wewenang mengajukan RUU terkait daerah, yang mencakup otonomi, perimbangan keuangan antara pusat dan daerah, hubungan pemerintah pusat dan daerah, pembentukan dan pemekaran serta penggabungan daerah, serta pengelolaan sumber daya alam," ucap Akil. Menanggapi putusan ini, Ketua DPD Irman Gusman mengaku gembira dengan putusan MK yang revolusioner. "Ini hari bersejarah, sehingga pelaksanaan tupoksi DPD mendapat tempat sebagaimana seharusnya," kata Irman. Kuasa Hukum Pemohon, Todung Mulya Lubis, mengatakan bahwa putusan MK meluruskan kembali pasal 22 D UUD 1945, setelah memberikan hak DPD ikut mengusulkan, dan merancang UU. "MK memberikan hak kepada DPD, bersama DPR dan Presiden membahas Prolegnas (Program Legislasi Nasional) meskipun DPD tidak ikut dalam persetujuan," katanya.

Figure 5. Original document entitled "DPD Sambut Baik Putusan MK Soal Kewenangan Legislasi"

on compression rate (CR), subjectivity measure (SM) of the accuracy and processing time (PT). However, some examples of text summarization result should be shown in advance as the proofs of the proposed measurement. Figure 1 is the original document and the results of its summarization is given in Figure 2 and 3 where Figure 2 is the summarization result using cosine similarity and figure 3 is the summarization result using euclidean similarity.

Thus, table II shows the comparison result of text summarization experiments using four different techniques. These experiments used seventy (70) articles that have twelve (12) sentences in average as data sets and were run with this following specifications as the environment of this research implementation:

- Processor Intel(R) Core (TM) i5-2430M CPU @ 2.40GHz
- Memory 4096Mbyte
- Hard Disk Drive 700Gbyte
- Operating System Windows 7 Home Basic 64-bit
- Database MySQL version 5.5.27 - MySQL Community Server (GPL)

MK: DPD Berwenang Bahas UU Terkait Daerah

Mahkamah Konstitusi memutuskan bahwa Dewan Perwakilan Daerah (DPD) berwenang untuk ikut serta mengajukan dan membahas Rancangan Undang-Undang yang terkait daerah.

Dalam putusan yang dibacakan Rabu (27/3/2013), MK mengabulkan sebagian permohonan uji materi atas UU 27/2009 dan UU 12/2011.

Menurut MK, DPD juga memiliki hak menyusun program legislasi nasional (Prolegnas) sebab kedudukan DPD setara dengan Presiden dan DPR.

UU 27/2009 adalah tentang Majelis Permusyawaratan Rakyat, Dewan Perwakilan Rakyat, Dewan Perwakilan Daerah, dan Dewan Perwakilan Rakyat Daerah.

Sedangkan UU 12/2011 merupakan UU tentang Pembentukan Peraturan Perundang-Undangan.

Uji materi atas kedua UU tersebut diajukan oleh Ketua DPD Irman Gusman, Wakil Ketua DPD Laode Ida, dan Wakil Ketua DPD Gusti Kanjeng Ratu Hemas.

Menanggapi putusan ini, Ketua DPD Irman Gusman mengaku gembira dengan putusan MK yang revolusioner.

"Ini hari bersejarah, sehingga pelaksanaan tupoksi DPD mendapat tempat sebagaimana seharusnya," kata Irman.

Figure 6. Summarization Result of Figure 1 using Cosine Similarity

MK: DPD Berwenang Bahas UU Terkait Daerah

Mahkamah Konstitusi memutuskan bahwa Dewan Perwakilan Daerah (DPD) berwenang untuk ikut serta mengajukan dan membahas Rancangan Undang-Undang yang terkait daerah.

Dalam putusan yang dibacakan Rabu (2732013), MK mengabulkan sebagian permohonan uji materi atas UU 272009 dan UU 122011.

Menurut MK, DPD juga memiliki hak menyusun program legislasi nasional (Prolegnas) sebab kedudukan DPD setara dengan Presiden dan DPR.

Sedangkan UU 122011 merupakan UU tentang Pembentukan Peraturan Perundang-Undangan.

Menanggapi putusan ini, Ketua DPD Irman Gusman mengaku gembira dengan putusan MK yang revolusioner.

Ini hari bersejarah, sehingga pelaksanaan tupoksi DPD mendapat tempat sebagaimana seharusnya, kata Irman.

Figure 7. Summarization Result of Figure 1 using Euclidean Similarity

- Programming Language Java(TM) SE Environment version 1.7.0_11

Table II
RESULT OF MEASUREMENTS COMPARISON

Text Summarization Techniques	SM	CR	PT
Statistical Approach w/ Word Frequency	75.96%	76.53%	00m01s
Semantic Analysis w/ Euclidean	85.00%	53.27%	03m54s
Semantic Analysis w/ Cosine	83.46%	46.66%	03m42s
Statistical Approach w/ Jaccard	85.18%	47.93%	00m03s

This research used human-based evaluation for measuring the subjectivity measurement (SM) since it is considered the best procedure for assessing text summarization. In order to evaluate based on SM, the evaluators were provided seventy data sets for each text summarization technique explained in table II. Thus, the evaluators assessed the summary given by each technique or method based on the relevancy between each sentence in summary, the relevancy between each sentence in summary and title, the clarity of whole summary and the

existence of irrelevant sentence in summary. Based on those factors, the evaluators should give scores for each article summary produced by each technique with the range of one (1) as the lowest score and ten (10) as the highest. By using that way, the subjectivity measurement value for each technique could be achieved with the accurate assessment.

Another considerable factors for evaluating this measurement are Compression Rate (CR) and Processing Time (PT). The compression rate value was calculated based on the ratio between the amount of the sentences in the original document and the amount of the sentences in the generated summary. On the other hand, processing time is the value of time required for processing each technique defined in Table II.

Based on the result shown in Table II, semantic analysis approaches have the better average accuracy at 84.23% than statistical approaches at 80.57% in spite of the considerably long processing time. However, statistical approach has the better compression rate average at 62.23% than semantic analysis at 49.97%.

It is noted that the implemented statistical approach techniques for this research include the sentence ranking and sentence position features. Thus, it explains the low magnitude of average result between statistical approach and semantic analysis, plus the good result from the subjectivity measure.

VIII. Conclusion and Future Works

As defined in Section 7, statistical approach has the better overall result in the previously stated environment specification and certain data sets. However, semantic analysis may overcome the result of statistical approach in different environment. Based on the concept, semantic analysis should generate a better summary quality than statistical approach. For proofing the concept even further, a qualify Indonesian WordNet should be implemented.

Text summarization measurement of matching the title with every sentence in the content plus inclusion of first sentence in paragraph or text is proved to be effective. Due to the variant of sentences meaning in the content, semantic analysis can be used for finding the most similar meaning of each sentences with title. Therefore, those extracted sentences are the summarize outputs of the respective documents. Unfortunately, since the proposed measurement requires title as the main feature, certain documents with short title may not give the best summarization. In conclusion, the best testing set for this measurement is articles.

One of the unimplemented features in this research is eliminating the junk sentences. Junk sentences are sentences that doesn't have relevancy with the document's meaning. Theoriticaly, discarding the junk sentences should increase the compression rate (CR) and subjectivity measurement (SM). The quote fragments in the summarization are also decreasing the SM. The inclusion of those quote fragments may reduce the subjectivity measure due to their lack of relevancy between sentences in the readable summarization. There is a suggestion for replacing those sentences with quotes into new sentences by paraphrasing them. Thus, sentences with quotes will be eliminated and the possible result of summarization should have higher subjectivity measurement value.

Acknowledgement

This research presented in this paper is supported by Information Technology Department of Swiss German University and Akon Teknologi company.

References

[1] Oi Foong, Alan Oxley, and Suziah Sulaiman. Challenges and trends of automatic text summarization. *International Journal of Information and Telecommunication Technology (ISSN: 0976-5972)*, 1(1), 2010.

[2] Vishal Gupta and Gurpreet Lehal. A survey of text summarization extractive techniques. *Journal of Emerging Technologies in Web Intelligence*, 2(3), 2010.

[3] H. P. Luhn. The automatic creation of literature abstracts. *IBM Journal of Research Development*, 2(2):159–165, 1958.

[4] P.B. Baxendale. Machine-made index for technical literatureŮan experiment. *IBM Journal of Research Development*, 2(4):354–361, 1958.

[5] Erik Wiener, Jan O. Pedersen, and Andreas S. Weigend. A neural network approach to topic spotting, 1995.

[6] James Hendler. Agents and the semantic web. *IEEE INTELLIGENT SYSTEMS*, 16(2):30–37, 2001.

[7] Rakesh Verma, Ping Chen, and Wei Lu. A semantic free-text summarization system using ontology knowledge. *Document Understanding Conference DUC 2007*, pages 1–5, 2007.

[8] Farshad Kyoomarsi, Hamid Khosravi, Esfandiar Eslamiand Pooya Khosravyan Dehkordy, and Asghar Tajoddin. Optimizing text summarization based on fuzzy logic. *ICIS '08 Proceedings of the Seventh IEEE/ACIS International Conference on Computer and Information Science*, pages 347 – 352, 2008.

[9] Mohammed Salem Binwahlan, Naomie Salim, and Ladda Suanmali. Fuzzy swarm based text summarization 1, 2009.

[10] Mohsen Pourvali and Mohammad Saniee Abadeh. Automated text summarization base on lexicales chain and graph using of wordnet and wikipedia knowledge base. *International Journal of Computer Science Issues (IJCSI)*, 9:343, 2012.

[11] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. Introduction to WordNet: an on-line lexical database. *International Journal of Lexicography*, 3(4):235–244, 1990.

[12] Christiane Fellbaum, editor. *WordNet: an electronic lexical database*. MIT Press, 1998.

[13] Desmond Darma Putra, Abdul Arfan, and Ruli Manurung. Building an indonesian wordnet, 2008.

[14] Gunawan and Andy Saputra. Building synsets for indonesian wordnet with monolingual lexical resources. In Minghui Dong, Guodong Zhou, Haoliang Qi, and Min Zhang, editors, *International Conference on Asian Language Processing, IALP 2010, Harbin, Heilongjiang, China, 28-30 December 2010*, pages 297–300. IEEE Computer Society, 2010.

[15] ChukFong Ho, Masrah Azrifah Azmi Murad, Rabiah Abdul Kadir, and Shyamala C. Doraisamy. Word sense disambiguation-based sentence similarity. In Chu-Ren Huang and Dan Jurafsky, editors, *COLING (Posters)*, pages 418–426. Chinese Information Processing Society of China, 2010.

[16] Palakorn Achananuparp, Xiaohua Hu, and Xiajiong Shen. The evaluation of sentence similarity measures. *DaWaK '08 Proceedings of the 10th international conference on Data Warehousing and Knowledge Discovery*, pages 305 – 316, 2008.

[17] Yuhua Li, David Mclean, Zuhair B, James D. O'shea, Keeley Crockett, James OŠShea, Zuhair Bandar, and Keeley Crockett. Sentence similarity based on semantic nets and corpus statistics. *IEEE Transactions on Knowledge and Data Engineering*, 18:1138–1150, 2006.

[18] Aristoteles, Yeni Hendriyeni, Ahmad Ridha, and Julio Adisantoso. Text feature weighting for summarization of documents in bahasa indonesia using genetic algorithm. *International Journal of Computer Science Issues*, 9:1, 2012.

[19] Gregorius S. Budhi, Rolly Intan, Silvia R, and Stevanus R.R. Indonesian automated text summarization. Master's thesis, Petra Christian University, Information Engineering Dept, 2009.

[20] Peter F. Brown, Peter V. deSouza, Robert L. Mercer, Vincent J. Della Pietra, and Jenifer C. Lai. Class-based n-gram models of natural language. *Computational Linguistics*, 18:467–479, 1992.

[21] Jelita Asian, Hugh E. Williams, and S. M. M. Tahaghoghi. Stemming indonesian, 2005.

[22] Mirna Adriani, Bobby Nazief, S. M.M. Tahaghoghi, and Hugh E. Williams. Stemming indonesian: A confix-stripping approach. *ACM Transactions on Asian Language Information Processing*, 6, 2007.

Bank Envi : Thailand's First Online Social Network For Environment

Todsaporn Chantarasukree
Department of Computer Science, Faculty of Science
Khon Kaen University
Khon Kaen, Thailand
todsaporn@kkumail.com

Sumonta Kasemvilas
Department of Computer Science, Faculty of Science
Khon Kaen University
Khon Kaen, Thailand
sumkas@kku.ac.th

***Abstract*—Creating Green Application for Social Network which applies for the environment in Thailand is another way of promoting environment solutions and save energy and the environment. Bank Envi is Thailand's first online social network for environment which was developed by following the principles mentioned above and projects which uses the Bank Envi Platform. This Platform consists of 3 layers namely: Application Layer, Process Layer and Resource Layer. This application consists of Bank Envi Real time Weather, Storm Path, Real Time Environment Problem Detect and Bank Envi Alert Environment Problem. The first time this project used was from 2011 to 2013 have user one thousand more than and growing every month. Environment problems in Thailand are divided into soil, water, wind, fire, forest, pollution and others. The results show that causes of environmental problems are pollution 35%, water 20%, forest 10%, fire 10%, soil 5% and other problems 20%.**

Keywords— Social Network, Bank Envi, Web 2.0, Online Social Network For Environment, Web Platform, Green Applications

I. INTRODUCTION

According to the latest report by the Ministry of Natural Resources and Environment(Thailand) [1] in the recent years there has been severe flooding in many areas nationwide causing widespread damage to life, property, natural resources and the environment which seriously affects the population. One alternative is to inform the population about the environment problems, and promote this information via social network [5]

Many Thai Web site developers use social network for environment. Thai people and the media tend to use social networks, such as facebook and Twitter to inform the public about environment problems occurring within Thailand. In the meantime if there is an earthquake, Thai network can not be used outside the country, but Intranet may be used inside the country to reliably inform the public about the earthquake. The best remedy is for Thailand to have its own social network system within the country, so that the public can rely on it when such problems occur.

Thai people must be able to use a social network for environmental problem which has been developed especially for them. Project Bank Envi (http://th.bankenvi.com) is similar in many ways to social networks such as facebook and Twitter [3] [7]. Bank Envi was created for Thai environment and Thai people in particular. Researchers believe that this project can be applied to other countries or to private enterprise. Thailand needs a social network for environment in particular in case of natural disaster in the future.

II. MATERIALS AND METHODS

Bank Envi was created in 2009 and used in 2011 when severe flooding occurred [6] News was reported by TV, radio and online network, such as facebook and Twitter, which gives information unspecific and slow. The platform was created to quickly inform the public about natural disaster in Thailand. The Bank Envi platform was created in the social network for particular environment.

Bank Envi has become a new social network for environment for Thai people in particular. Bank Envi is a (green application) which aims to share information on saving energy, guidance to conserve the environment and informing others about environmental problems occurring in Thailand.

A. Bank Envi Architecture

Main architecture of Bank Envi is divided into three main factors shown in Fig. 1.

1) Application Layer is the level of user interface level which works by login via the user interface. This interface provides three web operations using responsive web design (RWD) [5] and Web 2.0 [2].

a) Mobile Web support mobile device display.

b) Desktop Web support desktop and laptop computers display.

c) Other Device Web support other device display such as tablet.

2) Process Layer is divided into two parts

a) Social Network shows its internal operation which consists of analysis (1) Environment problem detect is the process verifying the incoming information whether it is on environment problem or general information. verifying the type of problem consists of soil, water, forest, pollution and others, and verifying the location of problems, (2) User contents is the part of users which consists of setting profile and privacy, updating state, informing the environment, adding friends, sharing pictures and others.

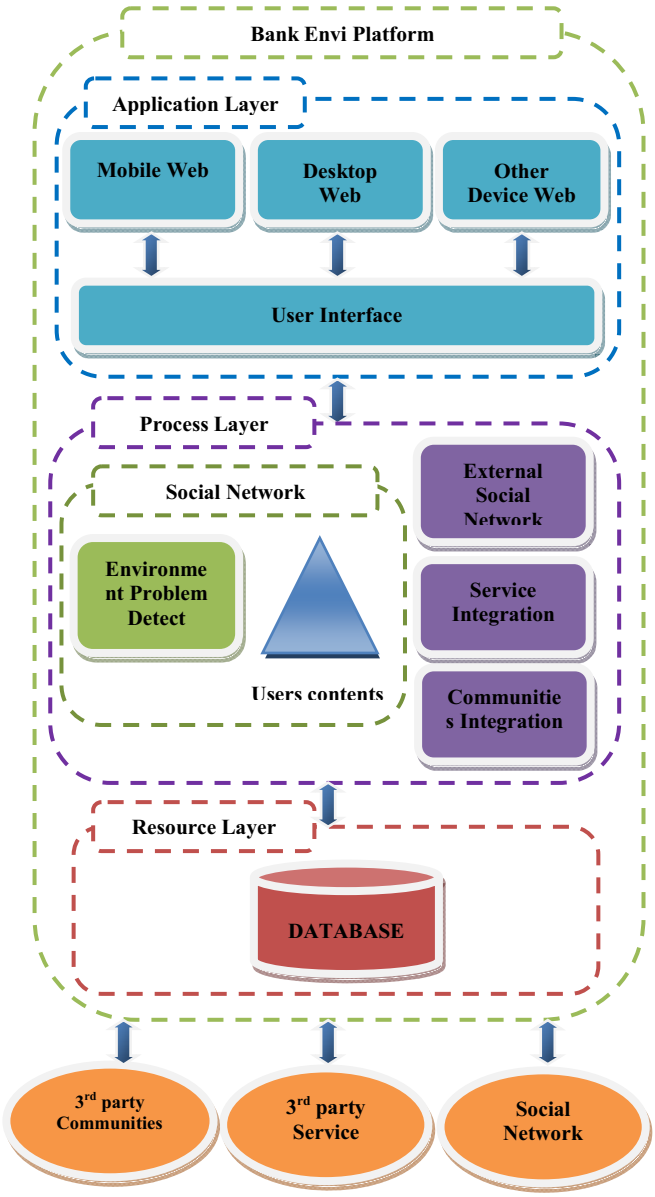


Figure 1. Application structure of *Bank Envi*.

b) Third party has three options available to contact third parties. (1) External social network which connects with other social network, such as facebook to collect users' information when applying or login. (2) Service Integration used for connecting with other service system, such as Google Calendar to send SMS, (3) Community Integration is the connection to other communities, such as the collection of climate information from the Department of Meteorology.

3) Application Layer is processed by using information base to collect all information which connects with Process Layer in reading and writing the information.

B. Hardware and Software

To design and develop the Bank Envi platform successfully and effectively, Bank Envi has used the software technology as listed in Table I.

Bank Envi was developed and tested by using Window XP SP3, evaluation unit of Intel Core2 Duo, speed 2.4 GHz, and memory unit 4 GB.

Bank Envi has been used in VPS (Virtual Private Server) on the Internet until now install in Linux CentOS which belongs to DellTM PowerEdge, model R210 E3-1230, evaluation Unit of Intel Xeon E3-1230 processor, speed 3l20 GHz, 8M cache, 16 GB (4x4GB) 1333MHz memory.

III. RESULTS

The work described in this paper has resulted in several releases of the Bank Envi web system. Currently, Bank Envi beta has come to replace the non-platform-based Bank Envi on the server in Fig 2-6 shown on screen of Bank Envi. Shows how it works in multi-functions as explained in Bank Envi below:

Table I. List of software and related technologies used in Bank Envi

Languages
• PHP
• Javascript
• SQL
• XML
• ActionScript 3.0
Web standards
• HTML 4.0, HTML 5
• XHTML 1.0
• CSS (version 2+, 3.0)
Web 2.0 and Framework technologies
• AJAX Technologies
• jQuery Framework
• URL Rewrite Framework
• Zend framework
Database
• Mysql
External APIs
• facebook API
• Google Map API
• Google Calendar API
External Services
• Thai Meteorological Department Service
Servers
• Hosting (Apache Server, MySQL server)

A. Bank Envi Platform

Applying for use in social network to be the centre for environmental monitoring in Thailand by creating the Bank Envi Platform have ability and qualification. (1) set profiles and privacy, (2) update status messages, (3) add friends and

create groups, (4) initiate and join discussions, (5) share photos and videos, (6) create user blog, and (7) set environment problems alert.

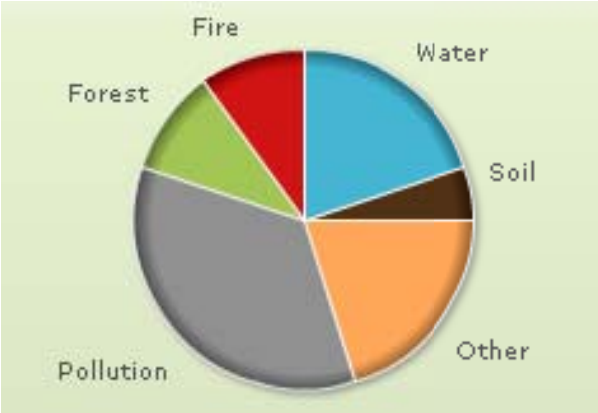


Figure 2. Environment problems reported in Bank Envi

Results of using Bank Envi, model Beta since 2011 until now including environment problems, such as soil, water, wind, fire, forest, pollution and others. In Thailand, informing about environment problems was divided into pollution 35%, water 20%, forest 10%, fire 10%, soil 5% and others 20% as shown in Fig. 2.

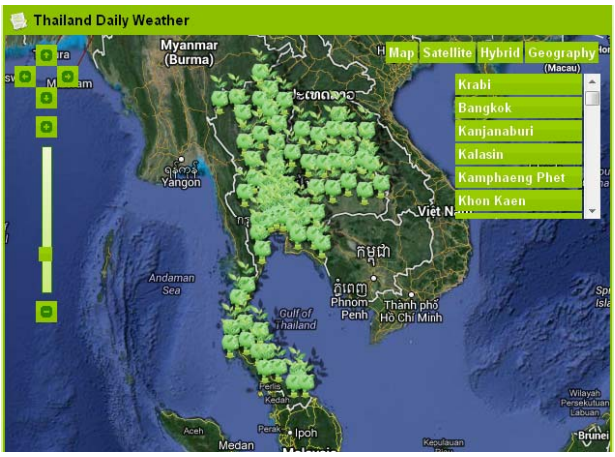


Figure 3. Bank Envi RealTime Weather

3rd party service by the Thai Department of Meteorology using the Bank Envi Platform has become the Bank Envi Real Time Weather system which displays climate daily in all 77 provinces of Thailand as shown in Fig. 3. It also shows the direction of a storm as shown in Fig. 4.

Figure 4. Direction of storms

Bank Envi platform in informing about environment of Real Time is able to inform via mobile, devices (Mobile Web and SMS) and desktop as shown in Fig. 5. and Fig. 6.

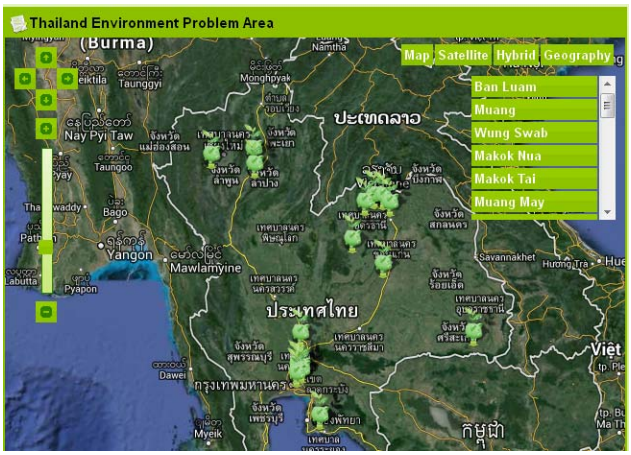


Figure 5. Bank Envi Real Time Environment Problem Detect

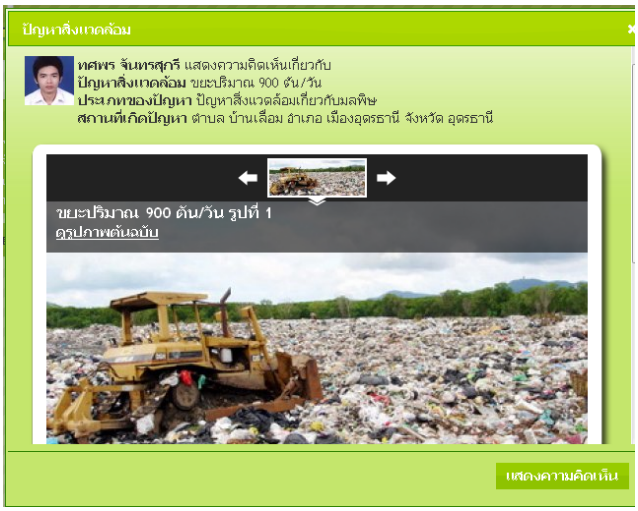


Figure 6. Bank Envi Alert Environment Problem

Bank Envi other features consists of (1) Inform action about environment problems, (2) Method which can save energy, and (3) Methods to conserve the environment. Information a nature disaster in Thailand is divided into province, district, and sub-district.

B. Mobile Access

To allow Bank Envi Platform tobetter permeate all Thai users, mobile is supported that let users access and update their profiles anytimeand anywhere. The underpinning technology here is the Wireless Access Protocol 2.0 (WAP 2.0) which is the scaled-down version of the Extensible Hypertext Markup Language (XHTML). Bank Envi platform is able to detect mobile devices accesses as well as mobile product vendors automatically.

IV. CONCLUSION AND FUTURE WORK

Since 2011 – Now Bank Envi has released the model Beta on the Internet at http://th.bankenvi.com. In Table II, shows the history of each model in this research.

Table II. Release version of Bank Envi

Release	Features
1	• Introducing online social network for environment • Users can add and write messages to friends
2, 3, 4, 5	• Contents directed to create environment problem, How to save energy, and environmental conservation.
Alpha	• Introducing Web 2.0 technologies • AJAX in every web page
Beta	• Bank Envi platform • Include 3rd party Communities • Include 3rd party Service • Include 3rd Social Network • Bank Envi Mobile Version • Graph Environment Problem • Bank Envi RealTime Weather • Real Time Environment Problem Detect • Bank Envi Alert Environment Problem

To achieve the goal of creating an online social network environment for Thai people, Bank Envi platform has resorted to Web 2.0 technologies to deliver effective contents and applications. From the start the strategy has always been to build an epicenter of Web 2.0 awareness among Thai developers and users and allow such that (1) potential and benefits of Web 2.0 technologies would be better appreciated, and (2) awareness and appreciation of Web 2.0's potential and benefits could be turned into rich Internet contents and applications suitable for sharing in the Bank Envi society with effective contents and applications, an effective online social network for environment.

By developing Bank Envi via social networks and educating the public on the application to the environment in Thailand. Bank Envi allows users to. (1) Access information on the problems of environment, (2) Access information method to save energy, and (3) Show how to conserve the environment. This will allow users to help each other find ways to solve problems of the environment and share ideas about environment matters in Thailand. This will promote a better society by enhancing knowledge and data about the environment in Thailand.

User notify environment problems to Bank Envi since 2011 – Now there are pollution 35%, water 20%, forest 10%, fire 10%, soil 5%, and others 20%. Planned future work includes developing to support other languages, mobile application version, and API Framework.

ACKNOWLEDGMENT

The authors would like to thank adviser for help. This work was supported in enhancing the strength to produce graduate scholarship by Department of Computer Science, Faculty of Science Khon Kaen University 2012.

REFERENCES

[1] Ministry of Natural Resources and Environment, "Annual Report 2011," http://lib.mnre.go.th/lib/Annual_report/Annual_Report_2554.pdf (last access: May 2013).

[2] P. Sharma, "Core Characteristics of Web 2.0 Services," http://www.techpluto.com/web-20-services(last access: May 2013).

[3] WikiPedia, "facebook," http://en.wikipedia.org/wiki/facebook (last access: May 2013).

[4] WikiPedia, "Responsive web design," http://en.wikipedia.org/wiki/Responsive_web_design (last access: May 2013).

[5] WikiPedia, "Social Network System," http://en.wikipedia.org/wiki/Social_network (last access: May 2013).

[6] Bank Of Thailand, "Thai Flood Survey 2011," http://www.bot.or.th/Thai/EconomicConditions/Thai/BLP/Documents/ThaiFloodSurvey2011.pdf (last access: May 2013).

[7] WikiPedia, "Twitter," http://en.wikipedia.org/wiki/Twitter (last access: May 2013).

Cascade Generalization for Breast Cancer Detection

[1]Kuntoro Adi Nugroho, [2]Noor Akhmad Setiawan, [3]Teguh Bharata Adji
Department of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
[1]kuntoro_ti08@mail.ugm.ac.id, [2]noorwewe@ugm.ac.id, [3]adji@ugm.ac.id

***Abstract*— Mammography is known as the preferred method for breast cancer diagnosis. Researchers have proposed machine learning based methods to improve the detection of breast cancer using mammography. In this study, cascade generalization is proposed for breast cancer detection. Four Bayesian Network based methods, SVM, and C4.5 are evaluated in loose coupled cascade classifier. The Bayesian based methods are evaluated in both base level and meta level. The evaluation results show the superiority of the proposed cascade strategy compared to Bagging and single classifier approach. Naive Bayes with SMO cascade demonstrated the best result in terms of ROC area under curve of 0.903. Bayesian Network using Tabu search with SMO cascade demonstrated the best accuracy of 83.689%.**

Keywords*— *cascade generalization, mammography, breast cancer detection, Bayesian Network, C4.5, Sequential Minimal Optimization

I. INTRODUCTION

Mammography is one among several methods to perform breast cancer screening and is preferred for detection. In mammography, X-ray image -called mammogram- is taken to examine the presence of tumor. Interpreting mammogram is not an easy task. Computer aided diagnosis can be used to improve the effectiveness and the efficiency of diagnosis [1].

Some researchers have proposed machine learning based method to improve breast cancer detection using mammography. Mokhtar and Elsayad [2] evaluated three data mining methods for classifying breast masses, namely Decision Tree, Artificial Neural Network, and Support Vector Machine. The decision tree used in the study is based on chi-square test, called Chi-squared Automatic Interaction Detector. The artificial neural network is implemented using pruning method and uses time as stopping criterion. The Support Vector Machine uses polynomial kernel. The method was validated using mammographic mass data set taken from Institute of Radiology, University Erlangen [3]. The data was split into two partitions, 70% data as training set and 30% data as validation set. Artificial Neural Network was superior in terms of accuracy and sensitivity. Support Vector Machine demonstrated the best result in ROC area under curve of 0.831.

In some other studies, artificial neural network is also proposed for mammography screening. Saritas [4] proposed three layer back-propagation network. In the study, ambiguous data from mammographic mass data set [3] were removed. The data was split into two partitions. One hundred and sixty nine instances were randomly selected as validation set, and the remaining instances were used as training set. The proposed method demonstrated 85.5% accuracy. Keleş et al. [5] proposed expert system for breast cancer diagnosis based on neuro-fuzzy rules. There are nine fuzzy rules which are represented in the hidden layer layer of the neural network.

Based on the previously discussed studies, Artificial Neural Network demonstrated good result in predicting breast cancer. However, the implementation of single learning algorithm can be improved further. Ensemble learning is one among several methods that can be used to improve the generalization ability of learning algorithm.

Ensemble learning was proposed by some researchers. Feature selection and ensemble learning were proposed by Luo and Cheng [6] for diagnosing breast masses. In the study, mammographic mass data set from Institute of Radiology, University of Nuremberg [3] was used. Forward and Backward selection were implemented and both of the method eliminated one attribute. Bagging, AdaBoost, and MultiBoost were evaluated. SVM and Decision Tree were used as base classifier for the ensemble methods. Based on the study, although some ensemble methods are not always better than single method, the proposed method is useful to improve the predictive performance of learning algorithm. Novakovic and Veljovic [7] proposed Principal Component Analysis and Rotation Forest ensemble. Principal Component Analysis was used to transform the data. Twelve variants of decision tree were used as base classifier for Rotation Forest. The evaluation results show that the performance of some decision tree algorithms were increased but some others are decreased. Based on the studies, classifier ensemble method did not always perform better than single classifier based method.

One of several alternatives for improving the accuracy of classification algorithm is cascade generalization. Cascade generalization is a multiple expert method in which classifiers are arranged in sequence. Training data for a classifier is extended by the prediction of the proceeding classifier. Based on the study from Gama and Brazdil, this method could improve the accuracy of classification algorithm. For example, the accuracy of C4.5 decision tree can be improved using cascade generalization with Naive Bayes as base classifier [8].

Several studies on cascade generalization have been carried out. Barakat proposed CGen-SVM, an ensemble based method which comprises three classifiers [9]. In the method, one of the classifier is cascade classifier and the other two classifiers are non-cascade classifiers. Two learning algorithms are used. One algorithm is used to construct the meta-level classifier and the other one is used to construct the base classifier. SVM is used

as base classifier. The evaluation results of CGen-SVM show that the proposed method improves the prediction performance of C4.5, RIPPER, and CART. Kotsiantis and Kanellopoulos [10] proposed C4.5 with M5' cascade. Evaluated on 24 benchmark data sets, the proposed method demonstrated better results compared to several ensemble methods.

Based on the discussed studies, cascade generalization is a good alternative to improve the predictive performance of learning algorithm. By using low variance algorithm like Naive Bayes as base level algorithm, the performance of C4.5 can be improved. Naive Bayes can be seen as special type of Bayesian Network in which all attributes are independent given the value of the class attribute [11]. Naive Bayes is only one among several joint probability model alternatives that can be created from data. Thus, there are more than one alternative graphical structures that can be constructed.

The objective of this study is to develop method to improve the detection of breast cancer diagnosis using cascade Bayesian Network. In this study, several search algorithms for constructing Bayesian Network are evaluated. The cascade generalization technique is also compared with some ensemble based methods.

The contribution of this study is to improve the quality of data mining based breast cancer detection using cascade generalization. In this study, cascade generalization with loose coupling strategy is proposed. In contrast to the study by Gama and Brazdil [8], four Bayesian Network based methods are evaluated. Different methods probably construct different network models. Thus, this study is also aimed to evaluate the performance of different Bayesian Network based methods in cascade strategy for breast cancer detection case. Besides the use of several Bayesian Network based methods, in contrast to the study by Barakat [9], SVM is proposed in this study for both meta-level and base-level algorithm in cascade generalization.

II. METHODOLOGY

A. Data set

Mammographic mass data set [3] which collected from the Institute of Radiology, University Erlangen-Nuremberg is used in this study. The data consist of 961 instances and 6 attributes. The summary of the attributes are described in Table I.

There are two class value for the instances, benign (516 instances) and malignant (445 instances). One instance has BI-RADS with value 55. This instance is considered error and the BI-RADS value is set to 5.

B. The proposed method

Gama and Brazdil [8] proposed two strategies of cascade method, namely loose coupling strategy and tight coupling strategy. In loose coupling strategy, training set for a classifier is extended by the proceeding classifier. In tight coupling strategy, the extension of data set is performed in divide and conquer algorithm, such as decision tree.

In this study, cascade generalization with loose coupling strategy is proposed. Based on some previous studies, loose coupling strategy could improve the performance of learning algorithms. This method is also compatible with non-divide and conquer algorithm.

In cascade classifier with loose coupling strategy, one classifier is trained and the classifier's output is used to extend the training data set for the following classifier. For example, let learning algorithm L1 and L2 are used in loose coupling strategy. Let L1 be used as base level learning algorithm. C1 is a classifier produced by applying L1 to the training set. C1 is applied to the training data and used to extend the training data using its posterior probability predictions for each class value. The extended data is then used to train L2 as meta level learning algorithm to produce meta level classifier C2.

TABLE I. ATTRIBUTE DESCRIPTION

No	Attribute	Description
1	BI-RADS	BI-RADS (Breast Imaging Reporting and Data System) assessment result. In the data set, the value ranges from 0 to 6. (nominal)
2	Age	Patient's age in years (integer, numeric)
3	Shape	Mass shape, 1 = round, 2 = oval, 3 = lobular, 4 = irregular (nominal)
4	Margin	Mass margin, 1 = circumscribed, 2 = microlobulated, 3 = obscured, 4 = ill-defined, 5 = spiculated (nominal)
5	Density	Mass density, 1 = high, 2 = iso, 3 = low, 4 = fat-containing (nominal)
6	Severity	Class attribute, 0 = benign and 1 = malignant

There are two values for the class attribute of the mammographic mass data set, namely benign and malignant. Base level classifier is used to extend the data set using one attribute which value is the posterior probability of one class value. The extension is performed using one posterior probability class value only as the sum of the probability of benign and the probability of malignant is 100%.

The cascade generalization strategy could be explained in flowchart as depicted in Figure I. The left flowchart illustrates the training procedure and the right flowchart illustrates the testing procedure.

Gama and Brazdil [8] evaluated three methods in loose coupling strategy, namely C4.5, Naive Bayes, and Linear Discriminant classifier. Based on the study, it is suggested that algorithm with low bias is used at meta level and algorithm with low variance is used at base level. Naive Bayes could improve the performance of C4.5 due to its low variance property.

Naive Bayes assumes that all attributes are independent given the class. In order to represents the joint probability of the variables without such assumption, Bayesian Network with several search algorithms could be used. Bayesian Network is a directed acyclic graph which represents the joint probability distribution over random variables [11].

There are some known search algorithms to construct Bayesian Network, some of which are

a) K2 [12]

b) Tabu search [13]

c) TAN (Tree Augmented Naive Bayes) [11]

Based on the previously discussed studies, only Naive Bayes that was proposed for cascade generalization strategy. In order to understand the impact of other type of Bayesian Network based methods in cascade generalization for the given mammography case, more methods are evaluated. Four types of Bayesian based methods are used, namely Naive Bayes, Bayesian Network with K2 search, Bayesian Network with Tabu search, and Bayesian Network with Tree Augmented Naive Bayes.

Gama and Brazdil [8] suggested that learning algorithm with low variance is used to construct base classifier. However, in order to evaluate the performance of Bayesian method as base classifier and meta classifier, the Bayesian methods are used both as base classifier and meta classifier.

C4.5 [14] and Support Vector Machine with Sequential Minimal Optimization (SMO) [15] were used in the study by Luo and Cheng [6]. These methods were also proposed in the studies of cascade generalization [9], [10]. The performance of C4.5 can be improved using cascade generalization as meta level learning algorithm. SVM was also proposed but it is used as base level algorithm. These methods are also proposed in this study to construct the cascade classifier as both base and meta level learning algorithm.

Luo and Cheng [6] proposed two algorithms to construct the base classifiers of ensemble, namely C4.5 and SVM (SMO). Based on the evaluation result, Bagging ensemble can be used to improve the accuracy of the algorithms.

There are twenty four methods which are evaluated in this study. The methods comprises learning algorithms in single mode, learning algorithms in cascade mode, and some ensemble methods. The combination for cascade based method comprises two group of methods, the Bayesian method in one group and C4.5 and SMO in the other group. If one Bayesian method is used as base learning algorithm, the other group is used as meta classifier, and vice versa.

In the study by Luo and Cheng [6], Bagging can improve the accuracy of C4.5 and SMO. For comparison purpose, the cascade methods are also compared with Bagging with C4.5 and SMO as base classifier.

The evaluated algorithms are summarized in Table II. NB denotes Naive Bayes. SMO and J48 denotes SVM with SMO and C4.5 respectively. K2BN, TabuBN, and TanBN denotes Bayesian Network with K2, Tabu, and TAN search algorithm respectively. The cascade methods are denoted using understrike. For example, NB_SMO denotes a cascade classifier with Naive Bayes in meta level and SMO in base level. Bagging with SMO and C4.5 as base classifiers are denoted as Bg-SMO and Bg-J48 respectively.

The algorithms used in this study are

a) Naive Bayes [16]

Naive Bayes algorithm is a learning method to construct Bayesian Network form data in which the class attribute becomes the root of the tree and all attributes are independent given the class attribute.

b) Bayesian Network – K2 search [12]

K2 search constructs Bayesian Network by processing attributes in sequence. Attribute is represented as node in Bayesian Network. For each turn, edges are tried to be added to a node from the previous nodes. The process moves to the next node if there is no improvement by adding edges to the current node. This method depends on the ordering of the nodes.

c) Bayesian Network – TAN (Tree Augmented Naive Bayes) [11]

Biasing the structure of Bayesian Network could improve the performance for classification. TAN is similar, but not the same, with Naive Bayes as TAN begins with Naive Bayes structure. Then, edges could be added between attributes to reduce the independence assumption.

d) Bayesian Network – Tabu Search [13]

One of search algorithms alternative for constructing Bayesian Network is Tabu Search. In this algorithm, arbitrary solution is selected and the search process continued by searching the neighboring solutions. A special data structure to avoid local optimum solution is used, which called Tabu list. If the neighbors of a solution do not optimize the given cost function, this solution is added to the list and the search process continues to other solution which mostly optimize the cost function. The best solution during the search process is returned as the final solution.

e) C4.5 (J48) [14]

C4.5 was developed by Ross Quinlan. This method is an improvement to ID3. Two general process in C4.5 are tree construction and tree pruning. Among the improvements are numerical attribute handling and missing value handling.

f) Sequential Minimal Optimization SVM [15]

Sequential Minimal Optimization (SMO) is one of optimization methods for Support Vector Machine in which quadratic programming optimization (QP) problem is broken into small QP problems. The problems are then solved analytically. This method has linear time complexity thus can handle large training set.

g) Bagging [17]

Bagging is among ensemble construction methods which is based on instance perturbation. Different classifiers are constructed using bootstrap samples of training data.

The evaluation is performed using 20 runs of 10-fold stratified cross validation. The final result is taken as the average across all runs. The accuracy, ROC area under curve, true positive rate, and true negative rate are measured for all methods.

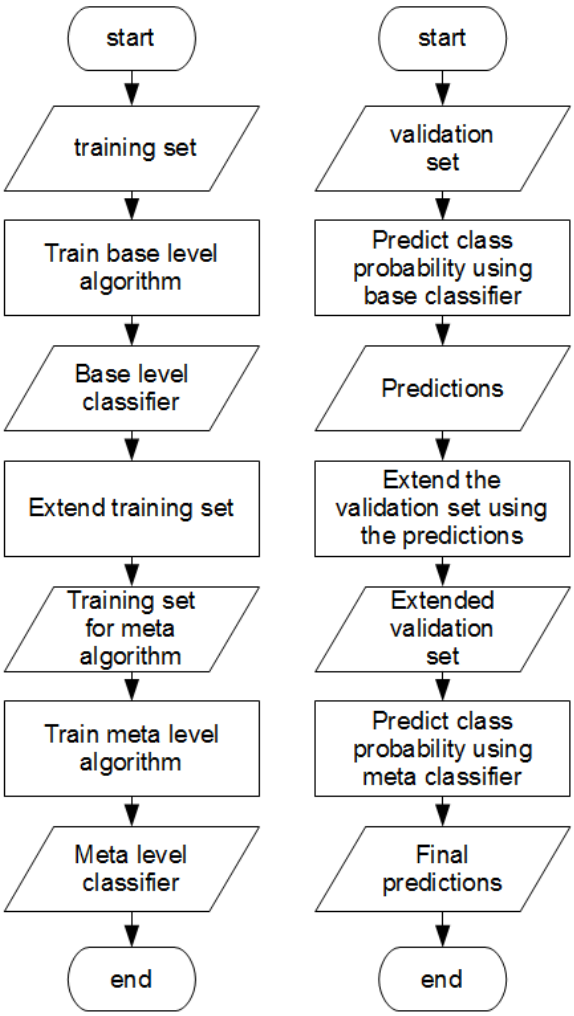


Fig. 1. Cascade generalization with two learning algorithms. Left: training procedure, Right: testing procedure

TABLE II. THE EVALUATED METHODS

Type	Methods
Single method	NB, K2BN, TabuBN, TanBN, SMO, J48
Cascade based method	NB_SMO, K2BN_SMO, TabuBN_SMO, TanBN_SMO, SMO_NB, SMO_K2BN, SMO_TabuBN, SMO_TanBN, NB_J48, K2BN_J48, TabuBN_J48, TanBN_J48, J48_NB, J48_K2BN, J48_TabuBN, J48_TanBN
Ensemble method	Bg-SMO, Bg-J48

III. RESULT AND DISCUSSION

A. Result

The evaluation result is summarized in Table III. On the table, Acc denotes accuracy in percentage, ROC AUC denotes Receiver Operating Characteristic Area Under Curve, TN denotes true negative rate, and TP denotes true positive rate.

The best results are printed in bold. TabuBN with SMO cascade demonstrated the best result in accuracy and Naive Bayes with SMO cascade demonstrated the best result in ROC area under curve. SMO performed the best in terms of true negative rate and C4.5 with Naive Bayes cascade demonstrated the best result in true positive rate.

B. Discussion

Based on the evaluation results, there is no single method that performed the best in all indicators. For example, Naive Bayes demonstrated good performance in terms of accuracy, ROC area under curve, and true positive rate, but low performance in terms of true negative rate. SMO demonstrated the best performance in true negative rate but very low performance in true positive rate. Naive Bayes with SMO cascade is superior in ROC area under curve and performed better in terms of true negative rate than true positive rate. Based on the evaluation results, a method which performed well in true negative rate tends to performs poorly in true positive rate, and vice versa. However, some methods performed quite balance in true positive and negative rate, such as Naive Bayes with C4.5 cascade and K2BN with C4.5 cascade.

Some cascade methods could improve the performance of the component learning algorithms. For example, Bayesian Network using tabu search with SMO cascade demonstrated better accuracy than both Bayesian Network using tabu search and SMO. Naive Bayes with SMO cascade demonstrated better ROC area under curve compared to both Naive Bayes and SMO. However, based on the result, there is a trade-off in terms of true positive rate and true negative rate. For example, Naive Bayes with SMO cascade performed better than Naive Bayes in terms of true negative rate, but performed worse in terms of true positive rate.

Based on the study by Gama and Brazdil [8], low variance method and low bias method are recommended for base classifier and meta level classifier respectively. The success and the failure of the cascade methods in this study probably because of the bias-variance characteristic of the methods.

ROC area under curve could be used as indicator to describe the performance of the methods in spite of the trade-off between sensitivity and specificity. Cascade generalization improves SMO in terms of ROC area under curve. All cascade methods based on SMO demonstrated better performance than single SMO. Cascade generalization also improves C4.5 in terms of ROC area under curve when C4.5 is used as base classifier. In terms of ROC area under curve, all cascade methods with SMO and J48 as base classifier demonstrate better result than Bagging with SMO and J48 respectively.

For breast cancer detection case, obtaining high sensitivity / true positive rate is important although the false positive rate becomes high. In breast cancer screening, false negative result, i.e. when a diseased case is detected as healthy case, could lead to more serious consequences [18]. Thus, the methods with high sensitivity or true positive rate are more considered for mammography screening application.

C4.5 with Naive Bayes cascade demonstrated the best result in sensitivity. However, the ROC area under curve is low compared to other methods. Naive Bayes with SMO demonstrated the best performance in terms of ROC area under curve, but with low sensitivity. To achieve higher sensitivity, the decision threshold of the classifier can be adjusted, although the adjustment will decrease the true negative rate.

TABLE III. EVALUATION RESULT

Bayesian Methods				
	Acc (%)	***ROC AUC***	***TN***	***TP***
NB	83.252	0.901	81.541	85.236
K2BN	82.253	0.899	80.736	84.011
TabuBN	82.336	0.899	80.698	84.236
TanBN	82.248	0.895	84.409	79.742
SMO related methods				
	Acc (%)	***ROC AUC***	***TN***	***TP***
SMO	81.483	0.808	90.300	71.258
NB_SMO	83.382	0.903	84.903	81.618
K2BN_SMO	83.673	0.902	86.986	79.831
TabuBN_SMO	83.689	0.901	87.054	79.787
TanBN_SMO	82.045	0.898	82.955	80.989
SMO_NB	82.331	0.822	84.138	80.236
SMO_K2BN	82.466	0.824	83.459	81.315
SMO_TabuBN	82.627	0.826	83.430	81.697
SMO_TanBN	82.451	0.822	85.891	78.461
Bg–SMO	81.717	0.874	89.845	72.292
J48 related methods				
	Acc (%)	***ROC AUC***	***TN***	***TP***
J48	81.733	0.859	83.895	79.225
NB_J48	83.205	0.898	82.936	83.517
K2BN_J48	83.091	0.899	82.345	83.955
TabuBN_J48	83.012	0.895	81.890	84.315
TanBN_J48	81.868	0.894	82.355	81.303
J48_NB	83.012	0.826	80.107	86.382
J48_K2BN	82.050	0.858	85.572	77.966
J48_TabuBN	82.112	0.859	85.610	78.056
J48_TanBN	81.530	0.833	82.141	80.820
Bg–C4.5	82.763	0.879	85.659	79.404

IV. CONCLUSION

In this study, classifier with cascade generalization strategy is proposed for mammography based breast cancer detection. Four Bayesian based methods, SMO, and C4.5 are proposed for the cascade method. The evaluation result shows that some of the cascade methods demonstrated better result in terms of accuracy and ROC area under curve compared to Bagging with C4.5, Bagging with SVM, and the single classifier methods. The highest ROC area under curve of 0.903 is achieved by implementing Naive Bayes with SMO cascade. The highest accuracy of 83.689% is achieved by implementing Bayesian Network using Tabu search with SMO cascade.

REFERENCES

[1] M. L. Giger, "Computer-aided diagnosis of breast lesions in medical images," Comput. Sci. Eng., vol. 2, no. 5, pp. 39–45, 2000.

[2] S. A. Mokhtar and A. Elsayad, "Predicting the Severity of Breast Masses with Data Mining Methods," Arxiv Prepr. Arxiv13057057, 2013.

[3] M. Elter, R. Schulz-Wendtland, and T. Wittenberg, "The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process.," Med. Phys., vol. 34, no. 11, pp. 4164–4172, Nov. 2007.

[4] I. Saritas, "Prediction of Breast Cancer Using Artificial Neural Networks," J. Med. Syst., vol. 36, no. 5, pp. 2901–2907, Aug. 2011.

[5] A. Keleş, A. Keleş, and U. Yavuz, "Expert system based on neuro-fuzzy rules for diagnosis breast cancer," Expert Syst. Appl., vol. 38, no. 5, pp. 5719–5726, May 2011.

[6] S.-T. Luo and B.-W. Cheng, "Diagnosing Breast Masses in Digital Mammography Using Feature Selection and Ensemble Methods," J. Med. Syst., vol. 36, no. 2, pp. 569–577, May 2010.

[7] J. Novakovic and A. Veljovic, "Interpretation of mammograms with rotation forest and PCA," in Applied Computational Intelligence and Informatics (SACI), 2011 6th IEEE International Symposium on, 2011, pp. 571–575.

[8] J. Gama and P. Brazdil, "Cascade generalization," Mach. Learn., vol. 41, no. 3, pp. 315–343, 2000.

[9] N. Barakat, "Cascade generalization: Is SVMs' inductive bias useful?," in Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on, 2010, pp. 1393–1399.

[10] S. Kotsiantis and D. Kanellopoulos, "Cascade Generalization with Classification and Model Trees," presented at the Convergence and Hybrid Information Technology, 2008. ICCIT '08. Third International Conference on, 2008, pp. 248–253.

[11] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," Mach. Learn., vol. 29, no. 2–3, pp. 131–163, 1997.

[12] G. F. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," Mach. Learn., vol. 9, no. 4, pp. 309–347, 1992.

[13] R. R. Bouckaert, Bayesian belief networks: from construction to inference = Bayesiaanse belief netwerken : van constructie tot inferentie. Utrecht: Universiteit Utrecht, Faculteit Wiskunde en Informatica, 1995.

[14] J. R. Quinlan, C4.5: programs for machine learning. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

[15] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," 1998.

[16] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," in Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, 1995, pp. 338–345.

[17] L. Breiman, "Bagging predictors," Mach. Learn., vol. 24, no. 2, pp. 123–140, 1996.

[18] S. H. Park, J. M. Goo, and C.-H. Jo, "Receiver operating characteristic (ROC) curve: practical review for radiologists," Korean J. Radiol., vol. 5, no. 1, pp. 11–18, 2004.

Clustering Top-10 Malware/Bots based on Download Behavior

Chaxiong Yukonhiatou [#], Surin Kittitornkun [*], Hiroaki Kikuchi [$], Khamphao Sisaat [#1], Masato Terada [*1], Hiroshi Ishii [$1]

[#] *International College, King Mongkut's Institute of Technology Ladkrabang, Thailand, chaxiong@fe-nuol.edu.la*
[*] *Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Thailand, kksurin@kmitl.ac.th*
[$] *Meiji University, Japan, kikn@meiji.ac.jp*
[#1] *Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Thailand, khamphao@fe-nuol.edu.la*
[*1] *Hitachi Ltd., Japan, masato.terada.rd@hitachi.com*
[$1] *Tokai University, Japan, ishii164@tokai.ac.jp*

***Abstract*—Malware can be spread over the Internet via especially download mechanism to the victim computers. This work tries to cluster malware/bots download behavior of Top-10 malware based on 2010 and 2011 CCC (Cyber Clean Center) datasets. The datasets contain more than one million download logs collected from several independent honeypots in Japan to observe malware/bot traffic and activities. Although the daily and hourly patterns are quite similar in 2010, those of 2011 are quite different. As a result, the proposed Integral Correlation Coefficient can cluster 3 and 4 groups of Top-10 malware/bots in 2010 and 2011, respectively.**

***Index Terms*—malware; botnet; honeypot; correlation; clustering;**

I. Introduction

A collection of infected hosts or zombie PCs is called a Botnet. Botnets in particular have had an eventful history, evolving to become increasingly sophisticated and more widespread. Today, a large scale of computers under control of some botnets is alarming, and many of recent cyber attacks are being launched by botnets for purposes of carrying out cyber attacks such as spam e-mail, malware downloads, Distributed Denial of Services(DDoS) and so on. Millions of systems can come under the control of a single malicious network. It is unsurprisingly difficult to pinpoint the moment where botnets became a reality, but a Trojan and a worm are seen as malware that helped initiate the rise of the botnet. It could run custom scripts in response to IRC events and it had access to raw TCP (transmission control protocol) and UDP (user datagram protocol) sockets, making it perfect for rudimentary Denial of Service attacks. Considering most botnets consist of a lot of bots which have been infected by a specific malware and they also try to propagate themselves into other vicim systems through the Internet. It is needed to better understand the basic infrastructure of botnets as well as the underlying malware so that heavy damages of botnet based cyber attacks can be mitigated. In addition, since the activities of botnets appear in various forms of traffic such as spamming email, scanning packets, exploit codes, etc., we need to analyze various kinds of data source to obtain more accurate and practical solutions.

In this paper, we investigate especially the download behavior of Top-10 malware/bots. We have analyzed the download logs, also known as CCC (Cyber Clean Center) dataset, from infected hosts around the world to the observed honeypots located in Japan. The rest of this paper is organized as follows. In Section II and III, we give a brief review of related work and the description of CCC datasets as well as our methods. In Section IV, results are presented and discussed in details. Finally, this paper is concluded and future work is suggested in Section V.

II. Related Work

Reference [1] explains the purpose of Cyber Clean Center Dataset in Japanese. The dataset has already been studied and analyzed as follows. Roysid et al. [2] brought up the ideas to detect sequential attack patterns using the PrefixSpan method for malware identification based on 2009 CCC dataset. The correlation analysis between 10 spamming botnets is presented and based on analysis of 3 weeks of spam email in their Darknets and Honeypots (CCC Dataset 2010) by [3]. By comparing members of the 10 spamming botnets with source hosts of darknet traffic and honeypot traffic,they found that 7.2% - 37.5% of spamming botnets has been infected by one of the following malware (i.e., Worm Rbot.SMA, Trojan Buzus.BEZ, Worm Palevo.SMJF, Worm Allaple.IK), and there is a high possibility that members of the 10 spamming botnets have been infected by additional malwares (e.g., Win32/Concker). Recently, Khamphao et al. [4] studied the shift of malware download frequencies relative to Japanese time zone based on CCC dataset 2010 and 2011 in order to locate the source IP addresses/countries.

III. Dataset and Methods

A. CCC(Cyber Clean Center) Dataset

In this research work, we rely on CCC dataset 2010 and 2011, which investigates more than 90 and 70 independent Honeypots to observe malware traffic on the Japanese tier-1 backbone network, respectively. A Honeypot is a virtual host running Windows XP as operating system with vulnerabilities. Each Honeypot is rebooted periodically in order to avoid an infection from being active for a long time. In other words, time to reboot is negligible short and hence, a Honeypot is supposed to be always online.

The observations are coordinated by the Cyber Clean Center (CCC). CCC dataset 2010 comprises the access binaries logs of the botnet attacks between May 1, 2009 and April 30, 2010 and CCC dataset 2011 is between May 1, 2010 and Jan 30, 2011. Some details of the dataset are listed in Table I. Each Honeypot records every packet as an item in an access log, comprising.

- *Download Timestamp*
- *Honeypot ID*
- *Source/Destination IP address*
- *Source/Destination port number*
- *Hash value (SHA1)*
- *Malware name(The Malware names are derived from the malware signature used by commercial anti-virus software; Trend Micro)*
- *Malware file name*

TABLE I: Numbers of logs recorded in CCC dataset 2010 and 2011

Details	2010	2011
Number of records	1,162,093	158,734
Based on TCP	1053,977	136,251
Based on UDP	108,116	22,483
IP addresses	176,522	82,691
Hash values	29,858	12,591
Unique malware names (excluding unknown Malware)	978	316
Number of Honeypots	92	72
Time period	12 months	9 months

B. Temporal behavior of malware downloading

We can investigate the download behavior of the malware in the CCC dataset 2010 and 2011. To analyze the download logs of each malware to the Honeypots, we can formulate the temporal behavior in terms of number of downloads per day/per hour of a given dataset. However, the variation of malware downloads can be too large to compare. Therefore, the normalized deviation of the number of malware downloads from its average should be investigated.

Definition 1: Let $l^{u,v}_{w,x,y}(d,h,m,s)$ be number of downloads of malware w which is originated from source IP address u, source port v, to destination IP address x and destination port y on day d, hour h, minute m and second s, repectively.

Therefore, *we can express some properties of the log as the following*.

- The total downloads of malware w on day d of a particular year:

$$l_w(d) = \sum_{h=0}^{23} l^{u,v}_{w,x,y}(d,h,m,s), \quad \forall u, \forall v, \forall x, \forall y, \forall m, \forall s \tag{1}$$

- The average daily downloads of malware w:

$$\bar{l}^d_w = \frac{\sum_{d=0}^{365} l_w(d)}{365} \tag{2}$$

- The daily correlation coefficient between malware $w1$ and malware $w2$:

$$X^d_{w1,w2} = \frac{\sum_{d=1}^{365}\left(l_{w1}(d) - \bar{l}^d_{w1}\right)\left(l_{w2}(d) - \bar{l}^d_{w2}\right)}{\sqrt{\sum_{d=1}^{365}(l_{w1}(d) - \bar{l}^d_{w1})^2}\sqrt{\sum_{h=1}^{365}(l_{w2}(d) - \bar{l}^d_{w2})^2}} \tag{3}$$

- The total downloads of malware w within hour h on any day of a particular year:

$$l_w(h) = \sum_{d=1}^{365} l^{u,v}_{w,x,y}(d,h,m,s), \quad \forall u, \forall v, \forall x, \forall y, \forall m, \forall s \tag{4}$$

- The average hourly downloads of malware w:

$$\bar{l}^h_w = \frac{\sum_{h=0}^{23} l_w(h)}{24} \tag{5}$$

- The hourly correlation coefficient between malware $w1$ and malware $w2$:

$$X^h_{w1,w2} = \frac{\sum_{h=0}^{23}\left(l_{w1}(h) - \bar{l}^h_{w1}\right)\left(l_{w2}(h) - \bar{l}^h_{w2}\right)}{\sqrt{\sum_{h=0}^{23}(l_{w1}(h) - \bar{l}^h_{w1})^2}\sqrt{\sum_{h=0}^{23}(l_{w2}(h) - \bar{l}^h_{w2})^2}} \tag{6}$$

- The integral correlation coefficient between malware $w1$ and malware $w2$:

$$X_{w1,w2} = X^d_{w1,w2} + X^h_{w1,w2} \tag{7}$$

- The deviation of hourly downloads of malware w from its average:

$$\Delta l_w(h) = l_w(h) - \bar{l}^h_w \tag{8}$$

- Maximum deviation of hourly download of malware w:

$$\Delta l^h_{w,max} = max \mid \Delta l_w(h) \mid \tag{9}$$

- The normalized hourly download deviation of malware w:

$$\Delta l'_w(h) = \frac{\Delta l_w(h)}{\Delta l^h_{w,max}} \tag{10}$$

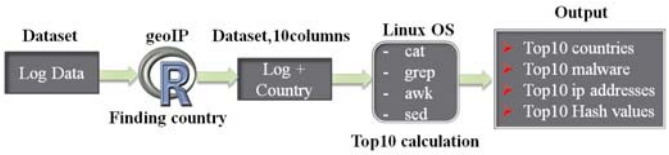


Fig. 1: Top-10 processing

C. Top-10 Processing

The dataset contains millions of malware download records uniquely identified by Timestamp, Source IP address/port number, Destination IP address (Honeypot ID)/port number, Hash value (SHA1), Malware name, and Malware file name. Based on a huge amount of logged information of malware/bot download activities, we can explore and enumerate the total downloads of malware w on day d, $l_w(d)$, and the total downloads of malware w at hour h, $l_w(h)$, using some powerful tools such as *R application (R is a free software environment for statistical computing and graphics)* and *cat, grep, awk, sed* commands as shown in Figure 1 to obtain Top-10 malware, Top-10 countries, and so on as follows.

- **geoIP:** Implements a lookup using the InfoDB API to determine point coordinates for a given IP address. We can define, *geoIP(x),* x is the IP address in IPv4 format. A list is returned containing the following elements:

IPaddress	Input IP address
statusCode	Returned status code from lookup
latitude	Point coordinate - latitude
longitude	Point coordinate - longitude
statusMessage	Returned status message from lookup
countryCode	Country code from IP lookup
countryName	Country name from IP lookup
regionName	State/region/province from IP lookup
cityName	City from IP lookup
zipCode	Postal code from IP lookup
timeZone	Timezone from IP lookup

In our work, we import initial logs in to R, and use geoIP function to lookup the *"countryName"* which matches the malware sources IP address. By geoIP processing, we try to insert any obtained countries to the existing logs data again in an extra column. Furthermore, we use the processed logs data for the next processing as shown in Figure 1.

- **cat:** It is a Unix and Linux command, concatenate FILE(s), or standard input. "cat" program is given files in a sequence as arguments, it will output their contents to the standard output in the same sequence. To extract any patterns of malware from the entire logs data 2010 and 2011, we used *cat* command to merge all monthly logs data of each year into single file and continue next step by using *awk* command.

- **awk:** The AWK language is useful for manipulating data files, text retrieval and processing, and for prototyping and experimenting with algorithms. An AWK program is a sequence of pattern action pairs and function definitions. To archive the Top-10, we locate to the logs data achieved from *geoIP and cat* . We use this command to *retract and count all unique malware names, IP addresses, countries, and etc.,* including number duplicated of them. From this we can use another sub-command such as *sort, head*, and especially *grep* to arrange the obtained output into Top-10 malware.

- **grep:** grep searches the input file name(s) for lines containing a match to the given PATTERN. By default, grep prints the matching lines. Hence, We used this command for 2 main purposes: First, search for quantities of unique malware containing in logs data. Second, create and save the greped lines into a new specific file for each Top-10 malware. Furthermore, we combine this command with others such as "*sed*" III-C to correct the data format before computing in R Project.

- **sed:** A stream editor is used to perform basic text transformations on an input stream (a file or input from a pipeline). We use *sed* to filter text, and symbol in a pipeline which particularly distinguishes of some output files.

IV. Results and Discussions

To finally cluster the Top-10 malware/bots, we gradually present the daily and hourly download behaviors in the following subsections. The integral correlations between them are summarized and eventually thresholded to a number of malware/bots clusters.

A. Daily Downloads of Malware/Bot

A botnet can exploit millions of compromised computers to perform malicious activities including malware/bot downloads in different periods of year. From our observation, Top-10 malware files are heavily downloaded at the beginning of 2010 and 2011 Japanese fiscal year as shown in Figure 2 and Figure 3, respectively. The Top-10 daily downloads, $l_w(d)$ of malware w are ordered according to their causalities and similarities. It can be observed that some of the reveal similar behaviors and probably dependencies in time domain.

B. Hourly Downloads of Malware/Bot

Based on Japanese local time of all honeypots, Figure 4 and Figure 5 show the normalized deviation of Top-10 malware hourly downloads from its average, $\Delta l'_w(h)$ in 2010 and 2011, respectively. Most malware behaviors in 2010 are similar to one another. On the other hand, the behaviors in 2011 are quite different from those of 2010 that no obvious sequences and patterns can be detected. This requires further investigations.

The average hourly downloads of malware w, $\bar{l}^h_w$ in 2010 and 2011 can be obtained from Tables II and III, respectively. It can also be observed that average hourly downloads, $\bar{l}^h_w$ of 2011 are much lower than those of 2010. This is due to shorter period (9 months) of logs and fewer (70) honeypots than those of 2010. In addition, there are fewer unique malware names in 2011 than 2010.

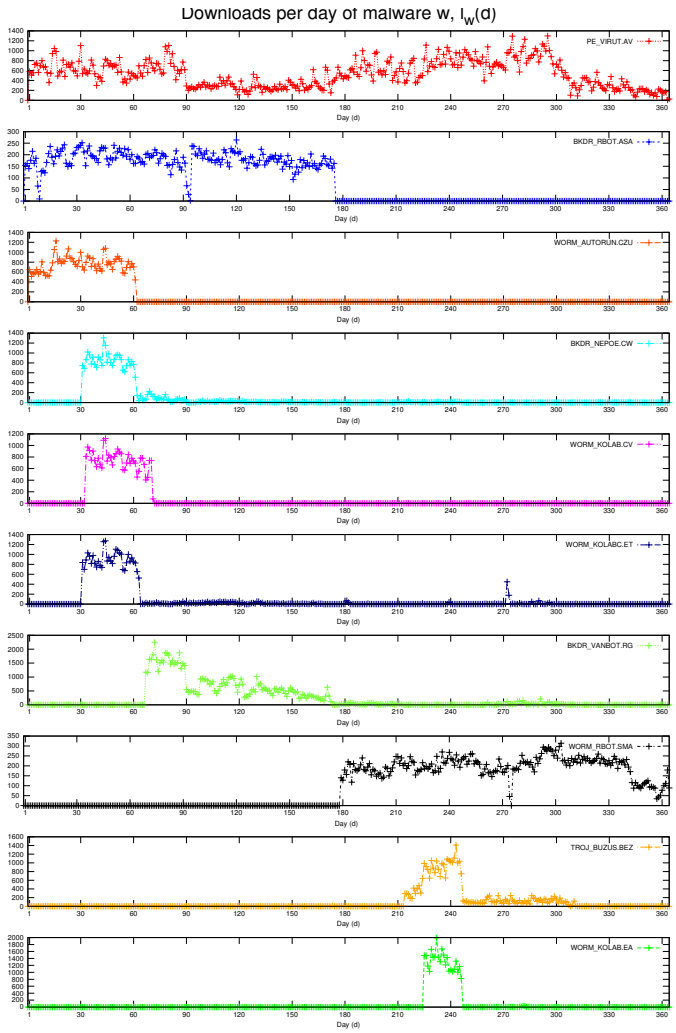


Fig. 2: Number of malware downloads per day, $l_w(d)$, in 2010 where w= PE_VIRUT.AV, BKDR_RBOT.ASA, WORM_AUTORUN.CZU, BKDR_NEPOE.CW, WORM_KOLAB.CV, WORM_KOLABC.ET, BKDR_VANBOT.RG, WORM_RBOT.SMA, TROJ_BUZUS.BEZ, WORM_KOLAB.EA, respectively.

TABLE II: Average Hourly Downloads of Top-10 malware in 2010 CCC datasets

Top	2010	
No	Malware	$\bar{l}_w^h$
1	PE_VIRUT.AV	8,106
2	BKDR_VANBOT.RG	3,489
3	WORM_AUTORUN.CZU	1,929
4	WORM_RBOT.SMA	1,507
5	TROJ_BUZUS.BEZ	1,340
6	WORM_KOLABC.ET	1,331
7	BKDR_RBOT.ASA	1,308
8	BKDR_NEPOE.CW	1,254
9	WORM_KOLAB.EA	1,204
10	WORM_KOLAB.CV	1,191

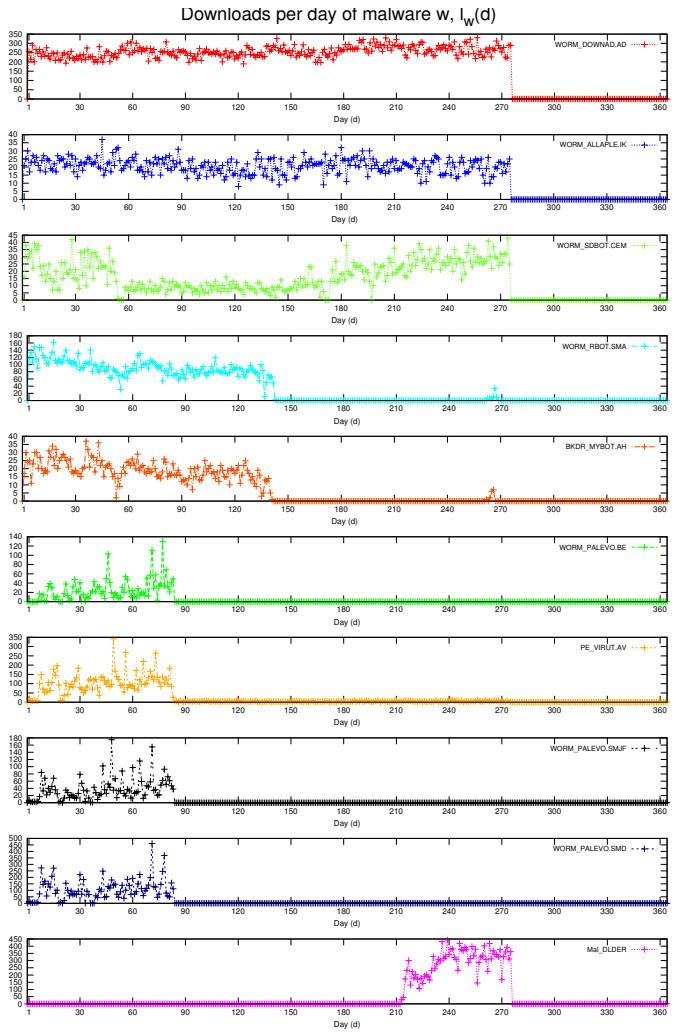


Fig. 3: Number of malware downloads per day, $l_w(d)$, in 2011 where w= WORM_DOWNAD.AD, WORM_ALLAPLE.IK, WORM_SDBOT.CEM, WORM_RBOT.SMA, BKDR_MYBOT.AH, WORM_PALEVO.BE, PE_VIRUT.AV, WORM_PALEVO.SMJF, WORM_PALEVO.SMD, Mal_DLDER, respectively.

TABLE III: Average Hourly Downloads of Top-10 malware in 2011 CCC datasets

Top	2011(9months)	
No	Malware	$\bar{l}_w^h$
1	WORM_DOWNAD.AD	2,919
2	Mal_DLDER	754
3	WORM_RBOT.SMA	532
4	PE_VIRUT.AV	390
5	WORM_PALEVO.SMD	366
6	WORM_ALLAPLE.IK	237
7	WORM_SDBOT.CEM	197
8	WORM_PALEVO.SMJF	130
9	BKDR_MYBOT.AH	113
10	WORM_PALEVO.BE	84

C. Clustering Top-10 malware/bots

Base on the daily and hourly download behaviors of Top-10 malware/bots in 2010 and 2011 as shown in the previous section. We can compute ***daily*** *and* ***hourly*** correlation coefficients, $X^d_{w1,w2}$ and $X^h_{w1,w2}$ using Equations 3 and 6, respectively. The integral correlation coefficient, $X_{w1,w2}$, defined in Eq.7

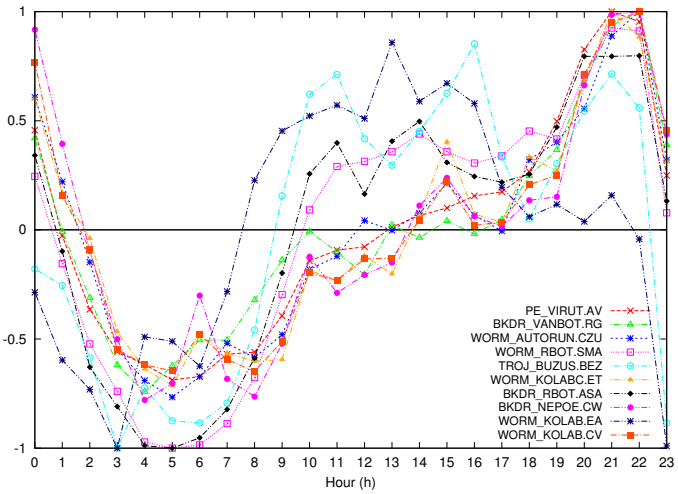


Fig. 4: Normalized deviation of hourly downloads from its average of Top-10 malware, $\Delta l'_w(h)$ in 2010.

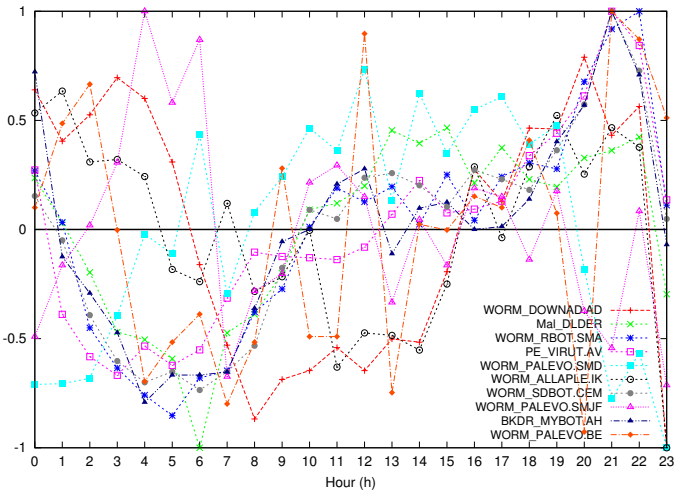


Fig. 5: Normalized deviation of hourly downloads from its average of Top-10 malware, $\Delta l'_w(h)$ in 2011.

between Top-10 malware/bots are listed in Tables IV and V. '

We have set $X^{th}_{w1,w2} = \mathbf{1.225}$ to be the threshold using human heuristics. The condition is that each malware in the same cluster must be individually correlated to one another. As a result, Top-10 malware of 2010 and 2011 can be clustered into 3 and 4 clusters, respectively as shown in Table VI.

V. Conclusion and Future Work

This paper has clustered Top-10 malware/bots based on their daily and hourly download behaviors in 2010 and 2011 CCC datasets. Due to the similarities and differences in daily and hourly download behaviors, the integral correlation coefficient, $X_{w1,w2}$, among Top-10 malware/bots is defined as

TABLE IV: Integral correlation coefficients $X_{w1,w2}$ of daily and hourly download for Top-10 malware/bot downloads in 2010

Top	2	3	4	5	6
1	0.910	1.157	**1.225**	1.008	1.028
2		0.720	0.414	0.409	0.785
3			0.483	0.430	**1.630**
4				**1.246**	0.547
5					0.421
Top		7	8	9	10
1		0.740	0.995	0.529	1.050
2		**1.374**	0.832	0.105	0.865
3		**1.373**	**1.617**	0.089	**1.560**
4		0.117	0.478	0.817	0.517
5		0.505	0.358	**1.753**	0.379
6		1.220	**1.947**	0.034	**1.878**
7			1.210	0.349	1.224
8				-0.030	**1.890**
9					-0.001

TABLE V: Integral correlation coefficients $X_{w1,w2}$ of daily and hourly download for Top-10 malware/bot downloads in 2011

Top	2	3	4	5	6
1	0.395	0.502	0.378	-0.118	**1.716**
2		0.537	0.570	-0.014	0.298
3			**1.550**	0.476	0.582
4				0.654	0.485
5					-0.046
Top		7	8	9	10
1		0.807	0.412	0.542	0.318
2		**1.427**	-0.564	0.493	0.161
3		1.086	0.058	**1.883**	0.893
4		1.080	0.184	**1.530**	1.002
5		0.075	**1.343**	0.430	0.472
6		0.767	0.282	0.649	0.424
7			-0.359	1.052	0.505
8				0.076	0.462
9					0.947

TABLE VI: Clusters of Top-10 malware/bots in 2010 and 2011 CCC datasets based on threshold $X^{th}_{w1,w2}$=1.225

2010		2011	
Cluster	Top	Cluster	Top
1	1,4	1	1,6
2	5,9	2	2,7
3	3,6,8,10	3	3,4,9
		4	5,8

the summation of daily, $X^d_{w1,w2}$ and hourly, $X^h_{w1,w2}$ correlation coefficients. Three and four clusters of highly correlated malware/bots can be identified by setting the threshold, $X^{th}_{w1,w2}$=1.225 in 2010 and 2011, respectively.

As future work, the common source IP addresses/countries can be identified for each cluster. As a result, their time zones can be investigated further in addition to their prefixes, names, file names, and hash (SHA1) values.

REFERENCES

[1] M.Hatada, Y.Nakatsuru, M.Akiyama, and S.Miwa, "Datasets for anti-malware research," *IPSJ anti Malware engineering Workshop 2010 (MWS2010)*, 2010.

[2] N.R.Rosyid, M.Ohrui, H.Kikuchi, P.Sooraksa, and M.Terada, "A discovery of sequential attack patterns of malware in bontents," *IEEE International Conference on Systems Man and Cybernetics (SMC)*, vol. Vol.2010-CSEC-48, No.37, pp. pp.2564– 2570, October 2010.

[3] J.Song, J.Shimamura, M.Eto, D.Inoue, and K.Nakao, "Correlation analysis between spamming botnets and malware infected hosts," in *IEEE/IPSJ 11th International Symposium on Applications and the Internet (SAINT)*, pp. pp.372 – 375, July 2011.

[4] K.Sisaat, H.Kikuchi, S.Matsuo, M.Terada, M.Fujiwara, and S.Kittitornkun, "Time zone correlation analysis of malware/bot downloads," *To be published in IEICE Transactions on Communications*, vol. Vol.E96-B, No.07, 2013.

Comparative Study of Attribute Reduction on Arrhythmia Classification Dataset

Anugerah Galang Persada, Noor Akhmad Setiawan, Hanung Adi Nugroho
Jurusan Teknik Elektro dan Teknologi Informasi
Fakultas Teknik Universitas Gadjah Mada
Yogyakarta, Indonesia
persada.ag@gmail.com

***Abstract*—The research presented in this paper is focused on comparative study of various attribute selections as one of pre-processing methods used in world machine learning applications. Using UCI arrhythmia dataset, nine combination of attribute selection, based on search methods (Best First, Genetic Search and PSO Search) and attribute evaluator (CfsSubsetEval, ConsistencySubsetEval, and RSARSubsetEval) are tested and compared. Those data of attribute reduction results are then classified by using eight classifiers (Naive Bayes, Bayes Net, MLP Classifier, RBF Classifier, Jrip, PART, J48 and Random Forest). The best overall results are achieved by the combination of Best First and CsfSubsetEval which has the accuracy of 81% when it is tested with RBF Classifier. PSO Search methods was also found not very effective to generate high quality subsets.**

Keywords—Comparative study; attribute reduction; classification; UCI Arrhythmia; WEKA

I. INTRODUCTION

Nowadays, there are a lot of diseases that caused death around the world. One of the most popular global killer is heart disease. Heart disease caused a lot of death in recent years. One of heart disease is caused by abnormal heartbeat. The heartbeat is stimulated by electrical signal which basically originated from the Sino Atrial node (SA), known as heart's natural pacemaker, and located at the top of the right chamber of Atrium (RA) in the heart. Any disruption of this pacemaker may lead heart to beat at an abnormal rate, thereby adversely impacting the circulation of the blood throughout the body. Arryhtmia is a kind of conduction system disease that causes an inefficient heart beat. There are several types of arrhtyhmia, which some of them could cause death of patients. Early detection and correct medical treatment of disease pertaining to hearth can increase probability of the patients and be saved from sudden death. [1]

There are many ways to early detect abnormal behaviour of heart. One of them is using expert system. The use of expert system as a solution to the needs of early detection of disease has been discussed more frequently the last decade. The expert system work by learning from the expertise, extract the knowledge using apropriate method and store the list of important knowledge in the database.

Machine learning can be used to expert system. Commonly, the data used to build the system need to be retouched, such as selecting the useful and important attributes of the data. After less important attributes have been removed, the data can be used to build the system effectively.

This works is focused on comparison between three atribute evaluator (CfsSubsetEval[2], ConsistencySubsetEval[3], RSARSubsetEval[4]) and and three search methods (Best First, Genetic Search[5] and PSO Search[6]). After the data have been reduced based on attribute selection combination (evaluator and search method), classification is performed using different classifier.

The classifier used in this research are bayes net, naive bayes, MLP Classifier, RBF Classifier, Jrip[7], PART[8], J48[9] and Random Forest[10]. All of the proccesses were done by using WEKA software[11]. Those evaluator was chosen based on work done by Selvakuberan, et al [12], which shows that different approach to reduce data may affect the classification result.

II. DATASET

The dataset used in this research is taken from UCI Machine Learning Repository, which is available in the internet [13]. The original aim of this dataset is to distinguish between the presence and absence of cardiac arrhythmia disease. If the disease found, the system shound able to determine which one is similar to one of sixteen group available. The dataset contains 279 attributes, 206 of which are linear valued and the rest are nominal. The detail information about 279 attribute can be found from UCI Arrhythmia dataset website.

Concerning the study of Guvenir: "The aim is to distinguish between the presence and absence of cardiac arrhythmia and to classify it in one of the 16 groups. Class 01 refers to 'normal' ECG classes 02 to 15 refers to different classes of arrhythmia and class 16 refers to the rest of unclassified ones. For the time being, there exists a computer program that makes such a classification. However there are differences between the cardiologist's and the programs classification. Taking the cardiologist's as a gold standard, we aim to minimize this difference by means of machine learning tools. To maintain confidentiality, patient ID and names were recently removed from dataset.

III. RESEARCH METHODOLOGY

This research divided into two sections. The first section is the data preprocessing by reducing the attributes using attribute selection. The other section is classification when the reduced dataset us classified using selected classifier. After all of preprocessing methods and classifiers combination has been tested, the result are compared to each other.

Attribute selection, in WEKA, consist of two parts. The first one is the attribute evaluator. Attribute evaluator are methods used to evalutate the data subset or each data attribute. This research focus on data subset evaluation methods, and used three of them, which are CfsSubsetEval, ConsistencySubsetEval and the last one is RSARSubsetEval. Detail parameters used in this research are shown below.

Table 1. Attribute evaluator chosen parameter

Methods	Options	Chosen Parameter
Cfs Subset Eval	locallyPredictive	True
	missingSeparate	False
Consistency Subset Eval	No editable properties	-
RSAR Subset Eval	No editable properties	-

Table 2. Attribute search chosen parameter.

Methods	Options	Choosen Parameter
Best First	Direction	Forward
	LookupCacheSize	1
	SearchTermination	5
	startSet	*NULL*
Genetic Search	crossoverProb	0.6
	maxGenerations	20
	mutationProb	0.033
	populationSize	20
	reportFrequency	20
	seed	1
	startSet	*NULL*
PSOSearch	individualWeight	0.34
	inertiaWeight	0.33
	Iterations	20
	logFile	Weka-3-7
	mutationProb	0.01
	mutationType	Bit-flip
	populationSize	20
	reportFrequency	20
	seed	1
	socialWeight	0.33
	startSet	NULL

Next part is attribute search methods. These methods search on original data, create new subset or make a rank of data attribute that is scored using evaluation methods. Because this research is focused on the subset methods, thus so the search methods that is able to create subset dataset are chosen. In this research, three methods are chosen, which are Best First, Genetic Search, and PSO Search.

After deciding which methods that are used, all of them are tested using all combinations. Each attribute evaluator is combined with all of search algorithms. Finally, nine different reduced attribute datasets ready to be classified.

IV. RESULT & DISCUSSION

The result, both from original and manipulated data, is shown below. To simplify and increase readability, the classifier's name are coded as follow :

- BN → Bayes Net
- NB → Naive Bayes
- MLP → Multi Layer Perceptron
- RBF → Radial Basis Function
- RF → Random Forest

A. *Classification using original data*

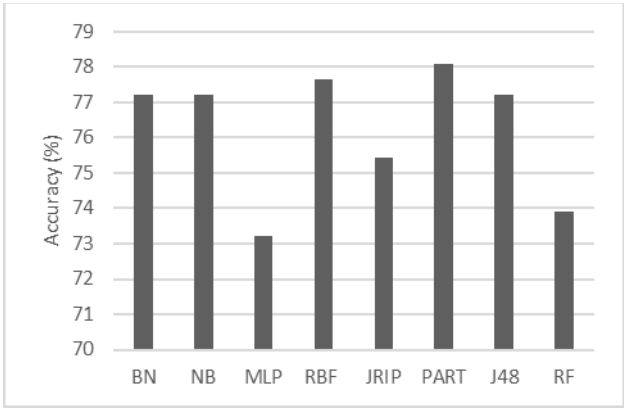


Fig. 1. Classification accuracy using original data

Using the original data acquired from UCI, classification result is shown in Figure 1. The best accuracy obtained is about 78%, using PART, and the worst accuracy coming from MLP classifier, only around 73% Average accuracy rate of classifier is around 75%. It means that there are still a lot spaces need to improve.

B. Best First & CfsSubsetEval

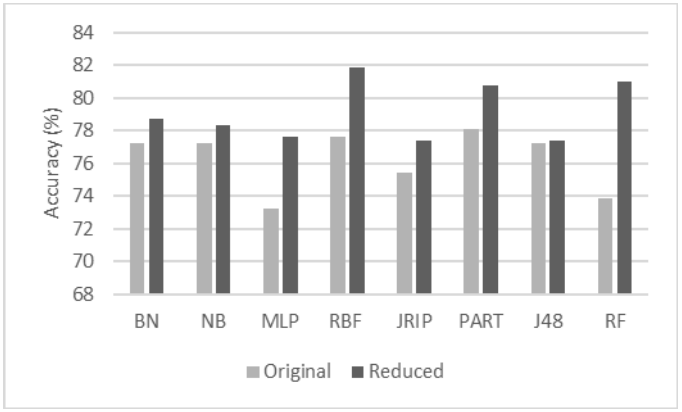


Fig. 2. Classification accuracy using Best First and CfsSubsetEval attribute reduction method

Using first combination, overall the Best First and CfsSubsetEval attribute evaluator, has better result than that of using original data. Random forest has the most significant improvement, with the increase of almost 7%. Therefore, the best result comes from RFB Classifier with the accuracy up to 82%.

C. Best First & ConsistencySubsetEval

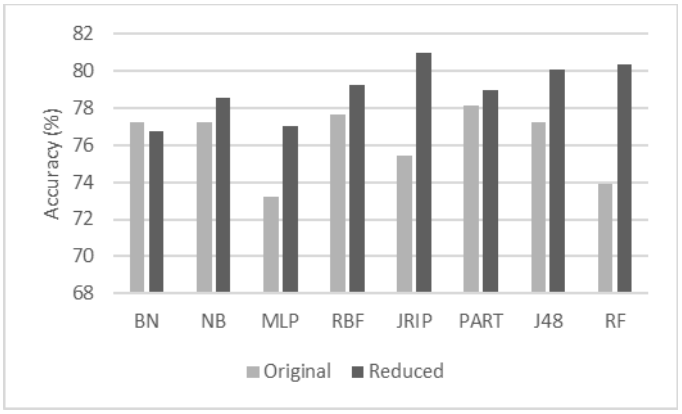


Fig. 3. Classification accuracy using Best First and ConsistencySubsetEval

Next combination that used in this work is Best First and ConsistencySubsetEval attribute evaluator. Almost all of selected classifier have increase of accuracy caused by data reduction. The best accuracy belongs to Jrip methods (80.9%), followed by Random Forest (80.3%) and J48 (80.1%). The Bayes Net, on the other hand, has worse result, which the accuracy decreases 0.5% compared to original data classification.

D. Best First & RSARSubsetEval

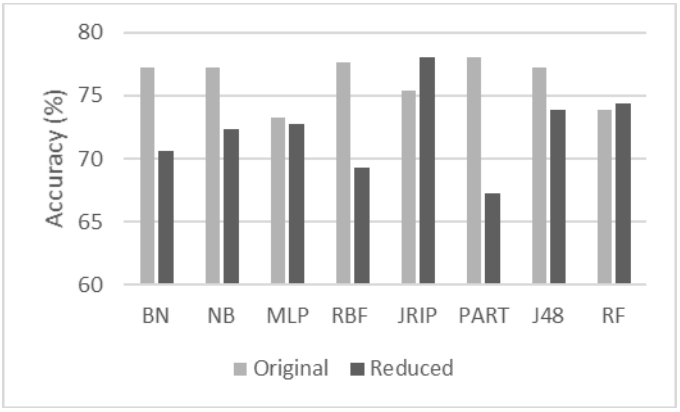


Fig. 4. Classification accuracy using Best First and RSARSubsetEval

Fig 4 shows the result of classification obtained by the combination of Best First and RSARSubsetEval data reduction. However, most of these results are not good enough. Almost all classifiers accuracy are decreased. PART algorithm is the classifier that has the worst result, in which the accuracy declines to almost 11%. Jrip and Random Forest still have accuracy increased, although it is not significant. The result shows that the combination of Best First and RSARSubsetEval, which is based rough set theory, is not good enough to select the proper attribute to increse classification accuracy.

E. Genetic Search & CfsSubsetEval

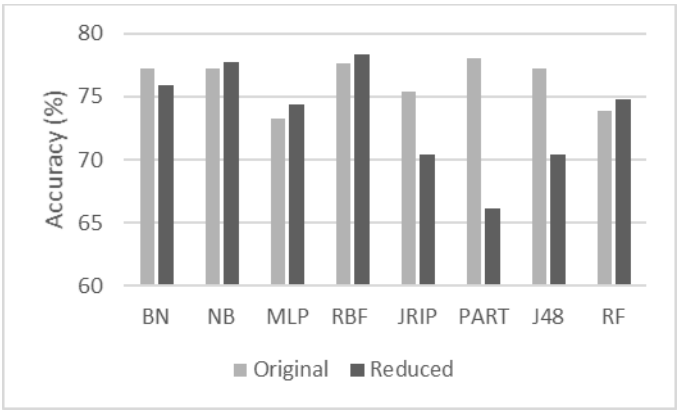


Fig. 5. Classification accuracy using combination of Genetic Search and CfsSubsetEval.

Figure 5 shows result of classification using data reduced by collaboration of Genetic Search and CfsSubsetEval. Overall result is not good. There are no significant improvement in classification accuracy. Some of them, the accuracy are getting worse. The best result obtained by RBF Classifier, in which the accuracy increses from 77.6% to 78.3% using original data.

F. Genetic Search & ConsistencySubsetEval

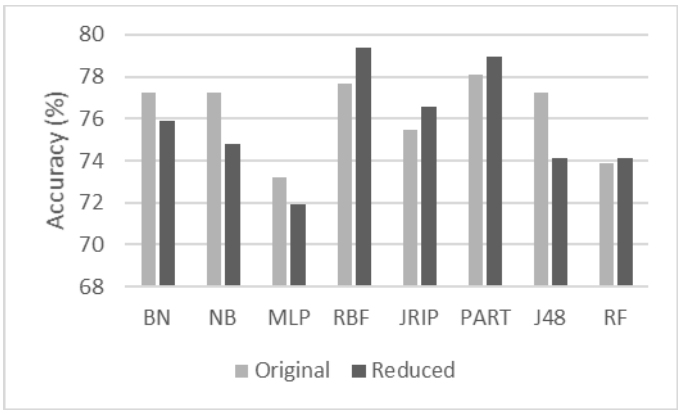


Fig. 6. Classification accuracy using Genetic Search and ConsistencySubsetEval

The next is combination of Genetic Search and ConsistencySubsetEval attribute evaluator. As shown on Figure 6, half of classifier has better accuracy with reduced data, but the others create worse accuracy. Overall, the result shows that this methods, the data reduction, is not very effective to select proper data that able to give important information to classifier.

Using this combination, the best result comes from RFB, with almost 2% increase of accuracy. Second best result is obtained by PART, in which the accuracy increses from 78% to 78,9%. The worst performance belongs to J48. Its accuracy drops from 77,2% to 74,1%, followed by Naive Bayes in which the accuracy decrease from 77,2% to 74.7%

G. Genetic Search & RSARSubsetEval

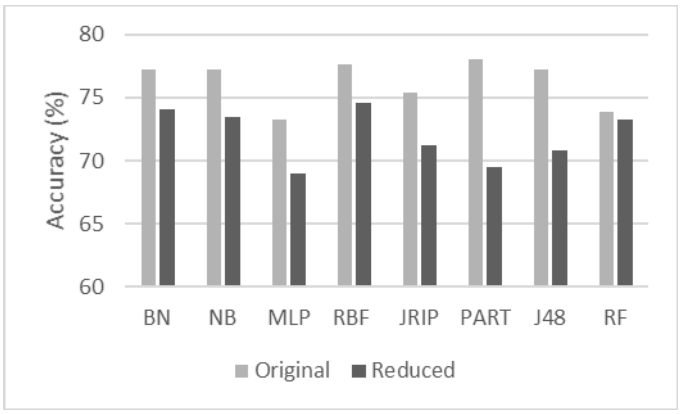


Fig. 7. Classification accuracy using Genetic Search and RSARSubsetEval.

Figure 7 shows the classification obtained by using reduced data that are produced by RSARSubsetEval and Genetic Search. As shown in Figure 7, the result is far from the expected. No classifier gets better accuracy. The most significant accuracy drops comes from PART algorithm. Its accuracy decreases from 78.1% to 69.4%, a quite large interval. The best accuracy, although all of them decresed, belongs to the RBF classifier, with 74.6%, instead of 77.6% if the original data are used.

H. PSO Search & CfsSubsetEval

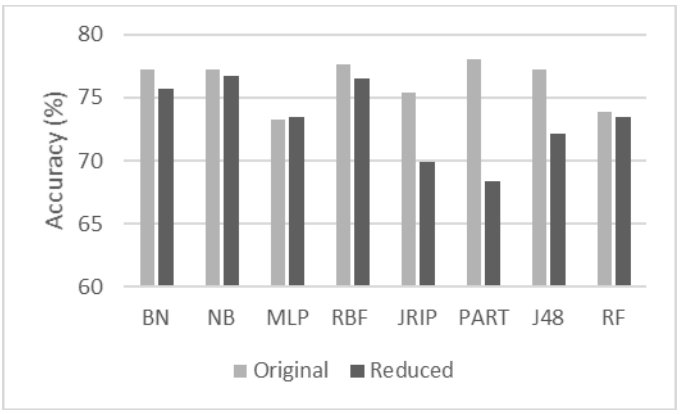


Fig. 8. Classification accuracy using PSO Search and CfsSubsetEval.

Figure 8 shows the classification accuracy result from reduced dataset using PSO and CfsSubsetEval. The results are poor, no classifier get better accuracy after using the reduced data. Even more, PART algorithm has the worst accuracy, which its accuracy is decreased by 10%, from 78.1% to 68.3%. The other classifier also have bad result, for example Jrip algorithm result decline around 5%. J48 methods result also drops almost 5%. From these results, it can be concluded that the use of PSO Search combined with CfsSubsetEval is not effective.

I. PSO Search & ConsistencySubsetEval

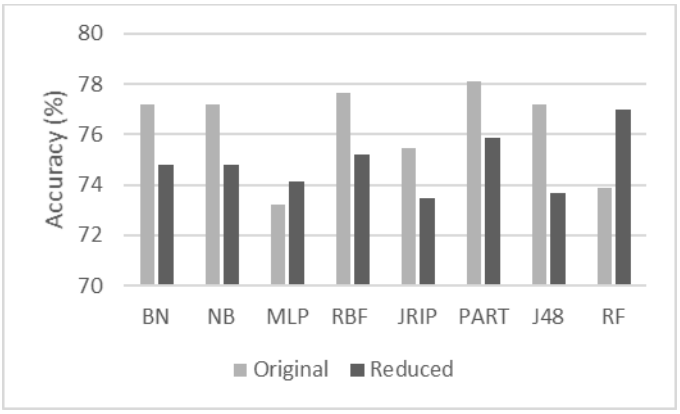


Fig. 9. Classification accuracy result using PSO Search and ConsistencySubsetEval.

As shown in Figure 9, the combination of PSO Search and ConsistencySubsetEval, is not effective. There is only one classifier, Random Forest, that is able to increase its accuracy from 73.8% to76.9%. The others failed to improve their accuracy, which is only up to near 5%. This result shows that the use of PSO Search combined with ConsistencySubsetEval as attribute selection procedure is not effective.

J. PSO Search & RSARSubsetEval

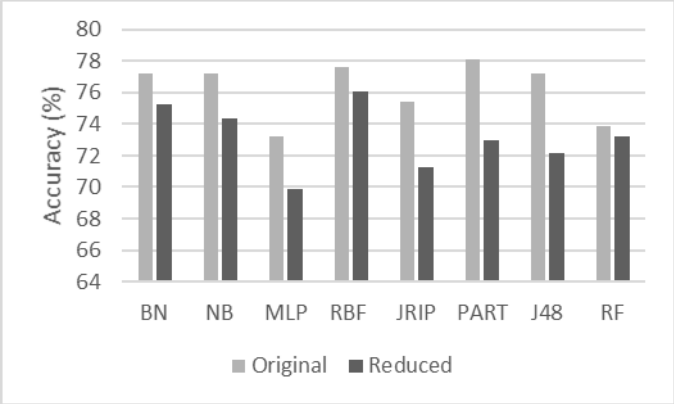


Fig. 10. The classification accuracy using reduced data generated by PSO Search and RSARSubsetEval.

The last combination of methods used in this paper is PSO Search and RSARSubsetEval. As ilustrated in Figure 10, the result shows that all selected classifiers have poor performance after using reduced data. RFB Classifier has the most minimal negavite effort. Meanwhile the Jrip and PART have large deficit of accuracy. RBF classifier can reach the accuracy of 76%, and put the RBF as the best classifier using this reduced dataset.

K. Result Summary

Table 3. Summary of Best Performance of each Classifier

Attribute Selection		Classifier	Accuracy
Attribute Search	Attribute Evaluator		
Best First	CfsSubsetEval	BN	78,76106
Best First	ConsistencySubsetEval	J48	80,0885
Best First	ConsistencySubsetEval	JRIP	80,97345
Best First	CfsSubsetEval	MLP	77,65487
Best First	ConsistencySubsetEval	NB	78,53982
Best First	CfsSubsetEval	PART	80,75221
Genetic Search	ConsistencySubsetEval	RBF	79,41176
Best First	CfsSubsetEval	RF	80,97345

As a summary, each best performance on each classifier shown in Table 3. Table 3 shows that the most effective attribute search belongs to the Best First Search methods, and the attribute evaluator has two dominant methods, i.e. the CfsSubSetEval and ConsistencySubSetEval.

V. Conclusion & Future Work

From all results of classification, it can be concluded that, PSO Search with its default parameter is not effective to search and create subset for evaluatoin. The most significant attribute search is Best First search, in addition, effective attribute evaluator are CfsSubsetEval and ConsistencySubsetEval.

The next works is focused on the adjustment parameter from the attribute selection and the classifier. This adjustment may lead to the best combination between search subset methods and subset attributes' selector.

References

[1] Y. Ozbay and B. Karlik, "A recognition of ECG arrhytihemias using artificial neural networks," in *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, 2001, vol. 2, pp. 1680–1683 vol.2.

[2] M. A. Hall, "Correlation-based Feature Selection for Machine Learning," no. April, 1999.

[3] H. Liu and R. Setiono, "A Probabilistic Approach to Feature Selection , A Filter Solution," 1993.

[4] A. Chouchoulas and Q. Shen, "Division of Informatics , University of Edinburgh Rough Set-Aided Keyword Reduction for Text Categorisation," no. May, 2001.

[5] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1989.

[6] A. Moraglio, C. Di Chio, and R. Poli, "Geometric Particle Swarm Optimisation," pp. 125–136, 2007.

[7] M. Avenue, M. Hill, W. W. Cohen, C. Of, and R. Pruning, "Fast E ective Rule Induction 2 PREVIOUS WORK 1 INTRODUCTION," 1994.

[8] E. Frank and I. H. Witten, "Generating Accurate Rule Sets Without Global Optimization," 1998, pp. 144–151.

[9] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.

[10] L. E. O. Breiman, "Random Forests," pp. 5–32, 2001.

[11] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The {WEKA} Data Mining Software: An Update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.

[12] K. Selvakuberan, M. Indradevi, and R. Rajaram, "Combined Feature Selection and classification – A novel approach for the categorization of web pages," vol. 3, no. 2, pp. 83–89, 2008.

[13] K. Bache and M. Lichman, "{UCI} Machine Learning Repository." 2013.

Computer Aided Diagnosis for Lung Tuberculosis Identification Based on Thoracic X-ray

(Preliminary result)

Ratnasari Nur Rohmah
Dept. of Electrical Eng. and Information Technology
University of Gadjah Mada, Yogyakarta, Indonesia
Dept. of Electrical Engineering
UMS, Surakarta, Indonesia

Adhi Susanto, Indah Soesanti
Dept. of Electrical Eng. and Information Technology
Maesadji Tjokronagoro
Dept. of Medical
University of Gadjah Mada, Yogyakarta, Indonesia

***Abstract*—This paper presents research on lung tuberculosis (TB) identification by using computer. This research was attempt to reduce patient waiting time in receiving X-ray diagnosis result on lung TB disease, due to mismatch ratio of radiologic experts to the number of patient, especially from remote areas in Indonesia. We used textural features calculated by computer to be used as descriptor in classifying image as TB or non-TB. We used statistical features of image histogram by calculates five features: mean, standar deviation (std), skewness, kurtosis, and entropy. These features were calculated from ROI images using pre defined ROI shape from *thresholding* method. Features calculated was then reduced down to one principal feature using Principal Componen Analysis (PCA) method. Finally, we used Mahalanobis distance classifier as classifier method based on one principal feature as descriptor. This research results show that it was possible to classify TB and non-TB image based on statistical feature on image histogram.**

Keywords—X-ray image; Lung Tuberculosis; textural features; PCA

I. INTRODUCTION

Rapid computer applications in medical area provide many advantages in healthcare application. Computer Aided Diagnosis (CAD) is one of those computer applications in medical field. The idea is using computer to assist radiologist in interpreting medical images by using medical analysis algorithm to provide 'second opinion' of some cases. This idea is related to the specific purposes of medical image that is one of diagnostic tools. In CAD, the computer performance does not have to be comparable to or better than physician diagnostic, rather than it is a complementary decision [1].

Thoracic X-Ray image commonly used by physician as diagnostic tool to find some diseases related to human lung, such as lung Tuberculosis (TB). Tuberculosis is a disease caused by an infection of *Myobacterium tuberculosis* [2]. It is a systemic disease that can strike any part of human organ with lung is primary infection location.

To make an X-Ray image, the patient will need several days before receiving the result. An X-Ray image need to be examined, diagnosed, and notated by radiologist before given to the patient. In Indonesia, the problem of uneven distribution of doctors is still a challenge [3]. Distribution of physicians, especially specialists, is also uneven in which more than two-thirds of specialists are in Java and Bali [3]. This condition makes the waiting time for patient to receive X-Ray images result relatively longer than those in advanced countries, especially for patient from remote areas in Indonesia.

This research was to develop CAD application to identify lung TB using computer. Identification using computer in this research was conduct using thoracic X-ray image of patient. The image classified as TB image and non-TB image, based on its textural feature of image. This application is expected to give benefit to TB lung patients to reduce the waiting time in obtaining the result.

II. UNDERLYING THEORY

Radiologic experts search specific features to diagnose Lung TB by examine X-ray image visually [2]. Imitate these work, this research calculated textural features based on statistical feature of image histogram. One of the simplest approaches for extract textural feature is to use statistical moments of the gray-level histogram of an image. Most common histogram features calculated in image features extraction are: *mean, standard deviation (std), skewness, kurtosis* and *entropy.* We calculated those five features and then select the most principal features to be used in image classification as discriminator.

III. MATERIALS AND METHODS

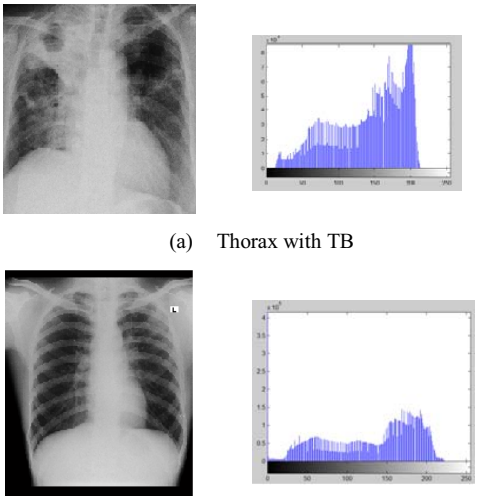
Data used in this research were copy of digital X-ray thoracic images from Dr.Sarjito Hospital Yogyakarta. We used 50 thoracic X-Ray images to design this application, 25 images without disease and 25 images with Lung TB. All images were proviously identified by radiologic experts. We also used 110 digital X-ray images to be used in systems performance test. Those were 45 images without diseases (normal), 45 lung TB indicated images, and 20 bronchitis indicated images.

A. *Pre-processing*

Raw data acquired from Radiology Unit of Sardjito Hospital were non-uniform data, in quality, size, and human body posed in X-Ray image. Fig. 1 shows data images and their histograms. Pre-processed was conduct before images feature extraction and conduct to reduce non-uniformities. Steps on this process were shown in Fig. 2.

Image pre-processing consisted of image quality enhancement and object isolation. Image quality enhancement was done using spatial filtering and histogram equalization. Spatial filtering was for noise reduction and pixel intensity transformation using histogram equalization was used to enhance image quality and make intensity level of all data in uniformity. Object isolation was the process of selecting a region of interest (ROI) from the image. ROI was the area of the image that its features will be measured. It was taken to address the non-uniformities in size and body pose of the data raw images.

Proper ROI image has an important role on calculating the best image feature for image classification. Getting ROI shapes in this experiment was done in three steps: cropping rectangular shape of thoracic image, resizing image and applying *thresholding* method to get ROI logical area. Fig. 3 shows process in getting ROI shape. We used normal image in this process to get ROI shapes as much similar as anatomic models of lung as shows in Fig. 4 [4]. The reason was that, TB lung affect the lung shape and in turn can defect the ROI uniformity in *thresholding* method result.



(a) Thorax with TB

(b) Thorax without TB

Fig. 1. Raw data images and their histograms

The best ROI shape from *thresholding* method result was then choosen to be used in image identification process. The prosedure in ROI selection were: (1) ploting image data based on descriptor for all ROI template used in image pre-processing step, (2) calculating Euclidean distance and Mahalanobis distances between TB class and normal class for each ROI, and (3) selecting best ROI by comparing Euclidean distance and Mahalanobis distances for each ROI. The ROI showed the longest distances between classes was chosen as the best ROI and will be used as ROI in the next process.

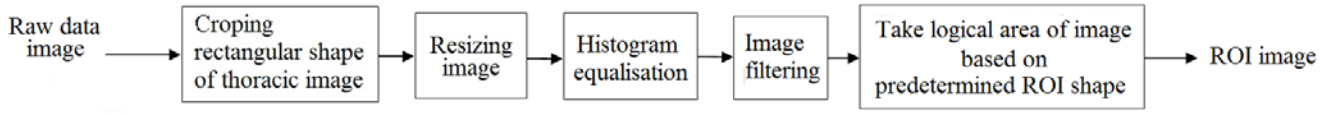


Fig. 2. Block diagram on pre-processing raw data images.

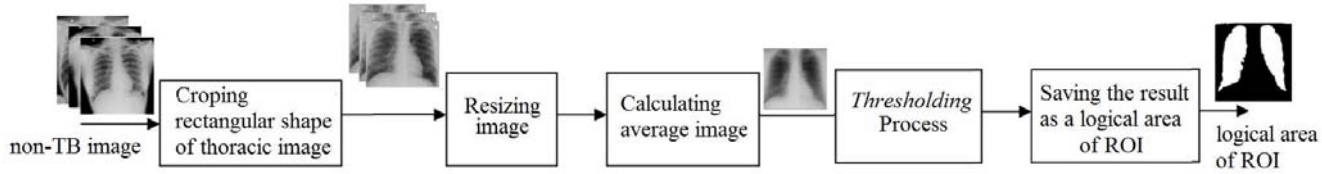


Fig. 3. Steps on image segmentation process to get ROI shapes.

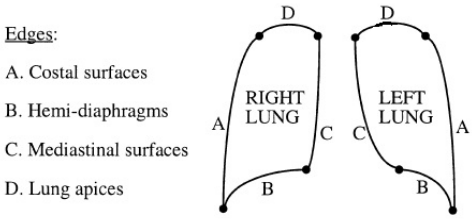


Fig. 4. Anatomic model of human lung[4].

B. Image features extraction

Five statistical textural image features were calculated from 50 image sample, that are *mean, standard deviation (std), skewness, kurtosis* and *entropy* [5]. From feature image calculation we have 5 dimension feature vectors. Feature dimension reduction and selection was then conduct to reduce feature vectors from 5 down to 1. We use Principal Component Analysis (PCA) method in this process to make sure that selected feature is the most principal feature on image [6]. This principal feature will be used as image descriptors. Fig. 5 shows the PCA transformation algorithm used in this research.

$$mean = \sum_{x=0}^{l-1} x \; p_u(x) \quad (1)$$

$$std = sqrt\left(\sum_{x=0}^{L-1}(x-m_1)^2 \, p_u(x)\right) \quad (2)$$

$$skewness = \sum_{x=0}^{L-1}(x-m_1)^3 \, p_u(x) \quad (3)$$

$$kurtosis = \mu_4 - 3 \quad (4)$$

$$entropy = -\sum_{x=0}^{L-1} p_u(x)\log_2 p_u \quad \text{bits} \quad (5)$$

Where:

$$p_u(x) \cong \frac{\text{the number of pixel with grey level } x}{\text{the number of pixel in an image}}$$

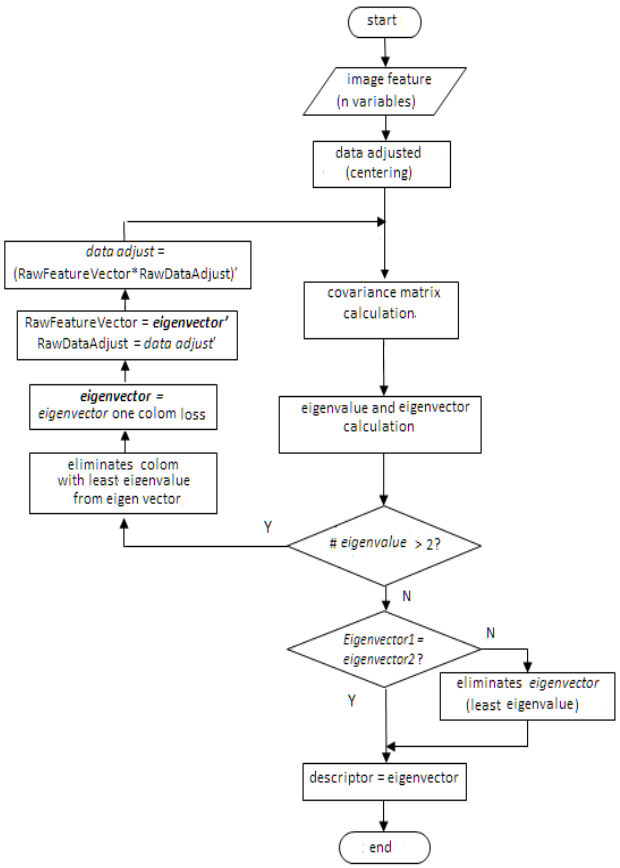


Fig. 5. Flowchart of PCA feature dimension reduction from n to 1.

C. Image identification

We use Mahalanobis distance classifier method in image identification application. Decision function for this classifier was made based on feature vector resulted from feature image extraction. Steps on image identification are shown in Fig. 6. Steps includes: pre-processing image, image feature calculation, PCA transformation, and image classification.

In this research, application performance was evaluated from its accuracy, fault acceptance rate (FAR), and fault rejection rate (FRR) [7].

$$\text{Accuracy} = \frac{\#\text{True identification}}{\#\text{Sampel}} \quad (6)$$

$$\text{FAR} = \frac{\#\text{Fault Accepted Elements}}{\#\text{Sampel}} \quad (7)$$

$$\text{FRR} = \frac{\#\text{Fault Rejected Elements}}{\#\text{Sampel}} \quad (8)$$

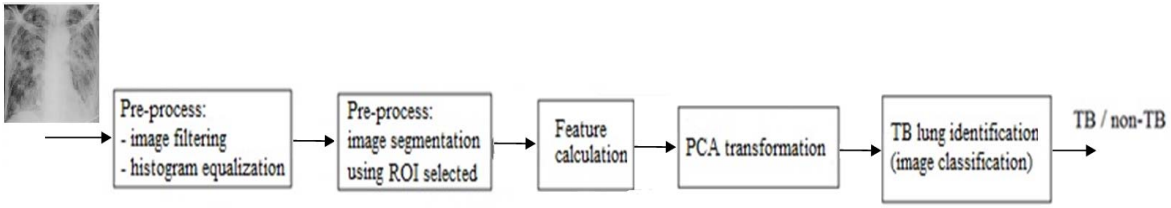


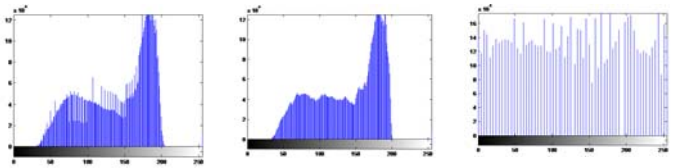
Fig. 6. Block diagram on image identification process.

IV. Results and Discussion

A. Pre-processing image

This process includes: filtering the image, adjusting the intensity of the image, and the image segmentation to obtain the ROI image. Filtering and histogram equalization results were shown in Fig. 7. Image histogram before the process was shown in Fig.7 (a). Fig. 7 (b) showed that the image filtering using *averaging* filters have reduced image noise by reducing spike on the image histogram. Fig. 7 (c) showed the histogram was evenly distributed and this will ensure uniformity in the intensity distribution of images samples.

Figure 8 showed three ROI shapes from our image segmentation process using threshold 120, 125 and 150. All of those three shapes were then tested by examining clustering images data on images plot based on their descriptors. We tried these three grey levels threshold values in *thresholdin*g process to get the optimum grey level threshold value. This optimum level will give the best ROI image to be used in image classification process.



(a) (b) (c)

Fig. 7. Result on pre-process: filtering and histogram equalization; (a) image histogram before pre-process, (b) image histogram after filtering, (c) image histogram after filtering and histogram equalization.

Fig. 8. Three ROI shapes resulted from image segmentation using thresholding method

B. Feature extraction

First experiment in image feature extraction was done by extract 50 raw data images, and using ROI1 as an ROI image. Five statistical features of image histogram were calculated in feature extraction: *mean, standar deviation (std), skewness, kurtosis,* and *entropy.* From this calculation we have feature vector matrix 50x5, with covariance matrix:

$$\begin{bmatrix} 242.7891 & 153.1577 & -1.201 & -3.9028 & -0.6807 \\ 153.1577 & 115.8762 & -0.2317 & -2.1057 & -0.3028 \\ -1.201 & -0.2317 & 0.0368 & 0.0633 & 0.0027 \\ -3.9028 & -2.1057 & 0.0633 & 0.1533 & 0.0071 \\ -0.6807 & -0.3028 & 0.0027 & 0.0071 & 0.0054 \end{bmatrix}$$

Those covariance matrix showed that all variables (calculated features) were correlated each other. Fig. 9 showed plots on images based on those five features. This plots showed that, agree with that condition, clustering in TB and non-TB images was still overlapping.

The nest experiment was transformation to reduce five dimension features vector down to one feature. This feature was the most principal feature and should be regarded as the descriptor of image. This transformation changed variable coordinates into a new one, and made a new uncorrelated variable data. From first transformation, we get new features vector matrix that has covariance matrix:

$$\begin{bmatrix} 0.0036 & 0 & 0 & 0 \\ 0 & 0.0986 & 0 & 0 \\ 0 & 0 & 13.5805 & 0 \\ 0 & 0 & 0 & 345.1773 \end{bmatrix}$$

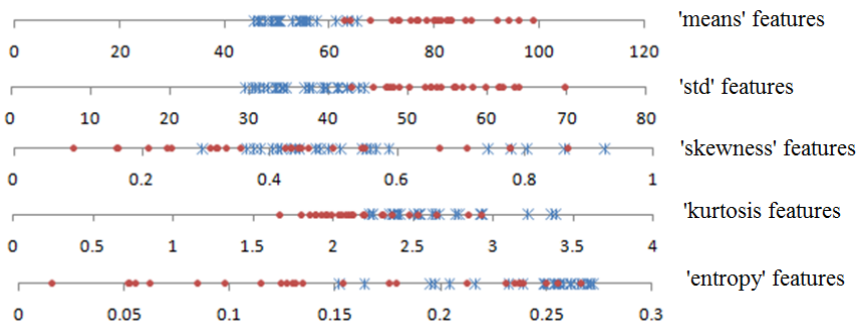


Fig. 9. Comparison of TB and non-TB image clustering on image plots based on each features (using ROI1 as ROI template).

This covariance matrix showed that data were not correlated each other anymore. This transformation proceeds until only one feature vector left. This is the most **principal feature** or **descriptor**. Fig.10 shows image plots based on this principal feature. From the plot we could see that images data split to two different clusters, without any overlapping data.

C. ROI selection

As we mentioned before, we used three ROI shapes resulted from *thresholding* method in pre-processing. Fig. 11 showed images plots based on one descriptor, for ROI1 and ROI2 images. Euclidean and Mahalanobis distance calculation results between two classes was shown in Table 1. This table showed that ROI1 gave the longest Euclidean distance, and ROI3 gave the longest Mahalanobis Distance. As we used Mahalanobis distance classifier for image identification, we chose ROI3 as ROI template for this application.

D. Image Identification

Images classification design in this experimen was done using ROI3 in pre-processing images. We used image feature from 50 images to build decision function. For system performance test, we used 110 test images, 45 images without disease, 20 images with bronchitis, and 45 images with Lung TB. Table II showed the system classification performance test. Test results were: accuracy: 95.4%, FAR: 4.5%, FRR = 4.5%.

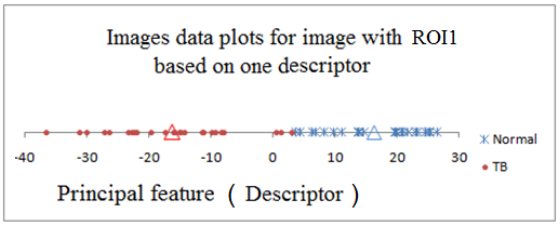


Fig. 10. Image data plot based on principal feature (descriptor) resulted from PCA transformation for image with ROI1 as ROI template.

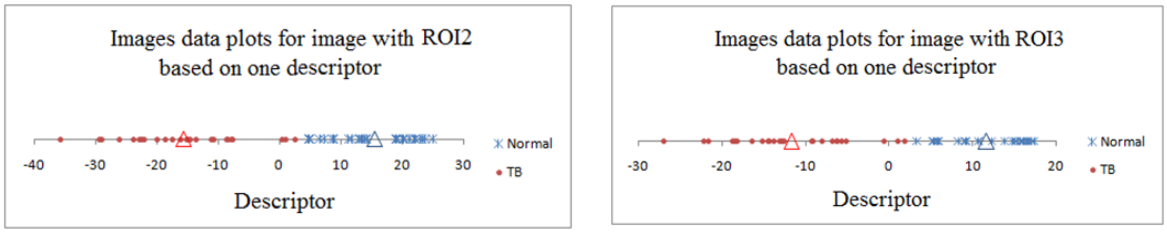


Fig. 11. Images data plot based on one descriptor, for each ROI2 and ROI3

TABLE I. COMPARISON ON EUCLIDEAN DISTANCE AND MAHALANOBIS DISTANCE FOR EACH ROI USED IN IMAGE SEGMENTATION

Distances	ROI1	ROI2	ROI3
Euclidean between mean descriptor two classes	32,5431	31,1842	23,2874
Mahalanobis distance between descriptor mean of TB class and all descriptors of normal class	21,4356	23,2359	25,2823
Mahalanobis distance between descriptor mean of normal class and all descriptors of TB class	10,2089	10,2994	10,088

TABLE II. RESULTS ON IMAGE IDENTIFICATION TEST

The number of test images	Identification results	
	TB positives	TB negatives
45 normal images	0	45
45 TB images	42	3
20 Bronchitis images	2	18

V. CONCLUSION AND FURTHER WORKS

We concludeed our research based on experiment and performance test result as follows:

1. Image pre-processing before feature extraction by mean of filtering, histogram equalization, and image segmentation, can be used to reduce non-uniformities on raw data images to improve images clustering.

2. Features dimension reduction using PCA assure that selected feature was the most principal feature of image.

3. The best ROI shape on image segmentation was ROI3 which was showed lung area with as few as possible of non-lung area.

4. The results on identification of 110 images tests in this research were: *accuracy: 95.4%, FAR: 4.5%, FRR = 4.5%.* These results showed the possibility in identifying lung TB automatically based on X-ray image textural features.

In the future, application should be complemented with the feature identification, especially for features that indicate a highly infectious stage of TB. This development should be supported by the availability of data, which is much larger than the number of data images used in this research, for each studied feature.

According to this research results, pre-processes that must be considered in further development include: handling non-uniformities of data, image enhancement, and selection of the image ROI in accordance with the characteristics of the image feature. Feature extraction process also should be expanded, in addition to textural features; other features must also be taken into account, for example, feature of shape.

REFERENCES

[1] Kunio Doi, “Computer-aided diagnosis in medical imaging: Historical review, current status and future potential ”. Computerized Medical Imaging and Graphics, No. 31, 2007, pp: 198-211.

[2] Pusponegoro, Hardiono D., “Standar Pelayanan Medis Kesehatan Anak”, first ed., Badan Penerbit IDAI, Indonesia, (2005).

[3] Research and Development Team Health Professional Education Quality (HPEQ) Project. “Potret Ketersediaan Dan Kebutuhan Tenaga Dokter”, Ditjen Dikti Kemendikbud, 2010. Available at: http://hpeq.dikti.go.id/v2/images/Produk/Potret_Ketersediaan_Dan_Kebutuhan_Tenaga_Dokter.pdf.

[4] Matthew S. Brown, at all, “Knowledge-based method for segmentation and analysis of lung boundaries in chest X-ray images”, Computerized Medical Imaging and Graphics, No. 22, 1998, 463-477.

[5] Gonzalez, R. C. and R. E. Woods, “Digital Image Processing”, third ed., Pearson Education, New Jersey, 2008.

[6] Partridge, M., & Calvo, R. A. 1998. “Fast dimensionality reduction and simple PCA”, Intelligent Data Analysis, 2(1-4), 1998, 203-214.

[7] Kachanubal, T., Udomhunsakul, S. 2008, Rock Textures Classification Based on Texturaland Spectral Features. World Academy of Science, Engineering and Technology 39, (2008) 110 – 116.

Content-Dependent Spatio-Temporal Video Watermarking using 3-Dimensional Discrete Cosine Transform

Iwan Setyawan[1], Ivanna K. Timotius[2]
Department of Electronic Engineering
Satya Wacana Christian University
Salatiga 50711, Indonesia
Email: [1]iwan.setyawan@ieee.org, [2]ivanna.timotius@ieee.org

***Abstract*—In this paper we propose a content-dependent spatio-temporal watermarking scheme for digital videos. Content dependency is achieved by incorporating the hash of the video sequence into the watermark. The video sequence is treated as a 3-dimensional spatio-temporal signal for the purposes of video hash computation and watermark embedding and detection. Our experiments show that the video hash algorithm has good discriminating power and robustness against various attacks. The watermark is also shown in the experiments to have good robustness against a variety of attacks, in particular when the watermark is copied from one video sequence to another.**

Keywords—digital video watermarking, content-dependent watermarking, copy attack, video hashing

I. INTRODUCTION

One of the biggest advantages of digital video is the ease with which video content can be reproduced without loss of quality. However, this particular advantage has also opened the possibilities for abuse, namely the unauthorized reproduction of content (illegal copying or piracy). Since reproduction is relatively easy and without quality loss, illegally copied content is both cheap and has comparable quality to the original. These features make such content very attractive to consumers. One solution to this problem is digital watermarking technology [1]. Using digital watermarking technology, a copyright owner can embed a perceptually undetectable signal into the digital video. When an illegal copy of this video is found, the copyright owner can assert his/her ownership using the digital watermark as evidence.

In embedding a digital watermark into a digital video, one can treat the video signal as a series of separate, 2-dimensional signals (i.e., treating each individual video frame separately). An example of this approach is proposed, for example, in [2]. However, one can also treat the video signal as having three dimensions (i.e., two spatial dimensions (width and height) and one temporal dimension). There are several proposed video watermarking techniques using this approach in the literature. For example, the authors in [3] proposed a video watermarking system based on 3-dimensional wavelet transform. Another example is the scheme proposed in [4], where the authors proposed a watermarking system based on 3-dimensional Fourier transform.

Since a digital watermark typically serves as a deterrent to illegal copying, it has to be robust against attempts (attacks) to remove it from the video. It may not be necessary to make the digital watermark impossible to remove. It is usually enough to force an attacker to reduce the quality of the video to an unacceptable level in order to remove the watermark (or render it undetectable). The approaches mentioned in the previous paragraph have been shown to have satisfactory performance against commonly encountered attacks, for example compression, filtering or temporal attacks such as random frame dropping [3][4]. However, there is another attack that is not aimed at watermark removal. Instead this attack is aimed at copying a watermark from one video onto another video. This attack—known as the copy attack [5]—is used for example to discredit someone by "proving" that a certain video belongs to him/her since it carries his/her watermark. The watermarking techniques presented in [3][4] are not robust against this type of attack. Robustness against the copy attack can be achieved by making the watermark signal *content dependent*, i.e., it must depend on the particular video into which it is embedded.

One way to achieve this goal is by incorporating the hash of the video into the watermark. Since different videos have different hashes, a watermark copied from one video onto another will not be correctly detected. Similar to a video watermarking technique, one can calculate the hash of a video by treating each frame of a video sequence separately. Alternatively, one can treat the video as a 3-dimensional signal. Video hashing techniques that use this approach can also be found in the literature. For example, the authors in [6] and [7] proposed video hashing schemes based on the 3-dimensional discrete cosine transform of the video sequences.

In this paper, we propose a video watermarking scheme in which the embedded watermark signal is made content-dependent by incorporating the hash of the host video. We treat the video as a 3-dimensional signal. First we transform the 3-dimensional signal using 3-dimensional discrete cosine transform (DCT). We then pick a subset of the resulting DCT coefficients for both hash calculation and watermark embedding. The hash is calculated from a set of pseudo-randomly chosen coefficients from this subset. The watermark is embedded in the remaining coefficients. Our experiments show that this scheme has a satisfactory performance against attacks, in particular the copy attack. The remainder of this paper is organized as follows. In section 2 we discuss the overview of the proposed system. Section 3 presents the proposed hashing and watermark embedding schemes in greater

details. Section 4 discusses our experimental setup and results. Finally, in Section 5 we present our conclusions and provide pointers to our future works.

II. Overview of the proposed scheme

Let V_H be a digital video sequence to be watermarked. This sequence has a spatial resolution of $W \times H$ pixels and contains T video frames. In order to embed a content-dependent watermark sequence $\mathbf{M}$ containing watermark payload $\mathbf{M_P}$ into the video, one must first compute the hash of the video sequence, $\mathbf{I_D}$. In the proposed scheme two secret keys, K_1 and K_2, are used.

The steps to embed a content-dependent watermark can be summarized as follows:

1) Compute the 3-dimensional DCT coefficients of V_H.
2) Choose a subset of DCT coefficients. From this subset, n_H coefficients are then pseudo-randomly chosen (using secret key K_1) and used to compute the video hash.
3) Compute the video hash, $\mathbf{I_D}$, and perform an XOR operation between $\mathbf{I_D}$ and the watermark payload, $\mathbf{M_P}$, yielding a new bit sequence $\hat{\mathbf{M}}_\mathbf{P}$.
4) Arrange the DCT coefficients that have not been used in step 2 into a vector, using K_1.
5) Construct the watermark sequence $\mathbf{M}$ by spreading $\hat{\mathbf{M}}_\mathbf{P}$ using a pseudo-random sequence constructed using the second secret key, K_2.
6) Embed the watermark sequence into the vector obtained in step 4.
7) Use 3-dimensional inverse DCT to reconstruct the watermarked video, V_M.

The embedded watermark payload is retrieved from a watermarked video (that may have been attacked), $\hat{V}_M$, using the following steps:

1) Compute the 3-dimensional DCT coefficients of $\hat{V}_M$.
2) Choose a subset of the DCT coefficients. Using K_1, pseudo-randomly choose n_H coefficients from this subset to compute the hash, $\mathbf{I'_D}$.
3) Arrange the DCT coefficients that have not been used in step 2 into a vector, using K_1.
4) Regenerate the watermark spreading sequence using K_2.
5) Using correlation between the spreading sequence obtained in step 4 and the vector obtained in step 3, retrieve the sequence $\hat{\mathbf{M}}'_\mathbf{P}$.
6) Decode the watermark payload by performing an XOR operation between $\mathbf{I'_D}$ and $\hat{\mathbf{M}}'_\mathbf{P}$.

As can be seen from the previous explanation, a correct hash is needed to actually decode the embedded payload, $\mathbf{M_P}$. This feature makes the proposed scheme robust against a copy attack. Since different video sequences will yield different hash strings, a watermark copied from one video to another cannot be correctly decoded to retrieve $\mathbf{M_P}$ even when the correct keys K_1 and K_2 are used. A more detailed discussion regarding the hashing and watermark embedding/detection algorithms is provided in the following section.

III. Hashing and watermarking algorithms

As explained in the previous section, the first step of the proposed scheme is to compute the 3-dimensional DCT coefficients from the video sequence to be watermarked, V_H. The DCT coefficients are computed by performing a DCT transform in the two spatial directions (the width and height of the frame) and one temporal direction. Given a video sequence with a spatial resolution of $W \times H$ pixels and having T frames, the result of the DCT transform will be a 3-dimensional matrix, C_M, containing $W \times H \times T$ coefficients. We then proceed to choose a subset of C_M, containing $w \times h \times t$ coefficients. These coefficients will be processed further for hash computation and watermark embedding. We call this new matrix C_S. In this paper, we choose the low- to mid-frequency coefficients of C_M to construct C_S. These frequencies offer good robustness and differentiating power of the resulting hash [6]. Furthermore, in our experiments we choose $w = h = t = 32$, thus the matrix C_S contains 32768 DCT coefficients.

A. Hashing algorithm

Given a matrix C_S containing DCT coefficients, the first step to compute the hash of the video sequence is to pseudo-randomly choose n_H coefficients for the hash calculation. The coefficients are chosen based on the secret key K_1. The number of coefficients chosen, n_H, determines the length of the video hash. In our experiments, we choose $n_H = 32$ to generate a 32-bit hash. We call the vector containing these chosen coefficients $\mathbf{C_H}$.

The next step is to compute the median of $\mathbf{C_H}$, designated by $\tilde{C}_H$. The hash vector of the video sequence, $\mathbf{I_D}$, is then computed as follows [6].

$$I_{Di} = \begin{cases} 1 & \text{if } C_{Hi} > \tilde{C}_H \\ 0 & \text{if } C_{Hi} \leq \tilde{C}_H \end{cases} \quad (1)$$

In this equation, the index $i = 1, 2, 3, \ldots, n_H$ represents the i-th element of $\mathbf{I_D}$ and $\mathbf{C_H}$.

B. Watermarking algorithm

After computation of the video sequence hash, we have $(w \times h \times t) - n_H$ coefficients in C_S that are not used to compute the hash. We then proceed to arrange these coefficients into a vector, $\mathbf{C_W}$. The order in which we arrange the coefficients into $\mathbf{C_W}$ is also pseudo-randomly determined using K_1. These coefficients will be used to embed the watermark. We choose to embed the watermark in the coefficients that are not being used to compute the video sequence hash in order to make sure that the hash will not change after watermark embedding.

The first step in the watermark embedding algorithm is to create a spreading sequence, $\mathbf{s}$, which is a pseudo-random, zero mean binary sequence containing the numbers 1 and -1. This sequence has a length L_s, which in our experiments equals 1000. In order to create this pseudo-random sequence, we use the second secret key, K_2. Next we create $\hat{\mathbf{M}}_\mathbf{P}$ by performing an XOR operation between the original payload, $\mathbf{M_P}$ and the video hash $\mathbf{I_D}$.

$$\hat{\mathbf{M}}_\mathbf{P} = \mathbf{M_P} \veebar \mathbf{I_D} \quad (2)$$

In this paper, the watermark payload is also of length n_H.

Finally, we create the watermark sequence $\mathbf{M}$ by modulating $\hat{\mathbf{M}}_\mathbf{P}$ using $\mathbf{s}$.

$$\mathbf{M} = \sum_{j=1}^{n_H} (2\hat{M}_{Pj} - 1)\mathbf{s} \tag{3}$$

In this equation, $\hat{M}_{Pj}$ denotes the j-th bit of the modified watermark payload. It should be obvious that the resulting watermark sequence, $\mathbf{M}$, has a length of $n_H \times L_s$.

The watermark sequence $\mathbf{M}$ is then embedded into the DCT coefficients in $\mathbf{C_W}$. Each member of $\mathbf{M}$, denoted by M_k, is added to one DCT coefficient according to the following formula.

$$\hat{C}_{Wk} = C_{Wk} + \alpha M_k \tag{4}$$

C_{Wk} represents the k-th DCT coefficient taken from $\mathbf{C_W}$. The parameter α is the strength of the watermark. Higher strength will provide higher robustness to attacks but also more severe degradation to the video, and vice versa. In our experiments, we choose $\alpha = 8$. The final step in the watermark embedding process is to put the modified coefficients back into C_M and then use 3-dimensional inverse DCT transform to reconstruct the watermarked video, V_M.

The watermark detection proceeds as follows. Given a video sequence $\hat{V}_M$ that may contain a watermark (and may have been attacked), we first perform the 3-dimensional DCT transform operation and create C'_S, a subset of the DCT coefficients for further processing. Then, using K_1, we pseudo-randomly choose the DCT coefficients to construct the vector $\mathbf{C'_H}$ and compute the video hash, $\mathbf{I'_D}$. The second step is to create the spreading sequence $\mathbf{s}$ using K_2. The third step is to arrange the DCT coefficients in C'_S that are not used in the hash computation process into a vector, $\mathbf{C'_W}$. The order in which the DCT coefficients are arranged in this vector is pseudo-randomly determined by K_1. The fourth step is to construct $\hat{\mathbf{M}}'_\mathbf{P}$. To achieve this, we first partition $\mathbf{C'_W}$ into $n_H \times L_s$ non-overlapping sub-vectors. Then, each bit of $\hat{\mathbf{M}}'_\mathbf{P}$ is determined using the following formula.

$$\hat{M}'_{Pj} = \begin{cases} 1 & \text{if } \mathrm{corr}(\mathbf{C_{Wj}}, \mathbf{s}) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

In this equation, $\mathrm{corr}(\mathbf{a}, \mathbf{b})$ represents the correlation of two vectors $\mathbf{a}$ and $\mathbf{b}$, while $\mathbf{C_{Wj}}$ is the j-th subvector of $\mathbf{C_W}$. The final step is to decode the watermark payload, $\mathbf{M'_P}$, by using the following formula.

$$\mathbf{M'_P} = \hat{\mathbf{M}}'_\mathbf{P} \veebar \mathbf{I'_D} \tag{6}$$

The watermark embedding and detection processes explained in this section are summarized in Figures 1 and 2, respectively.

IV. Experimental Setup and Results

The proposed system is evaluated using 8 different video sequences. Each video sequence has a QCIF resolution (176×144 pixels) and contains 200 frames. In our experiments, we use grayscale video sequences. Some examples of the frames taken from the test video sequences are presented in Fig. 3.

The experiments to evaluate the performance of the proposed system are divided into two main parts, namely experiments to evaluate the performance of the video hash and the experiments to evaluate the complete content-dependent video watermarking system.

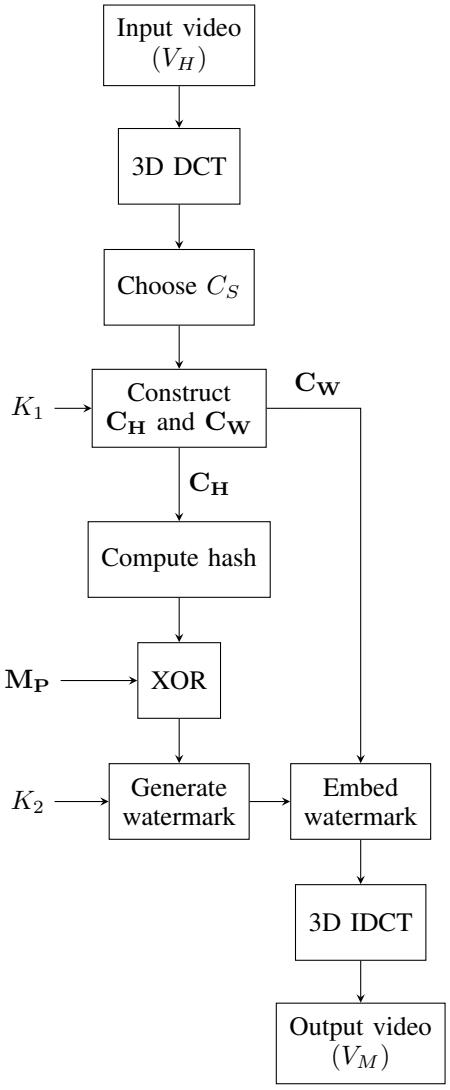


Fig. 1: Watermark embedding process

A. Evaluation of the video hash

Basically, the video hash is evaluated in two performance criteria, namely the intra-hash similarity and the inter-hash similarity. Intra-hash similarity measures how similar the hash between an original video sequence and attacked versions of itself is. In this test, good hashing algorithm should give hash sequences that have high similarity. Inter-hash similarity measures how similar the hash between a video sequence and other, different, video sequences is. A good hashing algorithm will give hash sequences that have low similarity. The similarity between two hash sequences are measured by their Hamming distances. Low Hamming distance indicates high similarity, and vice versa.

Furthermore, in order to evaluate the robustness of our video hash, we perform the following attacks to the video sequences prior to computing the hash:

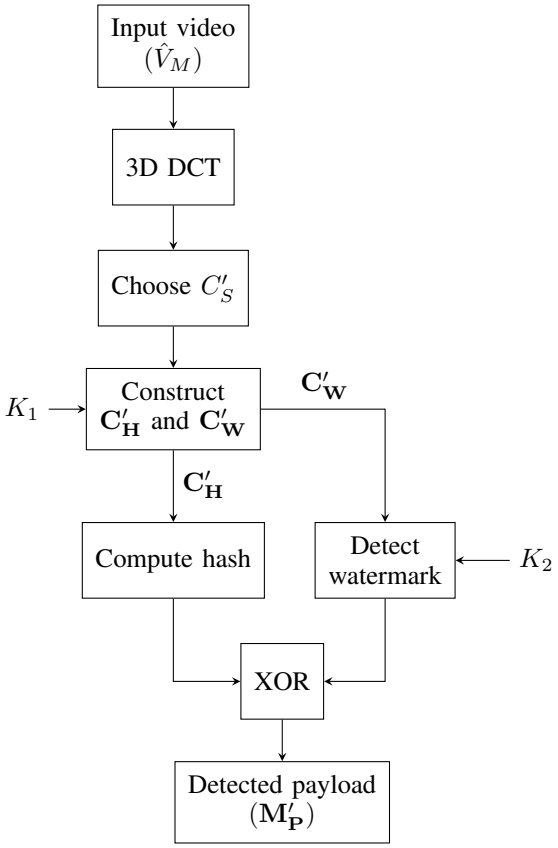


Fig. 2: Watermark detection process

1) Blurring (using a Gaussian low-pass filter with $\sigma = 1.0$)
2) Noise addition (using a Gaussian noise with variance 185)
3) Luminance adjustment by a factor of $\pm 10\%$
4) Frame resizing (shrinking by a factor of 0.9 or enlarging by a factor of 1.1)
5) Random frame dropping (5 frames)
6) MPEG compression (using Q scales for the I, P and B frames $Q_I = 4$, $Q_P = 5$ and $Q_B = 15$, respectively)

The results of the Intra- and Inter-Hash similarity are shown in Table I.

TABLE I: Average Intra- and Inter-Hash Hamming distance under various attacks

Attack	Avg. Hamming Distance	
	Intra-Hash	Inter-Hash
No attack	0.00	16.29
Blur	0.25	16.07
Noise	3.00	15.64
Lum. reduction	0.25	16.04
Lum. increase	0.75	16.14
Frame enlarge	2.00	16.50
Frame shrink	1.50	16.00
Frame drop	3.50	16.50
MPEG	3.00	15.93

Table I shows that the average inter-hash Hamming distance is close to 16 under various conditions. This indicates that the hash sequences between different video sequences have low similarity. This is because a Hamming distance close to 16 means that, on average, 50% of the hash bits are different when two hash sequences from two different video sequences are compared. In other words, the result is comparable to the one expected from comparing two random bit strings. Table I also shows that the average intra-hash Hamming distance is low when the video sequences are subjected to various attacks. This means that under various attacks, the hash sequence for a video sequence is not significantly altered. In other words, this result shows that the hash is robust against these attacks.

(a) "Coastguard" (b) "Hall"

(c) "Mother" (d) "Silent"

Fig. 3: Examples of the test video sequences

B. Evaluation of the complete scheme

The first performance parameter in the evaluation of the complete scheme is the quality impact of the watermark embedding. To evaluate this, we measure the average frame-by-frame PSNR values between an original and a watermarked video sequence. From the eight test sequences, we have achieved an average PSNR of 54.47 dB. This value indicates that the watermark embedding process does not degrade the quality of the video sequences significantly.

The second paramater is the robustness of the watermark against attacks. In this paper, our main concern is not watermark robustness against attacks that may remove the watermark (or render it undetectable) but against an attack that copies the watermark from one video to another. The attack is implemented as follows. Let V_H be an original video sequence and V_M is the watermarked version of this sequence. The watermark embedded is estimated by using the following formula.

$$\hat{\mathbf{M}} = V_H - V_M \tag{7}$$

The estimated watermark is then added to another video sequence, V_{H2} as follows.

$$V_{M2} = V_{H2} + \hat{\mathbf{M}} \tag{8}$$

It should be noted that this implementation is different from the one originally proposed in [5]. Furthermore, the attack as implemented here is unlikely to happen since the attacker would require the original, non-watermarked version of the video sequence. Rather, this implementation represents the worst-case scenario, in which an attacker is able to find an almost exact estimate of the embedded watermark.

We choose two video sequences, "Akiyo" and "Mother", and copy the watermark embedded in these sequences to the other test sequences. The average bit error rates for the watermark copied from "Akiyo" and "Mother" sequences into other sequences are respectively 0.48 and 0.49. This shows that on average, about half of the watermark payload bits are incorrectly detected, which is similar to the expected rate when we try to detect the watermark with a random key. This result shows that the proposed watermarking scheme is indeed robust against an attack in which watermark from one video sequence is copied to another sequence.

In addition to the specific robustness of the watermark against watermark copying, we also evaluate the general robustness of the watermark against other attacks. Therefore, we have subjected the watermarked videos to the attacks described in the previous sub-section. The result of the experiments are shown in Table II.

TABLE II: Average Bit Error Rate under various attacks

Attack	Average BER
No attack	0.07
Blur	0.08
Noise	0.16
Lum. reduction	0.07
Lum. increase	0.09
Frame enlarge	0.13
Frame shrink	0.11
Frame drop	0.16
MPEG	0.14

From Table II we can conclude that the watermark shows good robustness against the blurring and luminance adjustment attacks. The scheme has moderate robustness against noise addition. While this is in part due to the relatively high power of the noise added (the noise introduces, on average, a 34.63 dB quality degradation compared to the non-attacked watermarked video sequence), some robustness improvement against noise is needed. Robustness improvement against temporal attack, namely the frame-dropping attack, is also needed.

V. Conclusions and Future Work

In this paper, we present a content-dependent video watermarking scheme. The proposed method uses pseudo-randomly selected 3-dimensional discrete cosine transform coefficients in order to compute the video hash and also to embed the watermark. Our experiments show that the scheme is robust against copy attack. The scheme is also shown to have good robustness against other attacks.

In our future work, we will continue to investigate the performance of the proposed scheme for a wider range of video inputs and attacks. Furthermore, we will also investigate techniques to increase the robustness of the scheme against temporal attacks.

References

[1] G.C. Langelaar, I. Setyawan, and R.L. Lagendijk, "Watermarking digital image and video data: A state- of-the-art overview," IEEE Signal Processing Magazine, vol. 17, no. 5, pp. 20–46, 2000.

[2] H.H. Mirza, H.D. Thai, Y. Nagata, and Z. Nakao, "Digital video watermarking based on principal component analysis," Proc. Int. Conf. Innovative Computing, Information and Control, ICICIC 2007.

[3] C. Wang, C. Zhang, and P. Hao, "A blind video watermark detection method based on 3D-DWT transform," Proc. IEEE Int. Conf. Image Processing 2010, pp. 3693–3696, 2010.

[4] Y. Liu, D. Wang, and J. Zhao, "Video watermarking based on scene detection and 3D DFT," Proc. Int. Conf. Circuits, Signals and Systems, pp. 124–130, 2007.

[5] M. Kutter, S. Voloshynovskiy, and A. Herrigel, "The watermark copy attack," Proc. SPIE, Security and Watermarking of Multimedia Contents II, vol. 3971, pp. 371–379, 2000.

[6] B. Coskun, B. Sankur, and N. Memon, "Spatio-temporal transform based video hashing," IEEE Trans. Multimedia, vol. 8, no. 6, pp. 1190–1208, 2006.

[7] H. Mao, G. Feng, X. Zhang, and H. Yao, "A robust and fast video fingerprinting based on 3D-DCT and LSH," Proc. Int. Conf. Multimedia Technology, pp. 108–111, 2011.

Design and Implementation of Gaze Tracking Headgear for Nvidia 3D Vision®

Sunu Wibirama and Kazuhiko Hamamoto*, *Member*, *IEEE*

***Abstract*— The usage of Nvidia 3D Vision® is increasing rapidly, ranging from gaming to research purposes. However, researchers in human computer interaction and virtual reality are constrained by hardware configuration since current commercial gaze tracking systems are not specifically designed to be used with Nvidia 3D Vision®. In this paper, we present a novel prototype of gaze tracking headgear which can be used appropriately with Nvidia 3D Vision®. We explain design consideration and detail implementation of our gaze tracking headgear. We also evaluate our gaze tracking system by measuring gaze accuracy on stereoscopic display. Experimental result shows that the average gaze estimation error is less than one degree visual angle.**

***Index Terms*— gaze tracking, active 3D technology, virtual reality, Nvidia 3D Vision®**

I. INTRODUCTION

Virtual environment (VE) has been widely used for entertainment, gaming system, telepresence, and simulating real or imagined scenario, such as design and training processes in industrial and collaborative task [1]–[3]. Stereoscopic display technology is commonly used to enhance user experience in three-dimensional (3D) VE.

Stereoscopic display technology is developed based on stereopsis in human visual system. Since human eyes are separated horizontally, each eye has its own view of the world scene. Consequently, both eyes receive slightly different images. To achieve 3D experience, human brain has to perceive left and right images as a single image. Using information of screen disparity, which is the distance of corresponding point in the left and right images, the depth information can be extracted. Fig.1 shows a brief concept of stereoscopic viewing.

Recently, one of active 3D technology that gains wide attention is Nvidia 3D Vision®. Nvidia 3D Vision® system consists of two main hardwares: infra-red (IR) emitter and a lightweight active shutter glasses (50 gram of weight), as depicted in Fig.2. Nvidia 3D Vision® is widely used in various applications since it preserves color compared with anaglyph-based stereoscopic image. There are many applications that utilize Nvidia 3D Vision®, including 3D haptic-based modeling system [4], augmented reality with freehand interaction [5], navigation for visually impaired people [6], digital exhibition of archaeological structure [7], welding training for industrial worker [8], and performance investigation for sport player [9].

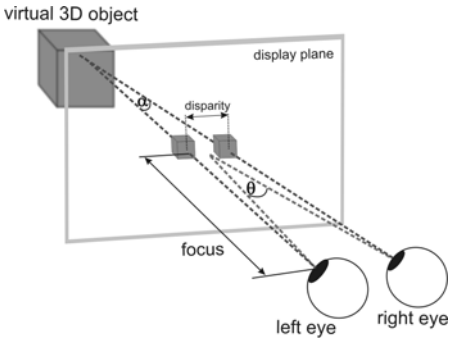


Fig. 1. Concept of stereoscopic viewing. By fusing left and right images, the brain perceives virtual 3D object. Parallax angle is defined as the difference between angle formed by both eyes-virtual object and both eyes-display plane (θ-α).

Fig. 2. Nvidia 3D Vision® system: infra red (IR) emitter (left panel) and active shutter glasses (right panel). The shutter glasses is synchronized with 3D display during stereoscopic images exposure with 120 Hz of shutter frequency.

On the other hand, research on gaze tracking as human-computer interface and visual search tools in 3D space is emerging. Gaze tracker, compared with other input devices such as mouse or keyboard, has been a choice to investigate user attention and visual perception in VE. Moreover, as an interactive interface, human gaze is considered faster than speech or gestures.

Essig *et al*. [10] used a commercial head-mounted gaze tracker to estimate position of stimulus presented in anaglyph stereogram. Lee *et al*. [11] suggested the usage of monocular gaze tracker to measure user gaze in 3D space. Daugherty *et al*. [12] used remote gaze tracker to compute point of gaze over anaglyphic stereogram video while the user has to wear blue and red filter to see the 3D video. Due to the nature of common gaze tracker which is not specifically designed for active 3D technology, the usage of Nvidia 3D Vision® system is normally excluded in previous research works. In case the user has to wear Nvidia 3D Vision® glasses, an appropriate gaze tracking headgear should be developed.

This research is supported by AUN/Seed-Net, JICA, and Tokai University under long-term research grant ID J1110112

*Sunu Wibirama and Kazuhiko Hamamoto are with Graduate School of Science and Technology, Tokai University, Tokyo, Japan 108-8619. Email: sunu@jteti.gadjahmada.edu, hama@keyaki.cc.u-tokai.ac.jp

In this paper, we propose a new prototype of gaze tracking headgear for Nvidia 3D Vision®. We consider proper camera installation to obtain clear eye image without obstructing user view. Proper Nvidia 3D Vision® glasses installation on gaze tracking headgear is also taken into account. Finally, we evaluate our system by measuring gaze accuracy on stereoscopic display.

II. DESIGN CONSIDERATION

There are three possible scenarios to use Nvidia 3D Vision® glasses with gaze tracking headgear, as shown in Fig.3. The first scenario is to directly put the camera in front of the Nvidia 3D Vision® glasses. The installation is not complicated. However, the camera cannot capture the left and right eye images simultaneously since the Nvidia 3D Vision® glasses alternately opens and closes the shutter of left and right eyes with 120 Hz of shutter's frequency. Moreover, straight camera installation obstructs user from viewing the stereoscopic 3D content on the screen.

The second scenario is to install the gaze tracker camera below the Nvidia 3D Vision® glasses. The advantage is similar to the first scenario, but it will limit the vertical field of view and obstruct the user when gazing the bottom part of screen.

The third scenario is to use additional hardware, such as infrared (IR) light and hot mirror (IR reflector). In this case, the vertical field of view of the user is preserved and the camera is able to capture the eye image clearly. However, since the glasses is slightly positioned farther from the eye than its normal usage position, the gaze tracking device has to take into account a mechanism to hold the Nvidia 3D Vision® glasses properly. Unfortunately, to the best knowledge of the authors, no commercial gaze tracking device is available for such purpose. Based on the third scenario, we have developed a new gaze tracking headgear that is suitable to be used with Nvidia 3D Vision® glasses.

III. HARDWARE IMPLEMENTATION

A. Gaze tracking goggle

The Nvidia 3D Vision® glasses is designed to be used appropriately with prescription glasses. Thus, it has wider width range than normal prescription glasses. We consider the size of the gaze tracking headgear to not exceed the maximum width range of Nvidia 3D Vision® glasses and for such purpose, a modified Famicom gaming goggle (Nintendo Co. Ltd., Kyoto, Japan) is used in this research. Fig.4 shows the schematic physical construction of the gaze tracking headgear.

The physical size of Famicom goggle is 17 x 4.5 x 6.5 cm with 365 gram of weight. Front cover and two small LCD displays of the goggle were removed and replaced by a plastic frame to hold hot mirrors. Head strap foam was added to provide comfortable usage for user. Extra surface was added to mount dual-cameras system while maintaining stability, preventing large movements of the goggle during experiment. To hold the Nvidia 3D Vision® glasses properly, a thermoplastic "Y"-shaped shutter glasses holder was designed and added at the middle between left and right hot mirrors. Fig.5 shows the Nvidia 3D Vision® glasses installed on gaze tracking headgear.

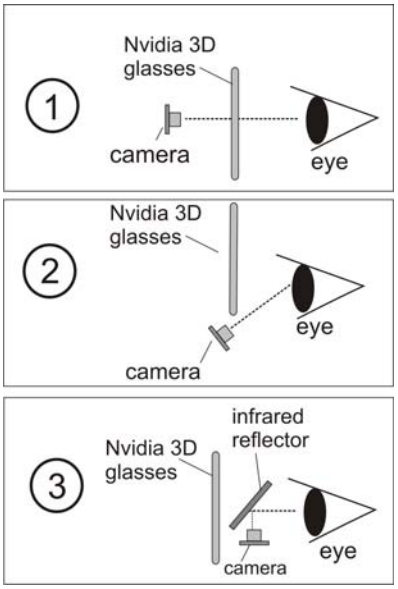


Fig. 3. Three possible scenarios of Nvidia 3D Vision® glasses and camera installation. The proposed gaze tracking headgear is developed based on the third scenario.

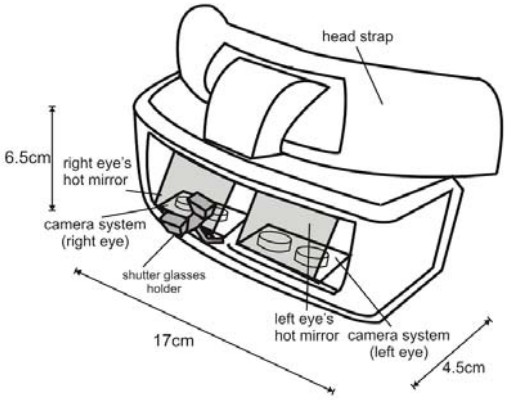


Fig. 4. Schematic physical construction of gaze tracking headgear.

Fig. 5. Front view of the proposed gaze tracking headgear.

B. Dual-camera system

The camera system, consisting of two mini CCD cameras (Analog Technologies, Inc., Santa Clara, United States), is

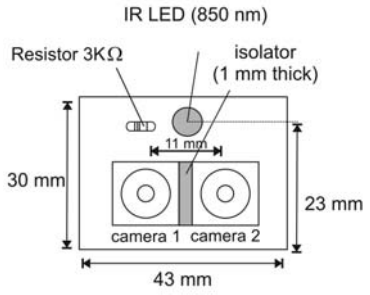


Fig. 6. Schematic configuration of dual-camera system. Camera 1 and camera 2 are installed parallel to obtain rigorous matching result on stereo pair images.

Fig. 7. Evaluation study of gaze tracking headgear.The subject was positioned 63 cm from the monitor. The head was stabilized using chinrest to avoid errors caused by head movement.

utilized as main capturing devices. The size of the circuit board is 43 x 30 mm. The usage of two cameras for one eye is to fulfill requirement of the gaze tracking software which is mainly based on our previous research work [13]. One of the major problems in multiple vision systems is the well-known "matching" problem, which is the determination of conjugate points between stereo pair images. To achieve rigorously matching result on stereo pair images, the two cameras were installed parallel in horizontal direction as shown in Fig.6.

The camera size is 10.7 x 10.7 mm. Each camera is equipped by a 1/4 " CCIQ sensor with 400 lines of resolution. The focal length of the lens is 3.9 mm. Despite its size, the camera provides adequate image quality in light sensitivity up to 0.6 lux for image thresholding, edge detection, and ellipse fitting algorithms provided by our gaze tracking software. The camera is ideal for capturing the eye because of its small size and sensitivity to minimum light exposure.

Obtaining high contrast pupil image in visible light is unlikely. Thus, the camera was modified into IR-sensitive camera by removing the IR filters installed inside the camera. We also attached an IR passing filter (Kodak, Ltd. Tokyo, Japan) to the front of the camera lens to pass IR light with wavelength longer than 750 nm. An IR LED (850 nm wavelength) is used to create illumination on eye region.

To minimize vertical gaze error while allowing the user to gaze at the designated scene, two 50 x 50 mm hot mirrors (Edmund Optics, New Jersey, United States) were installed such that the hot mirror and the optical axis of the eye

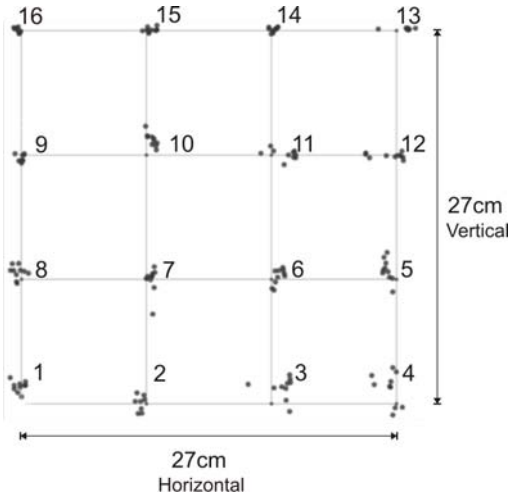


Fig. 8. Example of gaze validation result on sixteen target points.

formed 45° angle. The thickness of the hot mirror is 3.3 mm, allowing it to be installed neatly in the front of the goggle.

C. Supporting system

The supporting system is used to process and analyze the image stream in real-time. To support real-time gaze tracking while rendering stereoscopic content, a personal computer with Intel® i7-2600 3.4GHz processor, 4GB memory, and Windows XP® operating system is used. To support quad-buffered stereoscopic rendering, Nvidia Quadro 5000 graphic card is used.

We also consider the usage of special monitor that supports active 3D technology by providing ability to increase refresh rate up to 120 Hz. A stereoscopic I-O Data LCD-3D231XBR-S monitor with 23 inches of size and 1920 x 1080 pixels of resolution is used. The gaze tracking software was developed using Microsoft Visual C++ 2010, Open Frameworks C++ toolkit [14], Open Source Computer Vision (OpenCV) library [15], and Open Graphics Library (OpenGL) [16].

IV. EXPERIMENT

We evaluated the usefulness of gaze tracking headgear in practical usage by measuring 2D gaze accuracy on stereoscopic display. Gaze accuracy refers to the degree to which the gaze tracking system represents the true value of what is being gazed by the user.

To measure gaze accuracy, a subject was calibrated on a 27 x 27 cm sized square, divided onto 3 x 3 calibration points. The gaze tracking headgear was used with Nvidia 3D Vision® glasses as shown in Fig.7. The distance between the monitor and the subject was 63 cm. The subject was positioned such that the center of the screen was horizontally aligned at the middle point of both subject eyes. The subject's head was stabilized using a chinrest to minimize error in gaze estimation.

The validation session consisted of 3 trials. In each trial, the subject was asked to gaze on 4 x 4 validation points.

TABLE I
RESULTS OF VALIDATION

Experiment	Measurement errors (in degree)		
	Horizontal (H)	Vertical (V)	Mean (H and V)
Trial # 1	0.65	0.73	0.69
Trial # 2	0.68	0.77	0.72
Trial # 3	0.67	0.76	0.71
Average errors		0.70	

TABLE II
COMPARISON OF GAZE TRACKING ACCURACIES

Gaze tracking methods	Average error (in degree)
Gaze tracking using corneal model [17]	0.89
Gaze tracking based on 3D eye model [18]	1.2
The proposed gaze tracking system	0.70

For each validation point, 10 gaze points were automatically recorded by the system. Each trial took about 1.5 minutes of time. One minute rest time was allocated between each trial.

V. RESULTS

Fig.8 shows example of gaze estimation result in 4 x 4 validation points from the 1st trial. Observing Fig.8, it can be seen that when the subject gazed at eccentric position (i.e. target number 3, 4, 13), the errors were larger compared with other centralized positions (i.e. target number 6, 7, 10, and 11). This is due to unrobustness of the pupil detection algorithm of gaze tracking software. If the eye moves to eccentric position, the accuracy of pupil detection algorithm decreases due to occlusion of eyelid.

Gaze tracking validation result is shown in Table 1. In this experiment, average gaze error for 1st, 2nd, and 3rd trial are 0.69°, 0.72°, and 0.71°, respectively. The total gaze detection error averaged from all trials is 0.70°. Keeping the average gaze error below 1° visual angle is important since to see clearly a point of interest, people generally do not have to locate their eye more accurate than the width of the fovea region, which is about 1°-2° visual angle [19]. Table 2 shows comparison of gaze tracking accuracy between previously published methods [17], [18] and the proposed method. It is found that our system accuracy outperforms previously published gaze tracking methods.

VI. CONCLUSIONS

Researchers doing gaze tracking in virtual environment have expressed the need for custom, appropriate to use with Nvidia 3D Vision®, and practical gaze tracking system. Micro-lens cameras and hot mirrors have made it possible to develop gaze tracking headgear that can be used appropriately with Nvidia 3D Vision® glasses while allowing the user to view the designated stereoscopic 3D content. The evaluation results show that our proposed hardware yields slightly better performance than previous methods by 0.70° of average accuracy. In future, we intend to improve the pupil detection accuracy by minimizing noise coming from eyelid and imperfect pupil shape. We are also interested in conducting an experiment to investigate gaze behavior in virtual 3D space. A systematic larger participant group study, beyond the scope of this proof-of-principle paper, will allow us to estimate considerable statistical accuracy of the system, which is very useful for further development onto 3D gaze measurement system.

REFERENCES

[1] H. Y. Kan, V. G. Duffy, and C.-J. Su, "An Internet virtual reality collaborative environment for effective product design," *Computers in Industry*, vol. 45, pp. 197-213, 2001.

[2] T. Monahan, G. McArdle, and M. Bertolotto, "Virtual reality for collaborative e-learning," *Computers and Education*, vol. 50, pp. 1339-1353, 2008.

[3] P. Wang, A. A. Becker, I. A. Jones, A. T. Glover, S. D. Benford, C. M. Greenhalgh, and M. Vloeberghs, "A virtual reality surgery simulation of cutting and retraction in neurosurgery with force-feedback," *Computer Methods and Programs in Biomedicine*, vol. 84, pp. 11-18, 2006.

[4] S. Kamuro, K. Minamizawa, and S. Tachi, "3D Haptic modeling system using ungrounded pen-shaped kinesthetic display," in *IEEE Virtual Reality Conference*, 2011, pp. 217-218.

[5] H. Benko, R. Jota, and A. Wilson, "MirageTable: freehand interaction on a projected augmented reality tabletop," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Austin, Texas, USA, 2012.

[6] S. Vitek, M. Klima, L. Husnik, and D. Spirk, "New possibilities for blind people navigation," in *International Conference on Applied Electronics*, 2011, pp. 1-4.

[7] H. T. Tanaka, K. Hachimura, K. Yano, S. Tanaka, K. Furukawa, T. Nishiura, M. Tsutida, W. Choi, and W. Wakita, "Multimodal digital archiving and reproduction of the world cultural heritage "Gion Festival in Kyoto"," in *Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry*, Seoul, South Korea, 2010.

[8] S. A. White, M. Prachyabrued, T. L. Chambers, C. W. Borst, and D. Reiners, "Low-cost simulated MIG welding for advancement in technical training," *Virtual reality*, vol. 15, pp. 69-81, 2011.

[9] T. Hoinville, A. Naceri, J. Ortiz, E. Bernier, and R. Chellali, "Performances of experienced and novice sportball players in heading virtual spinning soccer balls," in *IEEE Virtual Reality Conference*, 2011, pp. 83-86.

[10] Kai Essig, M. Pomplun, and H. Ritter, "A neural network for 3D gaze recording with binocular eye trackers," *The International Journal of Parallel, Emergent and Distributed Systems*, vol. 21, pp. 79-95, 2006.

[11] J. W. Lee, C. W. Cho, K. Y. Shin, E. C. Lee, and K. R. Park, "3D gaze tracking method using Purkinje images on eye optical model and pupil," *Optics and Lasers in Engineering*, vol. 50, pp. 736-751, 2012.

[12] B. C. Daugherty, A. T. Duchowski, D. H. House, and C. Ramasamy, "Measuring vergence over stereoscopic video with a remote eye tracker," in *Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications*, Austin, Texas, 2010.

[13] S. Wibirama, S. Tungjitkusolmun, and C. Pintavirooj, "Dual-Camera Acquisition for Accurate Measurement of Three-Dimensional Eye Movements," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 8, pp. 238-246, 2013.

[14] J. Noble, *Programming Interactivity*, 2nd Edition: O'Reilly Media, 2012.

[15] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*, 1st ed. California: O'Reilly Media, Inc., 2008.

[16] Khronos. (2007, 22 January 2010). OpenGL - The Industry Standard for High Performance Graphics. Available: http://www.opengl.org/

[17] E. Lee and K. Park, "A Study on Eye Gaze Estimation Method Based on Cornea Model of Human Eye," *Computer Vision/Computer Graphics Collaboration Techniques*, vol. 4418, A. Gagalowicz and W. Philips, Eds., ed: Springer Berlin / Heidelberg, 2007, pp. 307-317.

[18] W. Jian-Gang and E. Sung, "Study on eye gaze estimation," *IEEE Transactions on Systems, Man, and Cybernetics*,vol. 32, pp. 332-350, 2002.

[19] TobiiEyeTracker. (2012, 5 September). An Introduction to Eye Tracking and Tobii Eye Trackers. Available: http://www.tobii.com/eye-tracking-research/global/library/white-papers/tobii-eye-tracking-white-paper/

Detection and Object Position Measurement using Computer Vision on Humanoid Soccer

Iwan Awaludin, Priyanto Hidayatullah, Jonner Hutahaean, Dewa Gede Parta
Computer Engineering and Informatics Department
Bandung State Polytechnic
Bandung, Indonesia
awaludin@jtk.polban.ac.id, priyanto@polban.ac.id, jonnerh@jtk.polban.ac.id, dewagp@jtk.polban.ac.id

***Abstract*—Bandung State Polytechnic (POLBAN) has participated twice in humanoid robot soccer competition. From those two participations, it was known that the weakness was in computer vision. Computer vision capability is constrained by robot hardware specifications so that it was impossible to embed our recent object recognition application.**

In this study, we propose a computer vision system that implemented the latest technology similar to that technology used in the humanoid soccer winner season 2011. The model uses a field where the object and size comply with the rules of humanoid soccer tournament 2011.

Some previous methods use off the field camera which is cannot be used in humanoid soccer tournament because the sensor used has to be attached to the robot. While the approach in this paper emphasized to the fact that goalkeeper's position tend to be static relative to the object in a competition field. Goal keeper through its vision system recognizes objects and measures ball position using image processing technique.

The process of ball position measurement was first carried out by recognizing three different objects in the competition field: ball, goal's bar, and field line. Recognition process utilizes back projection method based on HSV information.. After the three objects were detected, the measurement of ball position on the field was carried out by ANN model by considering ball position in the image, position of goal's horizontal bar, and the middle field line point. After 10,000 training, the result is encouraging with the average error is less than 1 cm.

***Keywords*—computer vision, humanoid soccer robot, image processing, object detection, ANN**

I. INTRODUCTION

Humanoid robot soccer league is a match among autonomous robots. It is aimed at preparing robots to compete with the real human soccer players by 2050. The game is divided into various classes which are kid size, teen size, and big size.

Polban has been participating in this game at regional level since 2011 by presenting kid size robot. In 2012 Polban's robot could not proceed to the next round due to some identified flaws. The flaws include control of the robot body to run and kick the ball, to get up, and to get the computer vision to determine the position of objects in the soccer field.

This research will be focused on one of the above mentioned flaws: computer vision of humanoid robot. Normally a human player would use his vision to determine the action within the game. For example human players actions to catch the ball, drive the ball, and also kick the ball are made based on information obtained from the human vision. Humanoid robots in the game are also trying to mimic human players necessary processing of the information from computer vision to determine the action to be performed by the robot.

There are some interesting issues in humanoid computer vision such as:

1. Self localization such as carried out by [1], [2], [3], [4]
2. Object detection and tracking [5], [6], [7], [8]

Back in 2012, most participants of regional II humanoid soccer robots league require a long time to find the ball and direct it towards the opponent's goal. This is due to the robot itself does not know the position of the ball in field coordinates, so it takes extra time to look over the goal net after getting the ball.

Similarity of computer vision problems in all robot participants has raised research questions:

1. Which method is suitable to detect objects in humanoid soccer game field?
2. How to determine ball position in field coordinate based on previously detected objects in the game field.

Both questions will be investigated in this research.

II. RELATED STUDIES

Object detection and discovery of ball's position are not a new research issues. Some publications already have those as their main investigation. However those publications have different data and purpose with our study.

For example is the detection of objects such as players, ball, goal line, and the goal of some off-field cameras (broadcast video) reported by [9]. In that paper, the detection is performed to determine the position of the ball and players on the field with the ability to detect bouncing ball. But of course, this result cannot be used by a humanoid robot as the sensor used has to be attached to the robot, instead of being off the field. In addition to the wide field of vision is very different

from the point of view of the humanoid robots. It is limited in terms of scope and angle of vision.

Some publications show the results of a study of goalkeeper's vision. In this case the computer vision should be able to identify the object i.e. ball in the middle of a digital image. Ball is identified by HSL information. Once the robot identified the ball, it will rotate it self so that the ball is perpendicular to the robot. After that, the position of the ball will be calculated based on the camera's angle and robot's height. The calculation is usually done by linear equations. The weakness of this study is that the nature of the camera which is not linear due to the production process. It means that the calculation of different cameras will also result in a different position. Thus different calculation approach is needed.

Objects in the robot soccer game is usually searched one by one. For example, at the beginning, we want to look for the ball. Then the ball's information will be looked for from the digital image. For example, orange ball is scanned. After getting the ball, the robot body will be moved so that the ball is in the middle of a digital image. After the position of the ball relative to the robot are calculated, then the robot's action will be determined according to the position of the ball. In this case the goalkeeper robot will flop, move to a certain angle, or kick the ball.

Some preliminary findings of related studies are:

1. Computer vision is used to find a single object from each image obtained.
2. Computer vision still requires the robot body to be perpendicular to the target object so that the position of the object relative to the robot can be calculated.

These findings are different from the research we want to do. In this research, we want to find out

1. How to detect 3 different objects from one image.
2. How to calculate the position of the target object relative to the field instead of to the robot
3. Body of the robot does not have to be perpendicular to the target object

To achieve this, we designed the system as explained in the following chapter.

III. SYSTEM ARCHITECTURE

Several assumptions need to be taken to simplify the architecture and algorithms that will be investigated. These assumptions are:

1. The position of goalkeeper in the game tend to be static. Goalkeeper lies under the goal which its distances from other static objects are known.
2. The ball is always in touch with the playing field. The ball does not bounce up.
3. The size and color of all objects in the field has been defined in accordance with the game rules.

For this purpose we prepared a playing field that is fit with game rules of kid size humanoid robot soccer[10]. The field size is 4 x 6 m. The position of the left and right foot of the goal is 1.25 m and 2.75 m. Field lines have a white spot in the middle. Ball position will vary in the field. If one assumes that the field is represented by a 400 x 600 point with the point (0,0) is on the left end of the field then

1. The goalkeeper position is (200, 600)
2. The middle point of the field position is (200, 300)
3. The goal foot 1 is at (125,0) and the goal foot 2 is at (275,0)

Some objects that present on the playing field are tennis ball (orange), playing field (green), field line (white), goals (blue and yellow). For every image, the system will do two tasks which are:

1. Detect ball, goal's feet, and middle spot.
2. Calculate the position of the ball

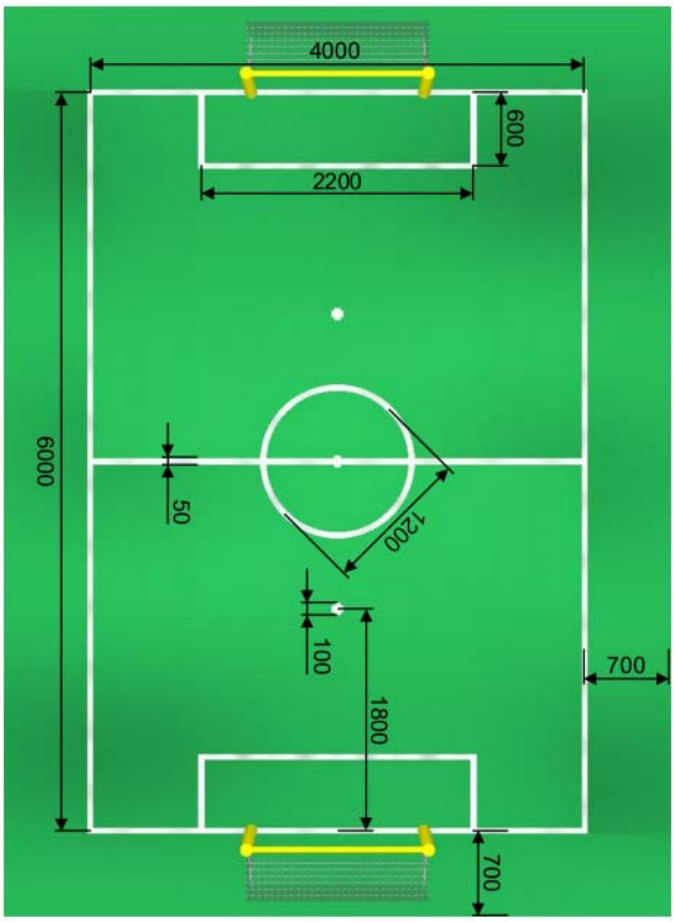


Fig. 1. Scale diagram of entire field (in mm)[10]

A. Object Detection

For this research we have to detect 3 objects at once: the ball, the goal, and the middle spot of the playing field. RGB color model is not good for object detection. Each RGB component is not distinctive enough to separate one object

part from another in an image. That is why we use another color model. The one we choose is HSV color model because it is clear that each object has different Hue.

There are some rules and formula for obtaining HSV value from RGB image. These rules must be follow so that no loss of information. The equations are following[12]:

$$V = Max(R,G,B) \tag{1}$$

$$S = \begin{cases} \dfrac{V - Min(R,G,B)}{V}, V \neq 0 \\ 0, V = 0 \end{cases} \tag{2}$$

If S = 0, then H = 0.
If R = V, then,

$$H = \begin{cases} \dfrac{60(G-B)}{V - Min(R,G,B)}, G \geq B \\ 360 + \dfrac{60(G-B)}{V - Min(R,G,B)}, G < B \end{cases} \tag{3}$$

If G = V, then

$$H = 120 + \frac{60(B-R)}{V - Min(R,G,B)} \tag{4}$$

If B = V, then,

$$H = 240 + \frac{60*(R-B)}{V - Min(R,G,B)} \tag{5}$$

The object detection itself is done semiautomatically. Some procedure is based on the object tracking algorithm in [11]. Detecting the goal for example, in simple way, the procedure can be explained as following. The image is converted from RGB to HSV. User then click any part of the goal then it grows until the edge of the goal using meanshift segmentation and region growing. At this stage, it stores the minimum and maximum HSV value to make a color mask. The process is continued with modeling the object using color histogram on Hue, Saturation and Value component. Each histogram is backprojected to the image. Then the backprojected images are combined using AND operator which results are illustrated in Fig 2 and Fig 3. As we can see from the Fig 3, the goal can be detected succesfully.

As mentioned above, for each image, three objects are detected silmutaneously. On the result image, each object will be shown as something that has value while nonobject will be declared as the value 0.

Ball is defined by its lowermost pixel. In other hand, goal is defined by its lowermost both feet pixel. While for field line, it is defined by midpoint of line that stretched from left to right. All those information are stored for further processing.

B. Finding the ball

This process is the process of finding the position of the ball relative to the coordinates of the field. Before capturing the image, we have noted the actual position of the ball on the field coordinate. The coordinate of the point will be fed into the system as a reference

Initially we expected to use linear equations to obtain the position of the ball. But the linear equation failed to give satisfying results. Each calculation has an error that is more than 300 cm. This of course can be said as fail. According to some references, calibration is needed before calculating the position of the ball.

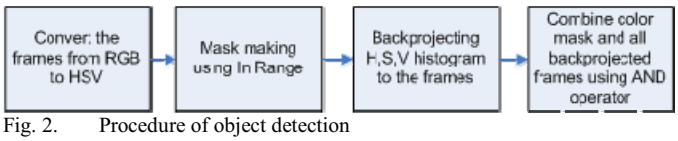
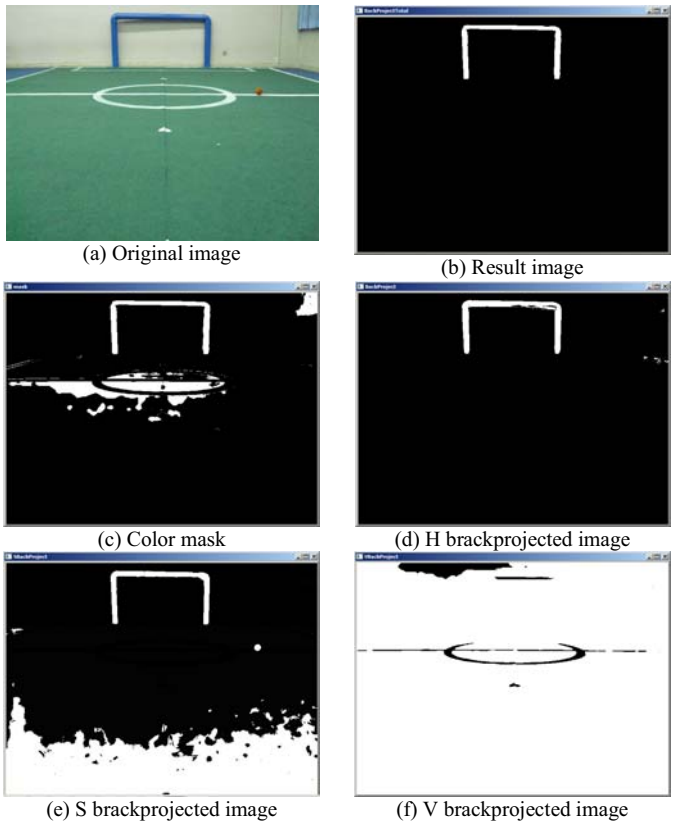


Fig. 2. Procedure of object detection



(a) Original image (b) Result image

(c) Color mask (d) H brackprojected image

(e) S brackprojected image (f) V brackprojected image

Fig. 3. Illustration of object detection process

To simplify the process of finding the ball, it was decided to use non-linear methods in order to accommodate the natural condition of the camera lens. The method we choose is the Artificial Neural Network (ANN). Information gathered from the process before capturing the image and the results of object detection will be made as reference for ANN. There are 46 image samples. Those images will be divided into training data and testing data. The proportion of training and testing data is 80% to 20%.

As the input is the data obtained from the object detection process, while the output is the data from the measurement position of the ball. So there are 8 inputs: x and y coordinates

of the two goal's feet, the ball, and the midpoint of the field, while the outputs are x and y coordinates of the ball.

The next step is determining ANN architectures such as: the number of hidden layer, the number of neurons in each hidden layer, activation function in each layer and learning algorithm to change the weight value. The ANN architecture used in this paper is given in Fig. 4.

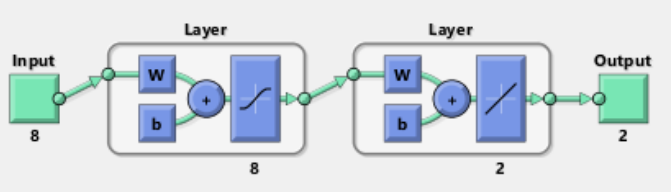


Fig. 4. The ANN architecture in this paper

Output of each hidden layer would be input to the next layer. The last layer output then compared to the manual measurement before capturing the image. The delta between those measurements will be returned to the previous layer to change the weight value according to the learning methods used. Each time weight value is updated is known as an epoch. The way ANN learns is repeated until the expected performance is achieved, e.g. specifying maximum error. In experiment, some of the following values will be used and reported the results in the next section:

1. The amount of hidden layer 1 and 2
2. The amount of Neuron from 2 to 16
3. Activation combination of tansig and purelin function.
4. Learning algorithm gdm, gd, and LM

TABLE I. TRAINING RESULTS AND FACTUAL BALL POSITION

Xtrain	Ytrain	Xori	Yori
-0.29946	0.082324	0	0
130.3548	-0.04763	130	0
199.7694	0.110398	200	0
285.5046	0.232881	285	0
326.7263	0.880777	325	0
397.7122	-1.2599	400	0
340.2095	144.8709	340	145
308.3807	219.1656	310	220
284.7753	299.8883	285	300
30.00935	70.21162	30	70
59.64023	144.5681	60	145
91.4443	220.1795	90	220
114.6627	299.6508	115	300
147.1512	370.2991	147	370
174.9325	446.8805	175	447
75.03499	300.6108	75	300
123.5894	299.9082	125	300
275.31	300.2149	275	300
199.4231	299.5733	200	300
199.8632	349.7338	200	350
200.138	450.2818	200	450
200.1244	500.083	200	500
198.3634	49.63603	200	50
199.9874	99.06983	200	100
199.7894	149.8717	200	150
200.3693	249.936	200	250
97.13111	163.1104	97	163
9.232697	184.6717	9	185
130.1855	237.2695	130	237
296.4492	34.39418	295	33
388.1055	142.7303	387	142
102.4084	347.7669	102	348
164.4385	410.2949	164	410
157.7839	485.9218	158	486
242.1545	409.9679	242	410
235.9171	485.9566	236	486
298.1714	362.2492	298	362

TABLE II. TESTING RESULTS AND FACTUAL BALL POSITION

Xtesting	Ytesting	Xori	Yori
75	0	77.48342	23.40226
130	0	157.9409	141.2816
285	0	290.5979	10.0102
400	0	390.8541	-32.2499
175	447	173.8582	447.6346
125	300	136.0052	289.5077
200	150	201.7087	225.2288
200	200	190.4663	228.9306

IV. RESULTS

Lavenberg Marquadt is chosen which can achieve the target minimum error after, on the average, 60 epoch. The use of Traingd and Traingdm methods makes the learning process very long. Even after 20,000 epochs, it still can't reach the learning target.

Activation function purelinear accelerates the learning process when it placed in the output layer while activation function tansigmoid improving accuracy when it placed in the hidden layer. Random initialization is also accelerates the learning process compared to the use of default values.

After 10,000 times training experiment the average error is less than 1 cm while from the testing experiment, the average error is 70 cm. Table 1 shows the original data and the results of training while Table 2 shows original data and the results of testing.

V. CONCLUSION AND FUTURE WORKS

In this paper, we have shown that we succeed to build a system that can be used to detect 3 different object simultaneously using HSV information and calculate ball position using ANN.

We also conclude if the minimum amount of neuron is 6 then it will give good accuracy while adding hidden layer do not improve the accuracy.

In the future we want to improve accuracy by increasing the number of training samples and develop the system so that it can fully automatic detect the objects.

REFERENCES

[1] Jen-Shiun Chiang, Chih-Hsien Hsia, and Hung-Wei Hsu, "A Stereo Vision-Based Self-Localization System," Sensors Journal, IEEE, vol. 13, no. 5, pp. 1677–1689, May 2013.

[2] Jen-Shiun Chiang, Chih-Hsien Hsia, Hung-Wei Hsu, and Chun-I Li, "Stereo vision-based self-localization system for RoboCup," Fuzzy Systems (FUZZ), 2011 IEEE International Conference on, pp. 2763–2770, 27.

[3] Jen-Shiun Chiang, Chih-Hsien Hsia, Shih-Hung Chang, Wei-Hsuan Chang, Hung-Wei Hsu, Yi-Che Tai, Chun-Yi Li, and Meng-Hsuan Ho,

"An efficient object recognition and self-localization system for humanoid soccer robot," SICE Annual Conference 2010, Proceedings of, pp. 2269–2278, 18.

[4] Shih-Hung Chang, Wei-Hsuan Chang, Chih-Hsien Hsia, Fun Ye, and Jen-Shiun Chiang, "Efficient neural network approach of self-localization for humanoid robot," Pervasive Computing (JCPC), 2009 Joint Conferences on, pp. 149–154, 3.

[5] B. Browatzki, V. Tikhanoff, G. Metta, H. H. Bulthoff, and C. Wallraven, "Active object recognition on a humanoid robot," Robotics and Automation (ICRA), 2012 IEEE International Conference on, pp. 2021–2028, 14.

[6] A. Fidjeland, M. Shanahan, and A. Bouganis, "Object Detection for a Humanoid Robot Using a Probabilistic Global Workspace," in Cognitive Vision, vol. 5329, B. Caputo and M. Vincze, Eds. Springer Berlin Heidelberg, 2008, pp. 135–148.

[7] J. Leitner, P. Chandrashekhariah, S. Harding, M. Frank, G. Spina, A. Forster, J. Triesch, and J. Schmidhuber, "Autonomous learning of robust visual object detection and identification on a humanoid," Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on, pp. 1–6, 7.

[8] Wei-Hsuan Chang, Chih-Hsien Hsia, Yi-Che Tai, Shih-Hung Chang, Fun Ye, and Jen-Shiun Chiang, "An efficient object recognition system for humanoid robot vision," Pervasive Computing (JCPC), 2009 Joint Conferences on, pp. 209–214, 3.

[9] Y. Ohno, J. Miurs, and Y. Shirai, "Tracking players and a ball in soccer games," Multisensor Fusion and Integration for Intelligent Systems, 1999. MFI '99. Proceedings. 1999 IEEE/SICE/RSJ International Conference on, pp. 147–152, 1999.

[10] http://www.tzi.de/spl/pub/Website/Downloads/Rules2012.pdf accessed on 16 June 2013

[11] Hidayatullah, P., Konik, H., "CAMSHIFT improvement on multi-hue and multi-object tracking," Electrical Engineering and Informatics (ICEEI), 2011 International Conference on , pp.1,6, 17-19 July 2011

[12] C. Hua-Wen, and L. Shu-Duo, Wavelet Analysis and Pattern Recognition, 2007. ICWAPR '07. International Conference 2007, 4 pp. 1585-1588.

Development of a Conceptual Model of E-commerce Adoption for SMEs in Indonesia

Evi Triandini
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
evi_nadine2006@yahoo.com
evi.triandini11@mhs.if.its.ac.id

Arif Djunaidy
Department of Information Systems
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
adjunaidy@is.its.ac.id

Daniel Siahaan
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
daniel@if.its.ac.id

Abstract—**E-commerce adoption provides many benefits for small and medium enterprises. In Indonesia, adoption of e-commerce by SMEs is still in low level because they meet such several problems to adopt e-commerce as difficulty to learn and use e-commerce, to understand the role of e-commerce, etc. Based on the literature review, a conceptual model that is capable of measuring the adoption level of e-commerce is proposed. The model–that consists of both functional and non-functional requirements of an e-commerce–can be used as a framework for raising their e-commerce adoption level.**

Keywords-conceptual model;e-commerce adoption; framework; SMEs

I. INTRODUCTION

Small and Medium Enterprises (SMEs) are small-scale firms that contribute a critical role in an environment characterized by rapid growth. They support industry development and reserve as economic growth drivers[1]. The globalization of the world economy accentuates SMEs as a backbone of the national economy. In Indonesia, SMEs play an important role in reducing the rate of poverty and unemployment. In the middle of 1997, when the crisis occurred in Indonesia, SMEs remained and even tended to grow [2].

In a changing and competing environment, notably in global market, SMEs need to improve their ability as well as to respond to the changes. E-commerce adoption among SMEs have experienced rapid growth in the past few years [3].E-commerce provides opportunities and benefits to organizations of all sizes, particularly to the small-business sector [4].

There were several works related to implementation and adoption of e-commerce for SMEs. Many works have been trying to identify factors influencing e-commerce adoption. Those works are expected to be a reference for implementing successful e-commerce in SMEs[4]–[11]. However, e-commerce adoption by Indonesian SMEs is still in low level. This study found several obstacles faced by SMEs when adopting e-commerce, particularly there were difficulty to learn and use e-commerce related with inadequacy of human resource capability [11]. Other study found that some SMEs have failed to adopt e-commerce because they did not know how to develop activities concerning the adoption of e-commerce in their firms. Moreover, they did not know how to understand a complex role in order to manage their e-commerce [12].

As already mentioned earlier that, although e-commerce provides many benefits for SMEs and there were several factors influencing successful adoption of e-commerce, there is still a problem on how to raise the level of e-commerce adoption in order to gain more benefits regarding each level they will achieve.

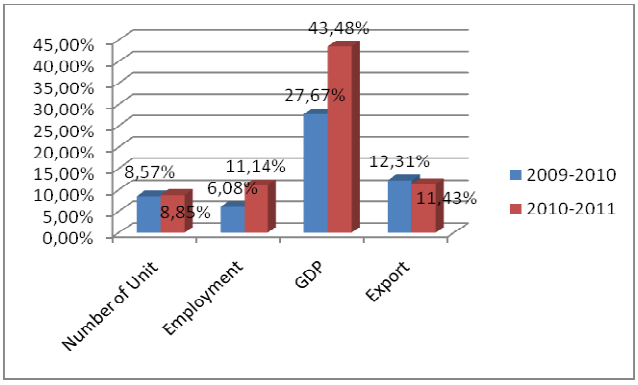
Based on the problem faced by Indonesian SMEs, this paper proposes a conceptual model for e-commerce adoption by SMEs. The proposed model is facilitated with a metric to measure the adoption level. The conceptual model also consists of functional and non-functional requirements which are provided in each e-commerce adoption level. In the future, this conceptual model can be used as a framework to develop e-commerce applications needed by Indonesian SMEs based on their benefits.

II. LITERATURE REVIEW

A. SMEs in Indonesia

SMEs in Indonesia play an important role in reducing the rate of poverty and unemployment in Indonesia economy. A report noted that SMEs contribute to the employment, gross national product (GDP), and export. In Indonesia, Indonesian's Central Bureau of Statics report indicated that the growth and the role of SMEs in Indonesia economy from 2009 to 2010 and 2010 to 2011 were significant, as show in Figure 1. This report indicated that Indonesian's SMEs should be more active

in the global marketplace in order to increase their revenue and advantage competition.



Source: Indonesian's Central Bureau of Statistic

Figure 1. Growth of SMEs Indonesia

B. E-commerce Adoption in SMEs

Many previous studies have examined the e-commerce adoption by SMEs. According to Hong[13], there were three drivers influenced SEMs to migrate toward e-commerce; i.e., technology integration, web functionalities, and web spending. The firms that have more web-compatible technologies are likely to be earlier to adopt e-commerce. In this context, there are several web functionalities that can specify the level of technological capabilities that are definitely related to the adoption of e-commerce [13]. A study on e-commerce adoption in New Zealand found that there were several factors affecting e-commerce adoption, such as innovative and involvement of Chief Executive Officers (CEOs), relative advantage, competition, information intensity of product, and supplier pressure [14]. In order to gain several advantages of e-commerce adoption, a study was conducted for evaluating implementation of e-commerce in Iran. Results of this study identified three factors that evoked positives relationship between the degree of e-commerce adoption and stages of e-commerce development in a firm; i.e., technical, organizational, and inter-organizational [15].

Currently, very limited studies that were already conducted in related with e-commerce adoption in Indonesia (see for example Yulimar[10]; Maryeni[16]). According to Yulimar, there were six factors effecting adoption e-commerce in Indonesia; these are, compatibility, top management support, organizational readiness, external pressure, and perceived benefit. In addition to this, results obtained from Maryeni's study indicated that there were several variables of technological and organizational factors that influenced the adoption of e-commerce by manufacturing SMEs in Indonesia, particularly in West Java province. These variables are relative advantage, compatibility, complexity, observability, planning, infrastructure, and security, IT skills of users, IT knowledge of owners/managers, management support, funding capacity, the required effort and turnover.

C. E-commerce Adoption Level

Rao proposed a four-stage model for the e-commerce development and its implication for SMEs[17]. The model could be used by SMEs to classify their position in compared to other SMEs. The model was developed based on the evolutionary nature of e-commerce. The characteristics of each stage in this model indicate a functional requirement of e-commerce. The model consisted of presence, portals, transaction integration, and enterprises integration [17].

A Study related to e-commerce adoption that measured the level of adoption was conducted by Wahid [11] and Govindaraju [18]. Wahid's study found that, in general, the e-commerce adoption by SMEs in Indonesia (particularly in Yogyakarta region) is still low. This level was measured using Knol and Stroekeng model [19].Govindaraju also found that most Indonesian SMEs participating inhis study are currently still adopt e-commerce at the lower level. There were three factors pursued the adoption of e-commerce such as push forces, man, and source of information. According to Govindaraju, Indonesian SMEs needed the support of their management, educated employee, and more sources of information about e-commerce. Moreover, although the adoption level of most SMEs are currently still at the lower level, they desire to migrate to higher level adoption.

III. THE PROPOSED CONCEPTUAL MODEL

In this study, a modified model of e-commerce adoption developed by Rao were used to measure the adoption level of e-commerce [17]. The initial design of the conceptual model is shown in figure 2. In this regard, a modification was made by involving the functional requirements and other attributes to measure the level of adoption by SMEs.

A survey was conducted to determine functionality requirements of e-commerce, where 15 best e-commerce platforms ranked in the Web Appers sites (http://www.webappers.com) as listed in Table 1. We analyzed and looked for to obtain the required functionalities. The analysis was conducted by determine the unique functionality had each platform. At least 10 platform supported the unique functionality. Based on this analysis, initial functional is

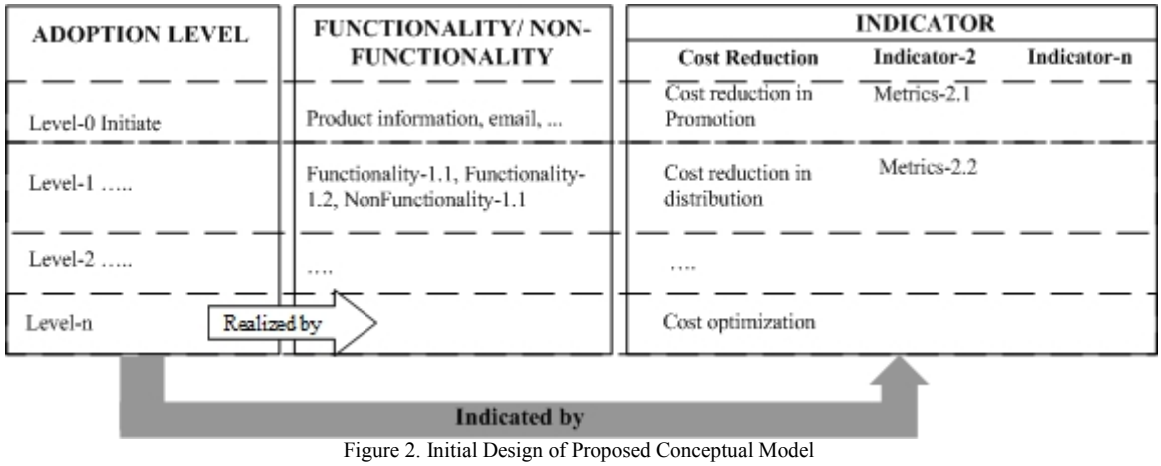


Figure 2. Initial Design of Proposed Conceptual Model

defined as shown in Table 2.

TABLE 1. E-COMMERCE PLATFORMS

ID	Platform	Website Address
ID01	Magento	http://www.magentocommerce.com
ID02	osCommerce	http://oscommerce.com
ID03	OpenCart	http://www.opencart.com
ID04	Spree Commerce	http://spreecommerce.com
ID05	PrestaShop	http://www.prestashop.com
ID06	VirtueMart	http://virtuemaart.net
ID07	Ubercart	http://www.ubercart.org
ID08	Zeuscart	http://www.zeuscart.com
ID09	Afcommerce	http://www.afcommerce.com
ID10	Zen Cart	http://www.zen-cart.com
ID11	SimpleCart js	http://www.simplecartjs.org
ID12	Tomato Cart	http://www.tomatocart.com
ID13	CuberCart	http://www.cubercart.com
ID14	RokQuickCart	http://www.rockettheme.com
ID15	StoreSprite	http://www.storesprite.com

TABLE 2. FUNCTIONALITY REQUIREMENTS

No.	Functionality	Website ID
1	Information product	ID01-ID15
2	Checkout	ID01-ID15
3	Payment	ID01-ID15
4	Shipping	ID01-ID15
5	Promotion	ID01-ID10, ID11, ID13, ID15
6	Searching	ID01-ID10, ID11, ID13, ID15
7	Search Engine Optimization (SEO)	ID01-ID10, ID11, ID13, ID15
8	Provide security	ID01-ID10, ID11, ID13, ID15
9	Customer account	ID01-ID10, ID11, ID13, ID15
10	Management order	ID01-ID10, ID12, ID13, ID15
11	Reporting and analytic	ID01-ID08, ID12, ID13, ID15
12	Multi language and currency	ID01, ID02, ID05, ID06, ID08, ID09, ID10, ID12, ID13, ID15

Non-functional requirements define the overall qualities or attributes of the resulting model[20]. There are a number of quality models that can be used as a basis for developing the software quality. The WebQual 3.1 [21] and the international standard ISO 9241-151 [22] were proposed to determine non-functionality requirements of e-commerce. The WebQual provides 12 aspects that are related to functionality of a Website; these are, informational fit-to-task, tailored information, trust, response time, ease of understanding, intuitive operations, visual appeal, innovativeness, emotional appeal, consistent image, on-line completeness, and relative advantage[21]. On the other hand, ISO 9241-151 provides guidance on the Web user interfaces with the aim of increasing usability. ISO 9241-151 is structured into five major areas: high-level design aspects, conceptual content model, content object and functionality, navigation and search, and content presentation[22].

The analysis was conducted toward Webqual 3.1 and ISO 9241-151. Non-functionalities requirement of e-commerce were determined based on the need of e-commerce platform, as shown on Table 3.

TABLE 3. NON-FUNCTIONALITY REQUIREMENTS (PROSESNYA, ATRIBUT DAN KUALITAS)

No.	Non-Functionality	Reference
1	Keeping content up-to-date	[22]
2	Communication with website owner	
3	Accepting online user feedback	
4	Providing a business policy statement	
5	Providing privacy policy statement	
6	User control of personal information	
7	Information fit-to-task	[21]
8	Trust	
9	Response times	
10	On-line completeness	
11	Relative advantage	

For the purpose of defining indicators required for quantifying the adoption level of e-commerce, a set of comprehensive literature surveys were performed. First, it was clear that the adoption of e-commerce provide opportunities such as increased number of customer, penetration of new market, faster communication with customer, competitive advantage over competitor, easy access to customer network, cost advantage and short delivery times [23]. In New Zealand, for example, 61.4% SMEs already used e-commerce to support their business processes. They gained benefits from adopting e-commerce such as effective advertising and brand building, increased customer based, increased sales, cost savings, increased profits, and better purchasing terms [24].

Second, it could be assured that SMEs were adopting e-commerce to improve their performance. The benefits gained from adopting e-commerce were increased revenues and reduced costs [25]. For examples, a study on implementing e-commerce by SMEs in Australia and Singapore was found the benefits such as increased sales, improved profitability, reduce costs associated with inventory, procurement and distribution, improving quality of service and guarantee competitive [26]. Moreover, SMEs in China and Malaysia were concerned with e-commerce related with the benefits their gained such as increase ability to compete, increase sales, reducing transaction costs and enhancing the corporate image [27], [28].

Third, in the context of Indonesia, only a few studies were done related e-commerce adoption. Those studies described several benefits for adopting e-commerce such as increasing penetration market, increasing advertising, reducing costs, increasing profit and increasing customer services [29].

Based on the above literature surveys, it is envisaged that the adoption of e-commerce will provide many opportunities and benefits for SMEs. However, although there are works in the literature which present implementation and adoption of e-commerce, there are no work that define the benefit of e-commerce adoption as an indicator to measure the benefit gained by SMEs. In this study, those benefits of e-commerce adoption are used as a basis for defining some indicators for quantifying the adoption level of e-commerce. Table 4 shows those proposed indicators.

TABLE 4. INDICATORS OF E-COMMERCE BENEFITS

No.	Attribute	References
1	Costs reduction	[23]–[27], [29], [30]
2	Global market	[23], [25], [26], [29], 30]
3	Market penetration	[23]
4	Increasing revenue	[24], [25], [29]
5	Increasing customer services	[11], [23], [24], [29]
6	Improving information availability	[24], [27]–[29], [31]
7	Rapid time to market	[23]

It is important to note that this study is part of a bigger research work, which is still in progress. In order to establish the

complete view of the conceptual model as a result of this study, data collection from Indonesian SMEs and e-commerce's experts will be conducted. A set of questionnaires will be delivered to a sample of SMEs that will be selected from the Ministry of Cooperative and Small and Medium Enterprises of Indonesia through its website (https://www.indonesian-product.biz).

Data collected from SMEs will be first analyzed using the factor analysis. The aim of this method is to find the relationships among a large number of variables by defining sets of variables that are highly interrelated [34]. Results of data processing will be used to create mapping between indicator of e-commerce adoption benefit and requirement of e-commerce. A hierarchy clustering method will be then employed to group data objects into a hierarchy of clusters[32]. This scheme was similar to the work performed by Jalozie, where a hierarchical clustering method was used to identify e-commerce business models [33]. Results of such a clustering procedure will be used to determine the number of stages of the e-commerce adoption model of SMEs.

IV. Conclusion and Future Reseach

A conceptual model of e-commerce adoption, which is equipped with a metric to measure the level of e-commerce adoption has been proposed. The proposed list of both functionality and non-functionality requirements of e-commerce will be used to characterize each level of the e-commerce adoption. Together with the proposed list of indicators of e-commerce benefits, the model can be used to assess the adoption level of e-commerce by SMEs. In order to acquire the numerous benefits of e-commerce adoption, this model can be utilized as a framework to guide SMEs to move from their existing adoption level to the higher-level ones. Therefore, our future research will focus on development of e-commerce adoption framework for SMEs based on this study. A survey instrument consisting of a set of questionnaires related to the benefits' indicators of e-commerce for each level of the model will be developed. In this future research, empirical tests will also be employed in order to validate the model being developed.

References

[1] L.-M. Chang, S.-I. Chang, C.-T. Ho, D. Yen, and M.-C. Chiang, "Effects of IS Characteristics on The E-Business Success Factors for Small- and Medium-Sized Enterprises," *Computers in Human Behavior*, vol. 27, pp. 2129–2140, 2011.

[2] T. Partomo, "Usaha Kecil Menengah dan Koperasi," Faculty of Economics University of Trisakti, Jakarta, 9, 2004.

[3] Y. Wang, "Achieving Market Responsiveness: The Enabling Role of IT and E-Commerce Adoption in SMEs," in *2010 International Conference on Management and Service Science (MASS)*, 2010, pp. 1 –4.

[4] N. Al-Qirim, *Electronic Commerce in Small to Medium-Sized Enterprises: Frameworks, Issues and Implications,*. USA: IDEA Group Publishing, 2004.

[5] J. Bao and X. Sun, "A Conceptual Model of Factors Affecting e-Commerce Adoption by SMEs in China," in *2010 Fourth International Conference on Management of e-Commerce and e-Government (ICMeCG)*, 2010, pp. 172 –175.

[6] J. Tan, K. Tyler, and A. Manica, "Business-to-business adoption of eCommerce in China," *Information & Management*, vol. 44, pp. 332–351, 2007.

[7] S. Nasco, E. Toledo, and P. Mykytyn, "Predicting Electronic Commerce Adoption in Chilean SMEs," *Journal of Business Research*, vol. 61, pp. 697–705, 2008.

[8] K.-Y. Oh, D. Cruickshank, and A. R. Anderson, "The Adoption of e-trade innovations by Korean small and medium sized firms," *Technovation*, vol. 29, pp. 110–121, 2009.

[9] M. Fathian, P. Akhavan, and M. Hoorali, "E-readiness assessment of non-profit ICT SMEs in a developing country: The case of Iran.," *Technovation*, vol. 28, pp. 578–590, 2008.

[10] V. Yulimar, "Analysis of factors affecting the adoption of electronic commerce and company (Study on small and medium company in Indonesia)," Universitas Diponegoro, Semarang, 2006.

[11] F. Wahid and L. Iswari, "Information Technology Adoption by Small and Medium Enterprises in Indonesia," in *The National Seminar on Information Technology Application*, Yogyakarta, 2007, pp. 75–79.

[12] E. Gide, "The Key Strategies for E-commerce Success Service SMEs in China", IEEE hal 229 -234," 2006.

[13] W. Hong and Ke. Zhu, "Migrating to internet-based e-commerce: Factors affecting e-commerce adoption and migration at the firm level," *Information & Management*, vol. 43, pp. 204–221, 2006.

[14] N. Al-Qirim, "The adoption of eCommerce communications and applications technologies in small businesses in New Zealand," *Electronic Commerce Research and Applications*, vol. 6, pp. 462–473, 2007.

[15] S. Elahi and A. Hassanzadeh, "A framework for evaluating electronic commerce adoption in Iranian companies," *International Journal of Information Management*, vol. 29, no. 1, pp. 27–36, Feb. 2009.

[16] Y. Y. Maryeni, R. Govindaraju, B. Prihartono, and I. Sudirman, "Technological and organizational factors influencing the e-commerce adoption by Indonesian SMEs," in *2012 IEEE International Conference on Management of Innovation and Technology (ICMIT)*, 2012, pp. 436 – 441.

[17] S. Rao and G. Metts, "Electronic Commerce Development in Small and Medium Sized Enterprises A Stage Model and Its Implications," *Business Process Management Journal*, vol. 9, pp. 11–32, 2003.

[18] R. Govindaraju and D. R. Chandra, "E-commerce adoption by Indonesian small, medium, and micro enterprises (SMMEs): Analysis of goals and barriers," in *2011 IEEE 3rd International Conference on Communication Software and Networks (ICCSN)*, 2011, pp. 113 –117.

[19] W. H. Knol and J. H. . Stroeken, "The Diffusion and Adoption of Information Technology in Small- and Medium-sized Enterprises through IT Scenarios," *Technology Analysis & Strategic Management*, vol. 13, 2001.

[20] A. Cremers and S. Alda, *Organizational Requirements engineering*. .

[21] E. Loiacono, "WebQual™: A Measure of Web Site Quality," *International Journal of Electronic Commerce*, vol. 11, pp. 51–87, 2007.

[22] *International Standard Organization, Egronomics of Human System Interaction – Part 151□: Guidance on World Wide Web User Interface*, First. ISO, Geneva, 2008.

[23] E. Santarelli and S. D'Altri, "The Diffusion of E-commerce among SMEs: Theoritical Implications and Empirical Evidence," *Small Business Economics*, vol. 21, pp. 273–283, 2003.

[24] P. McCole, "A Profile of Adopters and Non-adopters of eCommerce in SME Professional Service Firms," *Australasian Marketing Journal*, vol. 13, pp. 36–48, 2005.

[25] D. Johnston, M. Wade, and R. McClean, "Does e-Business Matter to SMEs? A Comparason of the Financial Impacts of Internet Business Solutions on European and North American SMEs," *Journal of Small Business Management*, vol. 45, pp. 354–361, 2007.

[26] S. Chong, "Success in Electronic Commerce Implementation: A Cross Country Study of small and Medium-Sized Enterprises," *Journal of Enterprise Information Management*, vol. 21, no. 5, pp. 468–492, 2008.

[27] F. Ying and L. Fengli, "Empirical Research on Interaction between Electronic Commerce Adoption and Effect Evaluation in China SMEs," presented at the International Conference on Information Management, Innovation Management aand Industrial Engineering, 2012, pp. 5–8.

[28] M. Shaharudin, M. Omar, S. Elias, M. Ismail, S. Ali, and M. Fadzil, "Determinants of electronic commerce adoption in Malaysian SMEs' furniture industry," *African Journal of Business Management*, vol. 6, pp. 3648–3661, 2012.

[29] A. Eva, "Persepsi Penggunaan Aplikasi Internet untuk Pemasaran Produk Usaha Kecil Menengah," in *The National Seminar on Information Technology Application*, Yogyakarta, 2007, pp. 13–16.

[30] E. Turban, *Introduction to Electronic Commerce*, 2nd ed. New Jersey: Pearson Prentice Hall, 2009.

[31] R. Ramanathan, U. Ramanathan, and H.-L. Hsiao, "The Impact of e-commerce on Taiwanese SMEs: Marketing and Operation Effects," *International Journal Production Economics*, vol. 140, pp. 934–943, 2012.

[32] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. USA: Morgan Kaufmann, 2012.

[33] I. Jalozie, H. . Wen, and H. . Huang, "A Framework for Selecting E-Commerce Business Model," *Unpublish*.

[34] J. Hair, W. Black, B. Babin, and R. Anderson, *Multivariate Data Analysis*, 7th ed. .

Development of Marine Radar Signal Acquisition and Processing System

Asif Awaludin, Ginaldi Ari Nugroho
Center for Atmospheric Science and Technology
National Institute of Aeronautics and Space (LAPAN)
Bandung, Indonesia
asif_aw@yahoo.co.id

Sahirul Alam, Dwi Fadila K, Rudy Yuwono
Department of Electrical Engineering
Brawijaya University
Malang, Indonesia
shah.teub06@gmail.com

***Abstract*—Marine radar commonly used for ship navigation. Nowadays, utilization of this radar to another application has expanded. For further applications, several marine radars have some limitations such as monochrome display and require new signal acquisition and processing. This research developed marine radar signal acquisition and processing system based on Furuno x-band marine radar 1932 Mark-2 . The developed tool consist of marine radar signal conditioner, ADC and marine radar signal processing using Matlab. This signal processing system display provide signal echo strength visualization, gain control, sea clutter control and rain clutter control. The simulation test was conducted to test the signal processing system. The result was the GUI display at full and half gain are quite similar with marine radar display unit view. The sea and rain clutter control test results were relevant with its function in reducing more close echoes rather than far echoes. While the time required for creating an image is still need for improvement.**

***Keywords*—marine radar; signal acquisition; signal processing; clutter control.**

I. INTRODUCTION

Marine radar is a radar employed to provide bearing and distance of ships and land targets in vicinity from own ship for collision avoidance and navigation at sea. Recently, x-band marine radar utilized also for marine wave observation [1], local area weather radar [2] and bird migration study [3]. For further applications, several marine radars have some limitations such as monochrome display and require new signal acquisition and processing. It is also used different clutter removal system compared to some application mentioned before.

This research was mainly motivated by the following two researches. The first was Li and Zhuang on development of navigation radar signal acquisition based on embedded system [4]. It consist of voltage conversion, signal filtering, ADC, amplifier, CPLD, and ARM9 microprocessor. The second was Wagner on real-time marine radar simulation using MacBook Air 2010 which running on 2.13 GHz Intel Core 2 Duo CPU, 4 GB RAM, and 256 MB NVidia GeForce 320M [7]. It was described that its gain test took time 3.1 s, rain control test took 3.2 s, sea control test took 3.0 s, and complex environment test took 3.8 s. Here in this paper, we bring both ideas together and construct a marine radar signal acquisition and processing system. The objective was to provide basic marine radar signal acquisition and processing tool which can be used for further marine radar development or another signal processing applications such as local area weather radar. Furuno x-band marine radar 1932 Mark-2 was used in this research which has operating frequency 9410 MHz with peak output power 4 kW and using 3.5 ft centre-fed waveguide slotted array antenna. Rotation speed of the antenna is 24 rpm.

The developed tool consisted of marine radar signal conditioner, ADC and marine radar signal processing using Matlab. This signal processing system display provide signal echo strength visualization, gain control, sea clutter control and rain clutter control.

II. SIGNAL ACQUISITION

The marine radar main outputs are heading, bearing, trigger and video signal. This four signals can be obtained from one of headers in marine radar display unit circuit board. Fig.1 shows the four signals displayed on oscilloscopes. These signals must be conditioned so as to be compatible with the ADC input range, hence a signal conditioner was built. The signal conditioner consist of active LPF for video signal and voltage converter for heading, bearing and trigger signal [4].

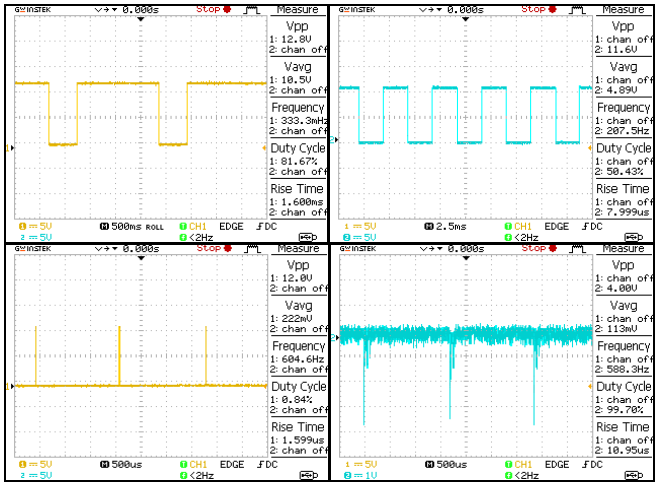


Fig 1. Marine radar signals displayed on oscilloscopes. Top left picture is heading signal, top right is bearing signal, bottom left is trigger signal and bottom right is video signal.

A. Active Low Pass Filter Sallen-key

Video signal is an analog signal containing echo and noise. Hence before it can be processed, the signal must be filtered to reduce its noise. A second order active low pass filter Sallen-key is employed for the signal filtering. In Butterworth low pass filter design with attenuation -20 dB/decade/order, the value of *n* and *m* are 3,3 and 0,229 respectively [5]. As the cut off frequency can be obtained from the signal rise time, the value of *R* and *C* can be calculated [5]:

$$f_c = \frac{1}{2\pi RC\sqrt{mn}} \quad (1)$$

The filter topology employed in this design is Sallen-key filter. Here R_1 value is mR, R_2 is R, C_1 value is nC and C_2 value is C. Where subscript 1 is for first stage and subscript 2 is for second stage in Sallen-key topology [5]. As the result, the voltage transfer function of the unity gain active low pass filter can be determined [5]:

$$H(s) = \frac{1}{s^2(R_1R_2C_1C_2) + s(R_1C_1 + R_2C_1) + 1} \quad (2)$$

The transfer function represents a second-order filter which has two poles and no zeros.

B. Voltage Conversion

Fig. 1 describe that heading, bearing and trigger signals have voltage range from 0 to 12 V. As the ADC has input range between -5 V and 5 V, it is necessary to convert the signal voltage into 0 to 5 V or -5 V to 0. The voltage conversion circuit design was able to use voltage divider circuit or if isolation between radar display unit circuit board and voltage conversion circuit is considered, an optocoupler circuit can be utilized as a voltage divider.

C. Signal Sampling

The radar signals should be converted to digital data for data processing using analog to digital converter. In Fig. 1, the trigger signal, which indicate the radar beam being transmitted, has frequency about 600 Hz. Since the video signal will be sampled 4000 times every trigger periode to obtain 4000 data, it will need an ADC with minimum sampling frequency 2.4 MHz. Therefore, 10 bit ADLINK PCI-9810 with maximum frequency sampling 20 MHz was employed. It has four input channels and an external digital trigger. The installment of this ADC is by attaching it on PCI slot of computer and the running control was done using Matlab.

Computer used in this signal sampling and processing application is low end in terms of computation and graphics performance by today standards. It run on Intel Core™2 Quad Processor 2,66 GHz, 3 GB RAM, and Intel GMA 3100 384 MB Integrated.

III. SIGNAL PROCESSING

In radar, the most important data for target detection are range and angle which yield the polar data. These data can not be achieved directly but should be extracted from heading, bearing, trigger and video data. Range (r) can be determined using two-way time delay (Δt) of echo in video data [6]:

$$r = \frac{c\Delta t}{2} \quad (3)$$

where c is speed of light. The angle of the polar data is extracted from heading, bearing and trigger data. Each heading period is equal with one rotation of scanner antenna in 360^o. Number of bearing periods in one heading period represents angle value of 360^o divided by their amount of numbers. Each bearing period consist of three until ten trigger signal period according to the pulse width as well.

If sampling interval is smaller than radar pulse width, there will be several polar data in each range resolution. Here, range resolution is determined using Eq. (3), where the two-way time delay is the pulse width. Hence, all polar data in each range resolution should be averaged.

Strength of the echo is indicated by its voltage amplitude in video data. The amplitude it self is not merely from a target but can be from clutter or combination of target plus clutter as well. Therefore, radar image composed by polar data need adjustment control to remove the clutter effect. The important clutter that should be removed in marine radar for ship navigation are sea wave and rain clutter. The adjustment control needed for the clutter removal are gain, sea clutter and rain clutter control.

A. Gain Control

In this digital signal processing, gain control method is different but has the same purpose with amplifier amplification method. It uses threshold value to determine which target should be drawn on the radar image. An amplitude value above threshold will be drawn otherwise will be neglected. Gain control adjust the threshold line to allow fewer or more targets to become visible. The higher the gain, the lower the threshold will become and allowing more targets and clutter to be displayed.

B. Sea Clutter Control

Sea clutter represents echoes from nearby waves that are strong enough to be displayed on the radar display. These echoes can be strong enough that they obscure main targets such as ships and coastlines. Sea clutter control is employed to remove most of this clutter and allow only main targets to be displayed. The characteristic of sea clutter is that it appears strong near the radar itself and quickly diminishes with distance. Sea clutter control use this fact to reduce only echoes that are close to the radar. Similar with gain control, the sea control adjusts the threshold line which determines if a signal is drawn on the radar display. But instead of adjusting the height of the entire line, changes in height is dependent on the

range. The curve applied by the sea clutter control is the inverse exponential where the control influences the exponent [7]:

$$T = \frac{Ct}{R^s} \quad (4)$$

with T is the new value of threshold t at range R, given a sea control setting s. C is a constant that differs per radar system and determines how strong the effect is. The minimum sea control setting is usually 0, equating to no change in the threshold value. The maximum setting differs per radar system.

C. Rain Clutter Control

A local rain shower that reflects radar beam will generate more echoes that obscure the main targets. The radar display will show rain showers as haze obscuring the real targets in the area. A lower gain would reveal real targets in a rain shower, but can simultaneously resulting very weak targets outside the rain shower become invisible. Preferably the gain should only be lower inside the rain shower itself. In resolving this problem, the rain control utilize the absolute value of the derivative of the signal. Given that the original threshold is defined by the function f(x), the threshold with rain control applied $f_r(x)$ can be defined [7]:

$$f_r = \left| f'(x) \right| \quad (5)$$

This derivate curve shows peaks whenever the original curve goes up or down and stays zero whenever it remains constant. The result is that large areas of uniform clutter only show up when they start and end, and that targets within the clutter can be clearly reappear. The disadvantage of applying the rain control is that it can make very large echoes or echoes that gradually increase in intensity appear to be weak.

IV. Results And Discussion

The result showed that the signal processing system utilizing Matlab GUI was able to display radar detection. The GUI was developed to resemble marine radar display. Therefore it has some features such as echo strength visualization, range adjustment, range unit selection (nautical mile or kilometers), auto gain check box, gain control adjustment, sea clutter adjustment and rain clutter adjustment.

The marine radar system was installed at the top of the 4th floors of LAPAN's office building in the middle of Bandung city. The developed system was tested by detecting target around the area and saving the data for simulation test. The simulation was conducted to test the gain control, sea clutter control and rain clutter control.

In gain control test, GUI display of full gain has many weak echoes appeared compared with half gain. Meanwhile, the GUI display at full and half gain are quite similar with marine radar display unit view. These result was described by Fig. 2.

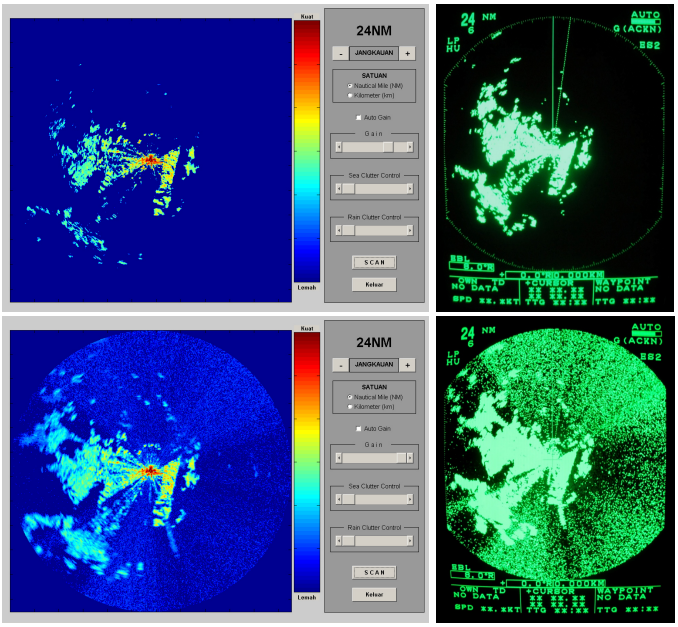


Fig 2. The GUI display (right) compared with marine radar display (left). Half gain view of both display (top) and full gain view of both display (bottom).

The sea clutter control test was conducted by data simulation for quarter, half, three quarters and full scale of sea clutter control. The results showed that by increasing the control scale, echoes closer to the radar were more diminished rather than echoes that far away from the radar, as mentioned by Fig. 3. This result is relevant with sea clutter control function in reducing echoes that are close to the radar.

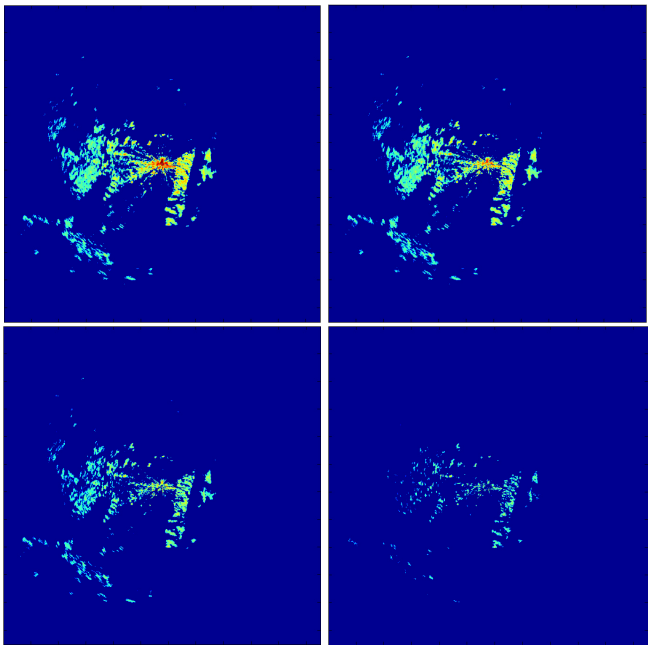


Fig 3. Sea clutter control test results. Quarter scale (top left), half scale (top right), three quarters scale (bottom left) and full scale (bottom right).

The rain clutter control test was conducted by data simulation for quarter, half, three quarters and full scale of sea clutter control as well. The results showed that rain clutter

control made very large echoes appear to be weak, as described by Fig. 4 compared with Fig. 3. This result is relevant with rain clutter control disanvantage effects. It is also can be seen that by the increasing of the control scale, echoes close to radar were more diminished than echoes far to radar.

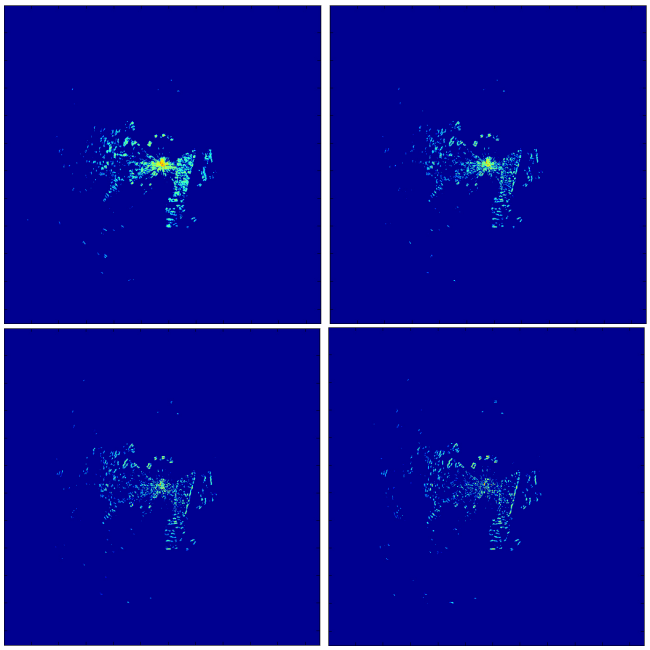


Fig 4. Rain clutter control test results. Quarter scale (top left), half scale (top right), three quarters scale (bottom left) and full scale (bottom right).

The time required for displaying image of full gain was 29 s. While both sea and rain clutter control were equally took time 30 s. The time was much longer than Wagner's real-time marine radar simulation took. The difference between the two is this signal processing system has polar data extraction from marine radar raw signals, while Wagner's real-time marine radar simulation does not have.

According to the time for creating an image, it is obvious that our signal processing still need for improvement, because the marine radar made a revolution every 2.5 s. Therefore, this results still only suitable for marine radar applications which need update data at least every 30 s.

V. CONCLUSION

Has been developed a marine radar signal acquisition and processing system based on Furuno marine radar 1932 Mark 2. It has conducted gain control, sea clutter control, and rain clutter control successfully. But its time required for creating an image was 29 s, much longer that the time needed by the marine radar for each revolution. Therefore this results still need for improvement.

ACKNOWLEDGMENT

The marine radar used in this research is belong to Indonesia National Institute of Aeronautics and Space (LAPAN). The authors would like to thank to Dr. Joko Suryana for giving some inputs for the system development.

REFERENCES

[1] J.P.Y Maa and H.K. Ha, "X-band radar wave observation System," Project report. Minerals Management Service. US Department of the Interior, 2005.

[2] L. Pedersen, I. Zawadzki, N.E. Jensen, and H. Madsen, "Assessment of QPE results from 4 kW x-band local area weather radar (LAWR) evaluated with s-band radar data," The Fifth European Conference on Radar in Meteorology and Hydrology, 2008.

[3] N.E. Jensen, L. Pedersen, and H. Skov, "Development and application of x-band radar systems for automatic detection and registration of migrating birds," 34th Conference on Radar Meteorology, 2009.

[4] S. Li and X. Zhuang, "Navigation Radar Signal Acquisition and Measurement System," The Ninth International Conference on Electronics Measurement and Instruments, 2009.

[5] J. Karki, "Active Low Pass Filter Design," Texas Instruments Corporated Application Report SLOA049A, 2000.

[6] B.R. Mahafza, "Radar Signal Analysis and Processing Using Matlab," Taylor and Francis Group LLC. Boca Raton. USA, 2009.

[7] P. Wagner, "Real-time realistic radar simulation," Master's thesis. Department of Mathematics and Computer Science. Eindhoven University of Technology, 2011.

DIGITAL COLOR IMAGE ENCRYPTION USING RC4 STREAM CIPHER AND CHAOTIC LOGISTIC MAP

Riah Ukur Ginting[1]
Facultyn of Mathematics and Natural Science
University of Sumatera Utara
Jl. Bioteknologi I Kampus USU Pd Bulan Medan
E-Mail: riahukur@gmail.com

Rocky Yefrenes Dillak[2]
Electrical Engineering Deparment, Kupang State Polytechnic
Jl. Adisucipto, Penfui, Kupang
E-Mail: rocky_dillak@yahoo.com

***Abstract*—Doing a digital image transmission over internet need a secure protection againts ilegall copying. Unfortunately, many current data encryption methods such as DES, RSA, AES, and other only suitable for test data, but not for digital image. In this paper, we propose new secure algorithm for image encryption, which based on RC4 stream cipher algorithm and chaotic logistics map. The proposed algorithm works as follows : (i0 converting the external key into initial value, (ii) using he initial value to generate a key stream using chaotic logistic map function, and (iii) processing a permutation and the result is then XOR-ed with bytes stream of digital image. The experiment results show that the proposed algorithm (i) is able to make the cipher-image can not be visually identified, (ii) can eliminate the statistical correlation between the plain-image and cipher-image, (iii) is very sensitive to small changes of key, and (iv) has no change in image contents (lossless encryption) during encryption and decryption process which is indicated by the hash value (MD5) of plain-image has the same hash value (MD5) wth decrypted image.**

Keywords—RC4 Stream Cipher; Chaotic Logistics Map; Digital Image Encryption.

I. INTRODUCTION

Digital image encryption is one of the secure methods to protect digital images against illegally copying when transmitted over unsecure channel. Unfortunately, according to [1][2][3] many popular encryption methods such as DES, RSA, AES, and others only work well for plaintext but not for digital images. In most of the natural digital images, the values of the neighboring pixels are strongly correlated (i.e. the value of any given pixel can be reasonably predicted from the values of its neighbors). This unique characteristic lead to huge changes of each pixel of plain-image is not going to drastically reduce the quality of the cipher-image which will makes the content of cipher-image can still be visually identified by human.

One solution to overcome these problems is using chaotic system (i.e. chaotic logistic map) in a cipher because according to [4][5], chaos is very sensitive to a small changes in the initial value and will produce the same effect as diffusion and confusion.

In this paper, we proposed a digital color image encryption using the combination of RC4 stream cipher and chaotic logistic map function (CLM for short. In the rest of the paper, we will use CLM and chaotic logistic map interchangeably). The proposed algorithm works as follows: (i) convert the external key into initial value X_0, then (ii) use that initial value X_0 on CLM function to generate pseudo random numbers, and the last step is (iii) XOR-ed the byte streams of plain-image with a stream of pseudo random numbers when doing an encryption process (or XOR-ed the byte streams of cipher-image with a stream of pseudo random numbers when doing an decryption process).

This paper is organized as follows. In Section 2, we discuss existing image encryption methods which based on chaotic logistic map. In Section 3, we present our proposed encryption method. In Section 4, we evaluate the performance of our proposed encryption method. In Section 5, we give a conclusion.

II. RELATED WORKS

Many attempts have been made in the past to encrypt digital images. In [1][6], they use two CLM functions, in which the first CLM is used to generate 24 pieces of real numbers which is then converted into integer form. Furthermore, 24 pieces of integer is used to generate initial value X_0 for the second CLM, which is used to perform image encryption process. Different approaches tend to do in [7], where they use a combination of CLM function and the genetic algorithm to encrypt the digital color image. CLM is used to generate 4 pieces of chaotic sequences which is then converted into 4 pieces of key stream. The generated key streams are then used to control the process of crossover and mutation. The use of CLM function is also used by [8] [9]. In [8], a CLM function is used to generate random numbers which will then be summed with the byte stream of the image. The result is then performed modulo 256. While in [9], the combination of cipher block chaining (CBC) method and the CLM function is used to encrypt the image. The main idea is

doing encryption on 4-bit of most significant bit (MSB) at each pixel and then operate them in CBC mode.

III. THE PROPOSED ENCRYPTION METHOD

The proposed encryption method should satisfy the following goal, where I be an image, E is the proposed encryption method, and D is the proposed decryption method:

Lossless: The encryption process has to be reversible, with perfect reconstruction of the image, D(E(I)) = I.

Secure: The proposed method has to be resistant to any known attack.

Low-complexity: The proposed method has to be based on low-cost operations.

The Structure of Proposed Encryption Method

Structure of the proposed encryption method (as shown in Figure 1) consist of three units, such as: (i) external key to initial value X_0 converter unit, (ii) CLM function unit, and (iii) permutation unit. The first unit will convert external key into initial value using Equation (3), (4), (5), (6), (7) and (8). The second unit (CLM unit) is using to generate 255 byte of key streams at array U[i] (as shown in Step 4 of encryption process), and the third unit (permutation unit) will swap array S[i] with array U[i] then the result will XOR-ed with byte streams of image (as shown in Step 5 and Step 6 of encryption process).

In this method, we assume that digital image is a byte streams (we do not encrypt the header part of image (*.BMP) but only the pixel part)[6].

Chaotic Logistic Map (CLM) Function

Chaos is a ubiquitous phenomenon existing in deterministic nonlinear systems that exhibit extreme sensitivity to initial conditions and have random like behaviors. Chaotic logistic map is one of the popular chaotic systems. Consider a CLM function in the general form of

$$X_{n+1} = \lambda X_n(1 - X_n) \quad (1)$$

where λ is a control parameter on the interval $\lambda = [0,4]$ and X_n is real number on the interval $X_n = [0,1]$. This system is said to be chaotic if λ has a value on the interval $\lambda = [3.569955672,4]$. In this paper, we use $\lambda = 4$ so the complate formula is:

$$X_{n+1} = 4X_n(1 - X_n) \quad (2)$$

Encrypt ion Algorithm

The proposed encryption method has 6 steps, such as:

Step (1). External key has 16 ASCII characters in length where every character K_i consist of 8-bit.

$$K = K_1, K_2, \ldots, K_{16} \text{ (ASCII code)} \quad (3)$$

Step (2). For every K_i, convert them into bit streams B_0.

$$\begin{aligned} B_0 = {} & K_{11} \ldots K_{18} K_{21} \ldots K_{28} K_{31} \ldots K_{38} \\ & K_{41} \ldots K_{48} K_{51} \ldots K_{58} K_{61} \ldots K_{68} \\ & K_{71} \ldots K_{78} K_{81} \ldots K_{88} \end{aligned} \quad (4)$$

$$\begin{aligned} B_1 = {} & K_{91} \ldots K_{98} K_{101} \ldots K_{108} K_{111} \ldots K_{118} \\ & K_{121} \ldots K_{128} K_{131} \ldots K_{138} K_{141} \ldots K_{148} \\ & K_{151} \ldots K_{158} K_{161} \ldots K_{168} \end{aligned} \quad (5)$$

where every K_{ij} from Equation (4) and (5) has binary representation (0 or 1), i refers to character position (i=1,2,...,16) and j referes to bit position of character (j=1,2,...,8).

Using binary representation K_{ij}, the real number X_{01} and X_{02} will be counted.

$$\begin{aligned} X_{01} = {} & (K_{11} \times 2^0 + \cdots + K_{18} \times 2^7 + \\ & K_{21} \times 2^8 + \cdots + K_{28} \times 2^{15} + \\ & K_{31} \times 2^{16} + \cdots + K_{38} \times 2^{23} + \\ & K_{41} \times 2^{24} + \cdots + K_{48} \times 2^{31} + \\ & K_{51} \times 2^{32} + \cdots + K_{58} \times 2^{39} + \\ & K_{61} \times 2^{40} + \cdots + K_{68} \times 2^{47} + \\ & K_{71} \times 2^{48} + \cdots + K_{78} \times 2^{55} + \\ & K_{81} \times 2^{56} + \cdots + K_{88} \times 2^{63}) / 2^{64} \end{aligned} \quad (6)$$

$$\begin{aligned} X_{02} = {} & (K_{91} \times 2^0 + \cdots + K_{98} \times 2^7 + \\ & K_{101} \times 2^8 + \cdots + K_{108} \times 2^{15} + \\ & K_{111} \times 2^{16} + \cdots + K_{118} \times 2^{23} + \\ & K_{121} \times 2^{24} + \cdots + K_{128} \times 2^{31} + \\ & K_{131} \times 2^{32} + \cdots + K_{138} \times 2^{39} + \\ & K_{141} \times 2^{40} + \cdots + K_{148} \times 2^{47} + \\ & K_{151} \times 2^{48} + \cdots + K_{158} \times 2^{55} + \\ & K_{161} \times 2^{56} + \cdots + K_{168} \times 2^{63}) / 2^{64} \end{aligned} \quad (7)$$

Furthermore, real number X_{01} in Equation (6) and X_{02} in Equation (7) is used to create initial value X_0.

$$X_0 = (X_{01} + X_{02}) \bmod 1 \quad (8)$$

Step (3). The entries of array S are set equal to the values from 0 through 255 in ascending order; that is; S[0] = 0, S[1] = 1,...,S[255] = 255. These preliminary operations can be summarized as follows:

```
for( i= 0;i <= 255; i++){
    S[i] = i;
}
```

Step (4). To strengthen CLM against attacks, we adopt the approach taken by [8][9][0] for eliminating the correlation between each X_n values of chaos by generating X_n value after a certain number of iterations. In this paper, we determine the number of iterations by taking two digits after the dot (decimal). For example, the initial value X_0 is 0.937696878979928 then the number of iterations required to get the first value of chaos X_1 is 93, thus after 93^{rd} iterations, X_1 value will be 0.8080204084200282. We can say

that the value of X_n which obtained at the end of iteration will act as a new "X_0" to calculate X_{n+1} and so on.

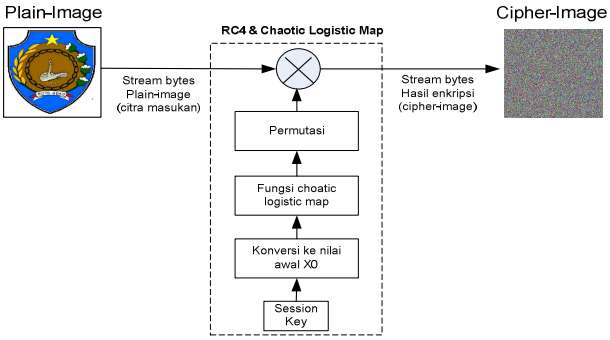


Fig 1. Structure of the proposed encryption method

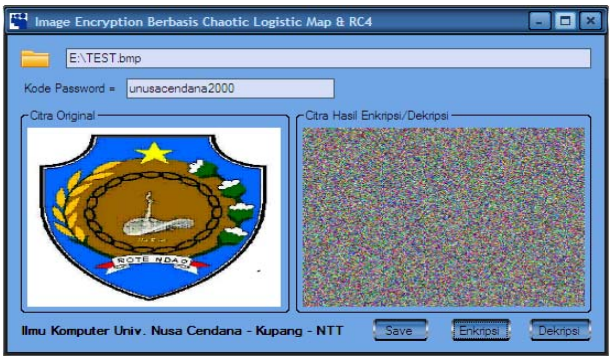


Fig 2. Image encryption result, plain-image (left side) and cipher-image (right side), using external key = ”unusacendana2000”

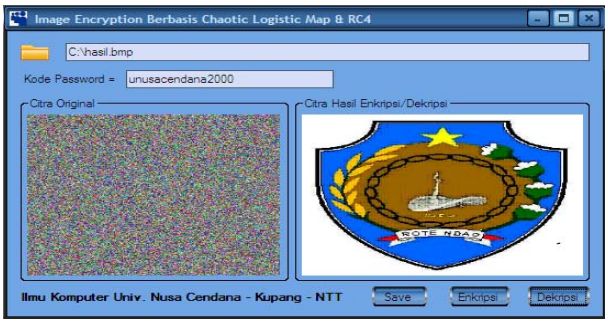


Fig 3. Image decryption result, cipher-image (left side) and plain-image (right side), using external key = ”unusacendana2000"

After X_n value is obtained, it will be converted back into integer form by taking eight points started after the dot (decimal) of real numbers. For example, assume that the value of X_n is 0.8080204084200282. After converting this value into integer form will yield 80802040, and then will be modulo with 256. The result value will be stored in array U[i] where i = 0,1, ..., 255. This process will be repeated until U[255] filled. The output of Step (4) is an array U[i] which also called “key streams” and consist of 256 pieces of pseudo random number sequence. The pseudo code of Step (4) is:

```
// X0 is an initial value //
x = 0;
hasil = 4 * X0 * (1 - X0);
do {
    hasilString = Str(hasil);
    iterasi = 0;
    for (i = 0; i < hasilString.Length; i++){
      if (hasilString[i] == '.'){
        iterasi = Int(Str(hasilString[i + 1]
                  +hasilString[i + 2]));
        exit for;
      }
    }
    for (i = 1; i <= iterasi; i++){
      hasil = 4 * hasil * (1 - hasil);
    }
    hasilString = Str(hasil);
    // ambil 6 angka setelah tanda titik  //
    randoms=hasilString[2]+hasilString[3]
            +hasilString[4]+hasilString[5]
           +hasilString[6]+hasilString[7]
            +hasilString[8]+hasilString[9];
    nilaiRandom = Int(randoms) mod 256;
    U[x] = nilaiRandom;
    x++;
} while (x < 256);
```

Step (5). Next we use U to produce the initial permutation of S. This involves starting with S[0] and going through to S[255], and, for each S[i], swapping S[i] with another byte in S according to a scheme dictated by U[i]:

```
j := 0
for (i:= 0; i <= 255;i++){
    j := ( j + S[i] + U[i] ) mod 256
  swap ( S[i], S[j] )
}
```

Because the only operation on S is a swap, the only effect is a permutation. S still contains all the numbers from 0 through 255.

Step (6). Once the S vector is initialized, the external key is no longer used. Stream generation involves cycling through all the elements of S[i], and, for each S[i], swapping S[i] with another byte in S according to a scheme dictated by the current configuration of S. After S[255] is reached, the process continues, starting over again at S[0]. The pseudo code is:

```
i := 0
j := 0
for (m:=54; m <= Length(ImageByte) -1 do
    i := ( i + 1 ) mod 256
    j:= ( j + S[i] ) mod 256
   swap ( S[i], S[j] )
   t:= ( S[i] + S[j] ) mod 256
  ImageByte[m] := S[t] XOR ImageByte[m]
}
```

The process of encryption can be seen in Figure 2.

Decryption Algorithm

The decryption algorithm is identical to the encryption algorithm discussed above except that the order of the basic operations is reversed.

IV. EXPERIMENTAL RESULTS

To empirically assess the performance of the our proposed method, we have carried out a number of experiments using C#. These experiments include encryption and decryption process, histogram analysis of plain-image and cipher-image, image content integrity test to find out whether there is an unperfect reconstruction of the image using message digest 5 (MD5) algorithm, and (iv) key sensitivity analysis.

Histogram Analysis and MD5

For this experiment, we have consider a 24-bit color image of size 270 x 272 and parameter $\lambda = 4$. This image is encrypted using external key = ”unusacendana2000”. The resulting encrypted image (cipher-image) is shown in Figure 2 (right side). The histogram of red, green and blue channel of plain-image and cipher-image is shown in Figure 5. Based on histogram analysis, we can see that the histogram of each channel (RGB) is uniform which make almost imposible for cryptanalyst to make statistical analysis to find the correct key. The results of histogram analysis also shows that there is a significant difference between cipher-image in Figure 5 dan plain-image in Figure 4.

We also test out proposed encryption method to find out whether there is a perfect reconstruction of image by using MD5 algorithm. Ideally, there is no difference between content size of plain-image and cipher-image. The results show that cipher-image (Figure 6.c) and plain-image (Figure 6.a) has the same hash value.

Key Sensitivity Analysis

We have carried out a key sensitivity test using a key that is one digit different from the original key to decrypt the encrypted image. We have encrypted plain-image using external key ”unusacendana2000” and then decrypted the cipher-image using: (i) wrong external key ”unusacendana2001”, and (ii) correct external key ”unusacendana2000”.

The resulting image is totally different from the original image as shown in Figure 7. This test demonstrates that the proposed algorithm is very sensitive to any change in the secret key value.

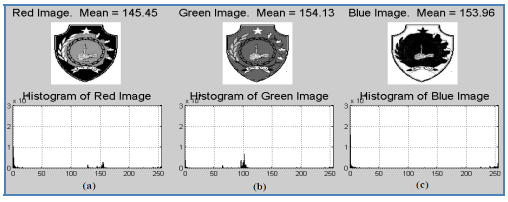


Fig 4. Histogram analysis of plain-image **(a)** histogram of red channel, **(b)** histogram of green channel, **(c)** histogram of blue channel

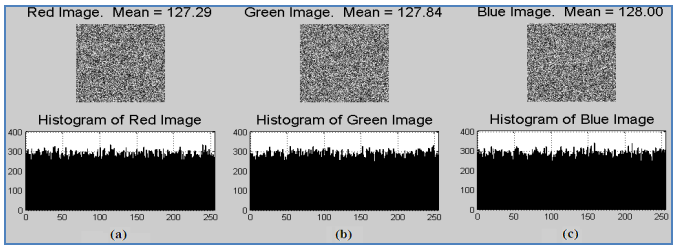


Fig 5. Histogram analysis of cipher-image using external key ”unusacendana2000” **(a)** histogram of red channel, **(b)** histogram of green channel, **(c)** histogram of blue channel

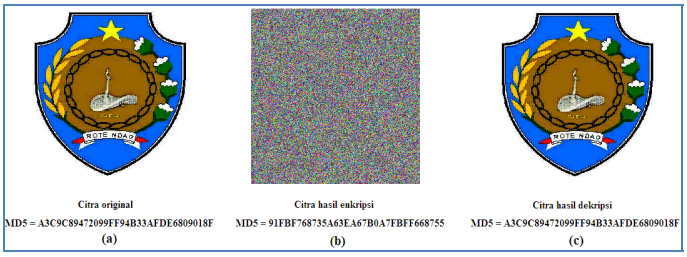


Fig 6. Image content size analysis using MD5 algorithm (a) MD5 value of plain-image (left side) = A3C9C89472099FF94B33AFDE6809018F, (b) MD5 value of cipher-image (middle) = 91FBF768735A63EA67B0A7FBFF668755, (c) MD5 value of decrypted image (right side) = A3C9C89472099FF94B33AFDE6809018F

Fig 7. Key sensitivity analysis (a) encrypted image (cipher-image) using external key ”unusacendana2000”, (b) decrypted image using correct external key ”unusacendana2000”, (c) decrypted image using wrong external key ”unusacendana2001”

255 255 255 255 255 255 255 255 255 255	(a)
10 10 97 121 113 52 188 210 122 186	(b)
233 92 206 1 167 244 8 151 129 40	(c)

Fig 8. Chosen plain-image attack **(a)** 10 bytes of plain-image starting from offset 80 until 103, **(b)** 10 bytes of cipher-image (offset 80-103) using key ”unusacendana2000”, **(c)** 10 bytes of cipher-image (offset 80-103) using key ”unusacendana2001”

Chosen Plain-Image Attack

We also have done some tests for simple chosen plain-image attacks and used two different keys. Figure 8 shows the complete results, where the output streams of images starting from offset 80 until offset 103 in Figure 8.a, Figure 8.b and Figure 8.c are totally different.

The difference of external key (even only 1 digit) will make significantly different of initial value X_0 and because

CLM is very sensitive to a small change in initial value so the outputs of our chaos system will be totally different.

V. CONCLUSION

We have proposed a color image encryption method which based on RC4 stream cipher and chaotic logistic map. Experimental results show that our method can be used as an alternative method to encrypt digital images because this method: (i) can encrypt image in such way so that cipher-image can not be visually identified by human, (ii) eliminates statistical relation between plain-image and cipher-image (histogram of cipher-image has a uniform distribution), (iii) is very sensitive to any changes in external key value, (iv) has no content size change between plain-image and cipher-image.

REFERENCES

[1] Pareek N.K, Patidar V, and Sud K.K, 2006. Image encryption usin chaotic logistic map, *Image and Vision Computing* 24 926–934.

[2] Mazloom, S., Eftekhari-Moghadam, A., 2011, Color Image Cryptosystem using Chaotic Maps, *IEEE*, 978-1-4244-9915-1.

[3] Liu Xiangdong, Zhang Junxing, Zhang Jinhai, He Xiqin, "Image scrambling algorithm based on chaos theory and sorting transformation", International journal of computer science and network security, Vol.8 No.1, January 2008.

[4] Schneier B. *Applied Cryptography*, Second Edition, New York: John Wiley & Sons, Inc. 1996.

[5] P.Xu; J. Zhao; D.Wang;, "A selective image encryption algorithm based on hyper-chaos," Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on, pp.376-379, 27-29, May 2011.

[6] V.Patidar, N.K.Parreek, G.Purohit, K.K.Sud, "Modified substitution-diffusion image cipher using chaotic standard and logistics maps", Communication in Nonlinear Science and Numerical Simulation 2010; 15(10); pp 2755-2765

[7] El-Alfy E.S and Al-Utaibi K. *An Encryption Scheme for Color Images Based on Chaotic Maps and Genetic Operators*.,2011. URL: http://www.thinkmind.org/index.php?view=article&articleid=icns_2011_5_10_10212, Visited on August 18, 2012.

[8] I.A.Ismail,Mohammed amin and Hossam Diab "An Efficient Image Encryption Scheme Based chaotic Logistic Map*", International Journal of Soft Computing*, 285-291,2007.

[9] Munir R. *Pengembangan Algoritma Enkripsi Selektif Citra dalam Ranah Spasial dengan Mode CBC-like Berbasiskan Chaos. 2012.* URL: http://informatika.stei.itb.ac.id/~rinaldi.munir/Penelitian/Makalah-SITIA-2012.pdf, Visited on August 18, 2012.

[10] Sathishkumar, G.A., Ramachandran, S., Bagan, B., 2012, Image Encryption Using Random Pixel Permutation by Chaotic Mapping, *IEEE Symposium on Computers & Informatics*.

[11] Wang XY, Yu Q., "A block encryption algorithm based on dynamic sequences of multiple chaotic system", Communications in Nonlinear Science and Numerical Simulation 2009; 14(2):574-81.

Distance Learning Lifecycle Management

An Agile Approach to Manage the Lifecycle of a Distance Learning

Ridi Ferdiana
Electrical Engineering and Information Technology Department
Universitas Gadjah Mada
Yogyakarta, Indonesia
ridi@acm.org

***Abstract*— the limitations of learning interaction on distance learning produces a variety of challenges. One of the major challenges in distance learning is to convince that distance learning produces the same quality as the conventional learning. The challenge can be solved if an organization has an implementation policy of distance learning. But in fact, a distance learning is only a limited e-learning software with digitized materials. This paper will propose a standard implementation policy called Distance Learning Lifecycle Management (DLLM). DLLM integrates Computer Supported Collaborative Learning (CSCL) concept and instructional design with the agile values. As a result, DLLM will guide any organization to manage and maintain the lifecycle of distance learning dynamically.**

Keywords— distance learning; e-learning; computer supported collaborative learning; instructional design; agile

I. INTRODUCTION

Distance learning or distance education has been part of current model of education. In contemporary form, distance learning means communications media to deliver instruction when teachers and learners are separated in place and/or time [1]. Distance learning has several benefits such as follows [2]:

- Student are no longer limited by geography. For example, students from rural area can get better education through distance education.
- Adult Students can get benefits to learn at their own pace at a time and location that is convenient to them.
- Educators praise the opportunities for interactivity offered by e-mail, discussion boards, and online resource centers.
- Educators get the possibilities for customizing and personalizing learning as well as for broadening the curricula

In Internet era, distance learning facilitates through many forms of digital material such as video, digital documents, audio, or even television. Therefore, a lot of researches that related with the use of e-learning or related technology to develop distance learning. For the example, the creation of live TV for distant education [2], the creation of "Platine" a synchronous collaborative environment [3], or the creation of distant VCR (Video Camera Recording) a virtual collaborative learning that integrates Moodle, DSpace, and Asterisk Voice [4]. The researches has similarity within each other's which is to enrich the experience of learner. The tools are not the only key of success factor of distant learning but also the foundation how to execute the distance learning.

However, the growing of Internet and computer software results a lot of choice in doing distance learning. Many tools provide many foundation to do the distance learning. Until now, there are few guidance to execute distance learning that independent within the tools. Several researches shows that the failure of distance learning is not only about the lack of good tools but also the lack of execution model of distance learning. The main distance learning issues are the fee of assessments, the readiness of technology, incentive and disincentive to participate in distance learning, institutional structure, and the class size [1].

Several researches shows that the growing of web application grows and the Internet communication tools make the distance learning has too many options with many tools [5]. Regardless of the tools, there are three models of distance learning execution which are [6]:

- Asynchronous distance learning. This distance learning uses e-learning that uploaded by the mentor and learnt by the student. The communication happens through online forum, posting comments, watching videos or sending email.
- Synchronous distance learning. This distance learning is done through direct interaction between mentor and participant such as video conferencing, instant messaging, or conference calling.
- Blended learning. The distance learning is embedded with traditional learning as an enrichment for the students.

The execution model shows the basic model how the distance learning is executed. The institution can choose a model that fit for them. Based on several researches [1] [2], the current distance learning execution provides several risks such as:

- Slow Internet connection provides challenges experience to follow distance learning.

- Students and teachers should have sufficient computer skill to follow distance learning activity.
- Lack of interaction between student and teacher. Distance learning has no means and little communication between each other's. It may not receive encouragement for their student learning needs
- Assessment issues. It is mentioned that the evaluation and assessment become an indicator of the quality of distance education. Therefore, the organization should develop quality examination
- Inflexible courses and technology. Most of the course management systems lack flexibility and require technologies that are not easily available to distance learner. Not only the technology but also some of the course might not be proper for distance learning such as several course that need direct interaction such as biologist, medical, and others.

The organization should make an effort to limit the risks. Therefore, many organizations use the e-learning tools that already familiar in distance education. However, the e-learning tools doesn't show how the organization execute the distance learning. Therefore after taking care the tools, organization should take care the process, the method, and the curriculum model that can be applied in distance learning.

In this paper, it makes an effort to summarize the successful of distance learning process and method. In order to make the process and method is legit, it combines a case study with the proposed lifecycle model for distance learning. As the novelty of this paper, it results a getting started process for any organization that want to execute the distance learning model in a procedural and agile way.

II. PREVIOUS RESEARCHES

Distance learning model evolves in fourth generation model [8]. It categorized with the use of Internet that enables participants to interact each other using virtual learning model. It is shown that the use of virtual learning model enhances the quality of interaction between student and teacher. Virtual learning model is about how to use Internet, devices, and instructional design model to provide better distance education learning experience.

The instructional design model has become popular to enhance the design of distance learning like e-learning or virtual classroom. Many important aspects exist between traditional learning and distance learning. For example, the transition from face-to-face to self-paced model using web application. Instructional design process on distance learning suggests three kind of interaction which are student to course content interaction, student to student interaction, and student to teacher interaction [1]. It means that the learning material as well as the distance learning tools should provide a value of interaction within a design.

The first previous research that related with this research is the development of the distance ecological model to support self/collaborative learning in the Internet environment [9]. In this research, it is proposed a distance ecological model that based on three representation which are learning goals, subject content in the designed subject model. And learning media. The research proposes a RAPSODY, a system that construct various kinds of learning form, design interactive and collaboration among participants. It is shown that a distance learning need a sufficient model to improve interaction between participants.

The usefulness of distance ecological model is supported by the others research like learning objects approach. The learning object approach is about a model that proposes the use of learning object repositories that can be easily reused or re-purposed by instructor, accessed by students, and shared among institutions [10]. The research captures many similar situation on the university that use distance learning on dual mode or blended. It proposes a learning object that works as intelligent learning content repositories that can be compatible with several standard such as IEEE learning object model (LOM), IMS content packaging (CP), and IMS digital repositories interoperability (DRI). Therefore, the model will give enhance sharing and interoperability between organization and system. In the research, it is shown that the learning object can be integrated and exchanged.

The learning object concepts is further discussed by the concept of Material-Framework-Objective (MFO) model [11]. In the research, it is shown that courseware based model caused a lot of problems such as dismissal, overlapped and desultory construction of learning material, and hardly adaptive to the diversity of learning objects. MFO provides learning system resource models based on four components which are Objective, Courseware, Knowledge Points, and Materials. These components are composed into a system in e-Dufe project. The system is composed by several sub-system which are materials management system, knowledge points system, courseware development system, and the intelligent guiding learning system. As a research contribution, MFO model erases the limitation of the mobility and pertinence of traditional learning model system, and improves the efficiency and the quality of study.

Based on the previous researches that already discussed. This research assumed that the better distance education will have several characteristics which are:

- Distance learning shall evolves with existing and next technology and it should adaptive to support traditional learning as well.
- The instruction design model still as a major solution to execute and manage the distance education.
- The learning materials is the first class of solution to improve the quality of distance learning. Therefore, several models are developed to improve the quality of learning materials to eliminate redundancy, to increase portability, and to enhance the interaction.
- The need of specific system that not only support the interaction of distance learning participant but also manages the learning object.

Based on that findings, Table I shows the position of the research among the others. It is shown in Table I that the research focused into a successful execution of distance

learning by integrating full lifecycle of distance learning with widely available of systems that support asynchronous and synchronous distance learning model.

TABLE I. RESEARCH POSITIONS

Year	Proposed Model	Contribution
2001	Distance Ecological Model [9]	RAPSODY system that constructs various form of learning, design interactive, and collaboration learning.
2004	Learning Object Model [10]	Model that provides an intelligent learning repositories that support various standard of learning object
2008	Material-Framework-Objective (MFO) Model [11]	Model that provides learning system resource management. It provides learning material management through it specific and open system
This research	Distance Learning Lifecycle Model	Model that provides end-to-end distance learning execution. It provides process and recommendation for a new organization who want to implement distance learning

III. RESEARCH METHOD

Learning is a process. In software engineering, process is a sequential activity that should be done to create a quality software. A quality software can be developed gradually using a concept that called lifecycle. Lifecycle is the entire activity on a process. It covers end-to-end activity of an engineering process. In this research, it is assumed that learning need a lifecycle to get quality learning.

The result of this research is a distance learning lifecycle model (DLLM). DLLM is a model for any organization who want to manage and execute distance learning. Therefore, the result will works as a fundamental model to execute the distance learning. In order to make this model flexible, it will follow an agile learning model (AGLEMO). AGLEMO is an instruction design model that is used to support ICT learning model for a people who want to learn ICT [12]. In AGLEMO, it is shown that the learning model should prepared using several phases which are

- Exploration phase. It identifies several backgrounds to execute the learning.
- Planning phase. It discusses contents that should be developed in the learning.
- Iteration phase. The iteration phase discusses how the planning phase is implemented as a learning activities.
- Production phase. In this phase, it is dedicated phase to deploy and test the learning activities to get feedbacks from the users.

In this research, it is done by following each phase of AGLEMO. As the result, each of phase of the AGLEMO will give deliverable for a DLLM initiative.

IV. RESEARCH RESULT

This research starts from the need of getting started guide for distance learning. In order to create a close relation between model and the real execution. This model is proposed based on lesson learned that taken from distance learning implementation in Indonesia. There are three case studies that are used to propose and evaluate the models. The three case studies as described below:

- Case study A. Case study A is a distance learning that is executed on a telecommunication organization. The purpose of the distance learning is to increase the use of mobile Internet connection for their educational customer. The case study is held by executing distance learning that involves several high school to execute virtual classroom.
- Case Study B. Case study B is a distance learning that is executed on an education organization. The purpose of the distance learning is to provide updated skill for a teachers through distance learning. The case study is held by executing distance learning that involves teachers from around Indonesia.
- Case Study C. Case study C is a distance learning that is executed on a community. The purposes of the distance learning is to update he product knowledge on the community. The case study is held by executing distance learning that involves community members from around the world.

After understanding the case studies, the research uses the AGLEMO phase on each case studies. In this research, it will discuss the activities that related within a phase for each case studies.

A. Exploration Phase

Exploration phase starts from the need of distance learning. It covers the main activities to prepare a distance learning. Based on the three case studies, it is found that the activities that are done on this phase is shown below:

- Composing key of stakeholders. In this activity, it is done by choosing the key of stakeholder. The stakeholder contains a person in charge for the distance learning, an organizational leader, and a senior consultant.
- The stakeholder will do a meeting to compose or to choose a technical team member. In Case study A and B, the stakeholder invite the capable vendor to join the distance learning initiative. In case study C, the stakeholder asked the volunteer for the technical team members. This technical team shall handle the technical issues such as technology solution, implementation solution, and execution support. This technical team contains engineer, IT professional, and education consultant.
- The technical team explore a technology review for the proposed distance learning tools. In this activity, it is proposed distance learning tools included with the alternatives. Table II shows the proposed learning tools for each case studies
- The technical team reviews the proposed learning tools to the stakeholder. It provides the advantage and disadvantages for each tools. In the end of the review, the technical team and the stakeholder make an

agreement to choose the technology that proper with their constraints

The exploration phase milestone is a commitment between stakeholder and the technical team to choose the learning tools for their distance learning. After choosing the tools, the team can go to the planning phase.

TABLE II. PROPOSED LEARNING TOOLS ON THE CASE STUDIES

Case Study	Proposed Learning Tool		Notes
	Asynchronous Activity	Synchronous Activity	
A	Office 365 Portal	Lync Online, Cisco Web-ex	The enterprise background make the organization prefers to choose enterprise solution. In the end of the exploration phase, they choose Office 365 to get less investment and better productivity by using Microsoft Office platform.
B	Moodle LMS, SharePoint	Skype, Lync Online	The massive users is the main consideration of this case study. In the end of exploration phase, the stakeholder choose Moodle and Lync Online.
C	Online forum, Facebook group	Google Hangout, Skype	The flexibility of the community make the stakeholder choose Facebook group and Google hangout

B. Planning phase

The planning phase focuses to deliver the plan to execute the distance learning. It covers several activities which are:

- The stakeholder creates curriculum plan with the proposed model and tools that proposed by the technical team.
- The technical team creates execution plan based on the proposed model and tools.
- The stakeholder discusses with the technical team to create milestone and execution plan for the distance learning execution.

The case study A, it starts with the initiative to do a simulation of distance learning. The stakeholder proposes to create distance learning simulation that mimics the real situation. The simulation is done by creating execution scenario with the tools and real material. The technical team should have backup plan if the simulation gives unexpected result.

The case study B starts with the pilot project that executed in the internal staff. It means that the pilot project is done by the stakeholder and staff itself. The purposes of this activity is to measure the execution plan that proposed by the technical team. The pilot project will show go or not go decision with proposed solution.

The case study C starts with the execution strategy. It differs with A and B, C has no simulation or pilot project during the limited budget and time. Therefore, case study C creates a solid plan without an evaluation. The solid plan is done by identifying the constraints that will be faced on the distance learning execution based on the experience intuition. The constraints examples are Internet connection, platform heterogeneity, and hardware limitation.

Based on the case study result. It shows that the proposed tools in exploration phase is evaluated through several approaches such as simulation, pilot projects, or experience intuition model.

C. Iteration Phase

The primary objective of the iteration phase is the execution of the distance learning. Iteration means repeated action to improve the result. Therefore, the activities will do in iterative manners such as follows:

- The technical team will monitor the distance learning execution. It creates a log book that records the execution of the distance learning. The log-book is created by the moderator when he joined the distance learning. The log-book is updated for each executed distance learning.
- The stakeholder team creates a feedback mechanism from the user to make the distance learning execution performs better. The feedback mechanism is done by creating an electronic feedback form to the users and it is used on any distance learning execution.
- The stakeholder and the technical team creates a review meeting to discuss the feedback and the technical issues from the users. The review meeting is done frequently to improve the distance learning experience.

In case study A, the technical team propose to do activity using plan-do-check-act discipline. It means that any issues that arrive in the previous of execution will be fixed on the next iteration of distance learning execution. For example, in the first meeting the users from Padang can't access the services using his mobile devices. In the next meeting, the stakeholder and technical team make a policy that related with the learning experience using mobile devices.

In case study B, the iteration phase done smoothly by creating a distance learning rules before the distance learning execution. In the execution, the stakeholder and technical team creates bi-weekly meeting to review the log-book and feedback users. The result of the bi-weekly meeting is a refined the distance learning rules, for the next iteration or class.

In case study C, the iteration phase is done by creating a forum topic that received the feedback from the users. The feedback form come up with two main forms which are pooling form, and comments form. The participant can submit the idea for the better experience, the reaction of the current distance learning, and do vote for the quality of the current distance learning. The feedback form itself is reviewed by the technical team to create better distance learning experience.

D. Production Phase

In software engineering, production phase means that the software is ready to deliver to the users. In the distance learning model, production phase means the organization have sufficient experience to create fruitful distance learning experience. The production phase is reached after the organization did several iterations. The result of the production phase varies as follows:

- Knowledge Base (KB): the KB provides the users to understand the basic of distance learning on the organization. It provides a reference document for any difficulties that might be happen in distance learning. It is updated frequently based on accidents, issues, or findings.
- User Manual: the user manual provides detail implementation to full lifecycle of distance learning. The user manual is distributed along with the learning material of the distance learning or portal that related with the organization distance learning. In several circumstances, user manual is divided into two type which are user manual for
- Getting started page: Getting started page is a web page that published on the web that works as one stop shopping to start and understand how the distance learning shall be executed.

As mentioned before, the production phase produces the artifacts that are needed to deploy the distance learning. The artifacts periodically refined on each iteration of the distance learning. In case study A, the initial artifacts are ready on the exploration phase. In case study B, the initial artifacts are ready on the planning phase. In case study C, the initial artifacts are ready on the iteration phase. The difference is due to the condition of the different needs of each organization. The artifacts are prepared earlier during the vital need of the organization. For example, the distance learning on case study A is fundamentally needed as a basic interaction between the participants and presenter. Therefore, the distance learning become essentially needed. It is different with the case study C that only used the distance learning for enrichment purposes.

E. Putting it All Together on DLLM

Based on the monitoring results with three different case studies. In this research, it is proposed DLLM model with the big picture as shown in Figure I. As shown in Figure I, the phase of DLLM follows the AGLEMO phase which are contains exploration phase, planning phase, iteration phase, and production phase. Each phase contributes deliverables that can be used for the next phase.

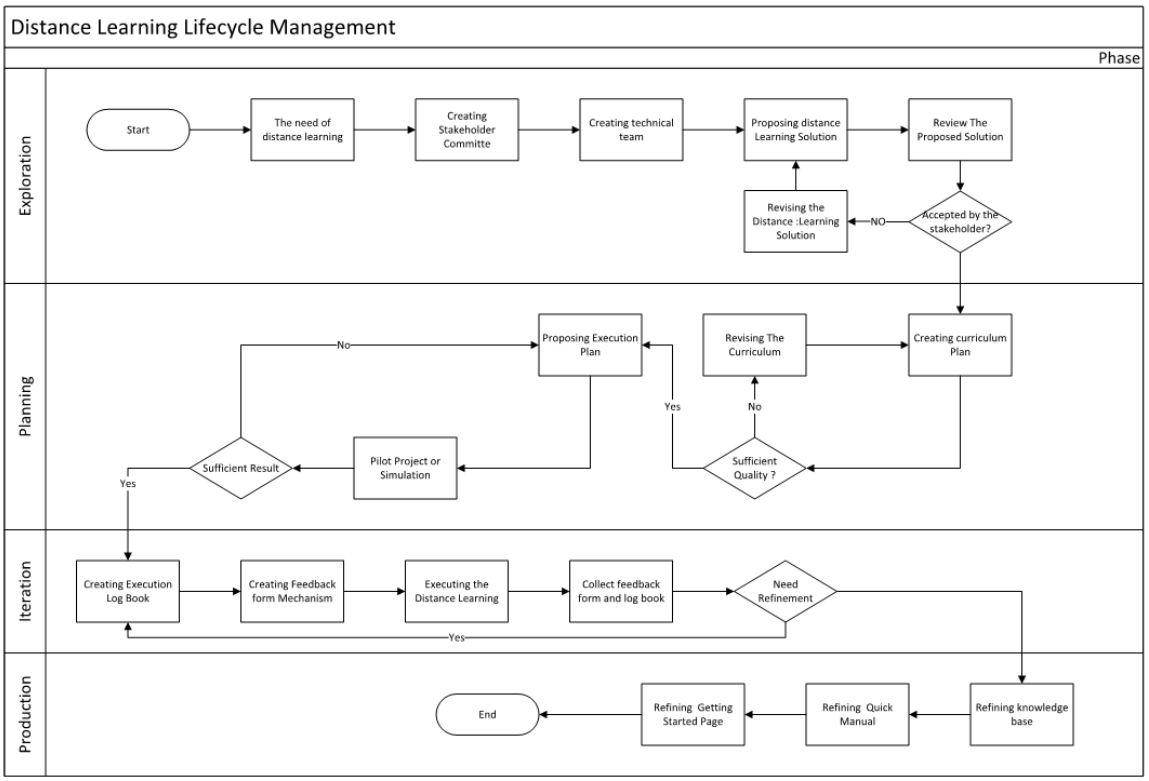


Fig. 1. DLLM big picture

The need of distance learning should become as a foundation how the distance learning is executed. For example, in University that held the distance learning as a blended learning for student, they will have a different purpose when the community use distance learning to enhance their product knowledge. However, the technical team and the

stakeholder are still the main actors of distance learning. Although, it has different foundation of the need.

The technical team can be a vendor or a dedicated division that have sufficient knowledge about distance learning. The technical team members should composed from the domain expert, technology expert, IT professional, developer, and tester. The composition can be categorized into two models which are implementer technical team and custom development technical team. Implementer means the technical team uses the existing technology without further development. Custom development team has a specific custom development for the distance education.

The technical team proposes a solution to the distance learning. It is validated and is discussed with the stakeholder about several aspects such as technology consideration, legacy content and system, and infrastructure investment. The proposed solution will be revised or agreed by the team based on the constraints.

The proposed solution should show the need of a digital content that can be used to deliver distance learning. Therefore, the next step after the proposed solution is to create the curriculum and content format (video, documents, audio, etc.). AGLEMO or instruction design model can be used here as a foundation to create distance learning curriculum. Curriculum and contents development is the hardest part on distance learning. Therefore, the team should review the content carefully and prepare the plan to evaluate the curriculum.

The evaluation of curriculum can be done through a pilot project or simulation. A pilot project is a way to execute the partial of participant to use the proposed curriculum, where the others using the legacy curriculum. Simulation is executing distance learning with mock users. The mock users are a fake user that can be technical team, stakeholder, or others persona that will join the distance learning as an evaluator. The result of the evaluation will be go or not go option. There is a condition when the evaluation is done by intuition during the limited budget and time.

If the organization tends to execute the distance learning. The organization should prepare the distance learning execution plan, the log book, and the feedback form mechanism. The plan-do-check-act can be used to create a refinement of distance learning execution. Any experiment will be logged to the log book. The valuable experience can be included in knowledge base, user manual, or getting started guide and it is refined iteratively.

Production phase is a stabilization of the distance learning model. The production phase should give the artifacts that can be used to the next distance learning refinement. The artifacts can be worked as a distance learning best practices on the organization. The artifacts such as knowledge base, user manual, and getting started guide can become an input for the long term evolution in distance learning.

Based on the result of this research, Table III shows the comparison between the research result and the previous researches. It is implicitly stated the properness of the implementation of this research result within the others. It also shows what the case that can be addressed on this research and what case that can't be addressed on this research but can be solved by the others research.

The DLLM is a management approach on distance learning. Therefore, it can't be compared directly with the previous researches since the previous researches focusing in distance learning solution. However, the DLLM provides management framework for any distance learning solution that proposed. It is already shown that the three case studies that already discussed have different solution but use the same DLLM.

TABLE III. RESEARCH RESULT COMPARISON

Researches	Implementation Model	Contribution
Distance Ecological Model [9]	Implementation Solution and Execution Constraint model	RAPSODY system that constructs various form of learning, design interactive, and collaboration learning.
Learning Object Model [10]	Concept Model and Implementation solution	Model that provides an intelligent learning repositories that support various standard of learning object
Material-Framework-Objective (MFO) Model [11]	Content model and implementation solution	Model that provides learning system resource management. It provides learning material management through it specific and open system
Distance Learning Lifecycle Model (This Research)	Getting started solution and agile execution management model	Agile learning model that provides end-to-end distance learning execution. It provides iterative and incremental distance learning execution

V. CONCLUSION AND FUTURE WORKS

In this research, it is proposed a lifecycle model for distance model called as Distance Learning Lifecycle Model (DLLM). The model adopts the blended learning model called Agile learning model (AGLEMO). As mentioned in the research result, the lifecycle model contains four phases which are exploration, planning, iteration, and production. In order to create sufficient model, the research uses three case studies as a conceptual model and lesson to learn. As the result, the research shows several findings which are:

- The main consideration of distance learning model is a purpose of the distance learning. The high distance learning needs is demonstrated through a set clear distance learning objectives.
- The distance learning team should have two sub teams which are technical team and the stakeholder team. Both team has a different role and members.
- Based on the case studies, the distance learning curriculum usually use the instruction design model through variety of digital learning materials and mediated instructions.

- The distance learning curriculum is evaluated through a simulation or pilot project. The evaluation results should indicate the quality before being distributed to the participants.
- The distance learning execution should have feedback mechanism. DLLM proposes the feedback mechanism through feedback form (questioner / pooling) and a log book. Both artifacts become main inputs to refine the distance learning.
- One of the signs that the distance learning model is well executed is the availability of a variety of artifacts that support the learning. Knowledge base document, getting started page, and user manual are the samples of variety artifacts on production phase.

Although this research shows the proposed model of distance learning through DLLM. The research still lack of evaluation after the model is proposed. Therefore, the further research should evaluate the proposed model to validate the properness and its contribution.

ACKNOWLEDGMENT

The authors wish to thank for the support of author family, Microsoft Innovation Center, ISV Partners, and Universitas Gadjah Mada electrical engineering and information technology department. This work was supported as part of a lecturer research 2013 program.

REFERENCES

[1] R. G. Fuller, G. W. Kuhne and B. Frey, Distinctive Distance Education Design, USA: IGI Global, 2010.

[2] M. Hentea, M. J. Shea and L. Pennington, "A perspective on fulfilling the expectations of distance education," in *Proceedings of the 4th conference on Information technology curriculum (CITC4 '03).*, New York, 2003.

[3] S. Li, S. Chen and M. Shyu, "A Live TV-Quality Distant Learning Multimedia Presentation System for Education," in *34th Annual Hawaii International Conference on System Sciences (HICSS-34). IEEE.*, Washington, DC, USA, 2001.

[4] D. Raymond, V. Baudin-Thomas, K. Kanenishi, T. Gayraud, K. Matsuura, M. Diaz and Y. Yano, "Distant e-learning using synchronous collaborative environment "Platine"," *Multimedia Software Engineering, 2004. Proceedings. IEEE Sixth International Symposium* , pp. 88-95, 2004.

[5] T.-A. Su, "Distant VCR - A Virtual Classroom for Distance Learning," in *Information Technologies and Applications in Education, 2007. ISITAE '07. First IEEE International Symposium* , USA, 2007.

[6] L. K. Lau, Distance Learning Technologies: Issues, Trends and Opportunities, USA: IGI Global, 2000.

[7] T. Le and Q. Le, Technologies for Enhancing Pedagogy, Engagement and Empowerment in Education, USA: IGI Global, 2011.

[8] L. A. Gerhardt, "The Future of Distance Learning - The Process and the Product," in *Information Technology Based Higher Education and Training, 2005. ITHET 2005. 6th International Conference*, Dominican Republic, 2005.

[9] T. Okamoto, M. Kayama, A. Cristea and K. Seki, "The Distance Ecological Model to support self/collaborative-learning in the Internet environment," in *Advanced Learning Technologies, 2001. Proceedings. IEEE International Conference*, 2001.

[10] P. Mohan and B. Daniel, "A new distance education model for the University of the West Indies: a learning objects' approach," in *Advanced Learning Technologies, 2004. Proceedings. IEEE International Conference*, 2004.

[11] Y. Yang, Q. Yang and K. Deng, "A Learning Material Management Model for Distance Education," in *Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference* , 2008.

[12] R. Ferdiana and O. Hoseanto, "Agile Learning Model for ICT Self-Paced E-Learning in Heterogenic Users Environment," in *ISODEL 2012 International Conference* , Bali, Indonesia, 2012.

Efficiency Factor and Risk Factor Based User Case Point Test Effort Estimation Model Compatible with Agile Software Development

Abu Wahid Md. Masud Parvez, Senior Project Manager (IT)
IT Services, RMIT International University.
niloy.cit1@gmail.com, masud.parvez@rmit.edu.vn

Abstract— **In software outsourcing industry, to win a contract, outsourcing companies need to consider the price that they will introduce to customers. Therefore, estimation has played an important role in any type of software project and as well as for test projects. Particularly, the estimation of testing has been more critical and magnitude as companies need to present or bid their prices on the project to win the outsourcing contract. Wrong estimation can negatively affect the company sales in both ways. On the one side, the profit margin could be reduced or hurt when the bidding is less than that real effort. On the other side, the company could lose the project when the bidding is higher than that of competitors. Therefore, it is extremely necessary to apply economic and systematic estimation model or process to enhance the accuracy in estimation. The integration of highly skilled workforce and accurate test effort estimation software will provide significant benefits to companies. Up to now, there are various existing models of test effort estimation, yet use case point estimation is one of the most popular models in the agile software development industry. The present paper aims to investigate the use case point estimation model. Following the analysis, we have developed two important elements namely, efficiency and risk factor based on a new layer in the model developed by us. The benefit of the new model is to increase the performance and effectiveness of the use case point estimation model.**

Keywords- Software estimation; use case point estimation; efficiency factor; risk factor; agile development; Outsourcing.

I. INTRODUCTION

Regarding the software development industry, the culture of 'outsourcing' has become more popular to take advantage of economic factors and skills. In terms of this culture, companies have been decentralizing its developing or testing unit through a contract or outsourcing process into different zones. Test estimation is really important for software project planning and software bidding to win a project in software industry. In order to be successful in software test project, a proper execution and estimation significantly important in the software development life cycle. [1,3] Such an example in agile scrum software development, the team needs to estimate the test. Normally team does it manually by an open consultation. And then the scrum master will need to measure test effort by a standard process and find out the estimated effort from team really feasible or not. If not then SM will request the team to re-estimate the test again. At present software outsourcing is a really popular approach in software development industry. So when stakeholder decides to build any software product then normally ask the industrial companies to bid for the project. In the selection phase the stakeholder give importance in the companies previous project's track record, company profile, development team's skill set and as well as on bided price. So between two companies, the one who have the better price will have more chance to win the project. And this is happening so often as software industry become so much competitive. Beside there is multi level outsourcing. Such an example, one American stakeholder outsource project to a renowned company at Europe and then that company outsource the project to a medium size company at Southeast Asia. So for the multi levels outsourcing the biding become more and more important. In many software companies, people utilize non-standardized but conventional estimation methods to make things work. At present software companies mainly follow the following test estimation models [2,5,6] to estimate the test:

- FIA (finger in the air) or best guess
- Use case point estimation method
- Functional point estimation method
- Ad-hoc method, WBS
- Delphi technique
- Experience Based - Analogies and experts
- Three-point estimation (successive calculation)
- Function points / Test point Analysis
- Percentage of development effort method
- Percentage distribution

II. USE CASE ESTIMATION METHOD

A use case captures a contract between the stakeholders of a system about its behavior. The use case describes the system's behavior under various conditions as it responds to a request from one of the stakeholders, called the primary actor. The primary actor initiates an interaction with the system to accomplish some goal. The system responds, protecting the interests of all the stakeholders. Different sequences of behavior, or scenarios, can unfold, depending on the particular requests. This use case estimation model collects together those all different scenarios /use cases and convert them in to numbers and drive that to ward total estimated effort hours. The First step for every type of estimation is to determine the size and the next step is to do effort estimation. In Use case methodology, there are 4 major steps, [3,4] which actually finally give the estimated effort:

- Determine and compute the unadjusted use case point (UUCP).
- Determine and compute the technical complexity factor (TCF).
- Determine and compute the Environmental complexity factor (ECF).
- Calculate final Use case point (UCP).

III. OUR MOTIVATION

There are significant amount of offshore test studio in Southeast Asia. The main duty of these offshore test centers is to run the test process for the product developed by the development team, which is located in a developed country. Beside many offshore companies does the both development and testing. In offshore test studios agile development is most popular and mostly follow use case test estimation model for test effort estimation. After reviewing use case point estimation model, we realized that there is a gap between estimated effort and real implement effort. So we decided to review the whole use case point estimation system and found the following main issues, which were preventing use case estimation model from being more efficient.

1. After receiving a project, we assign a team to run the test cycle. The size of the team depends on the size of the project and the deadline to complete that test cycle as well. In the team we had all professional testers but not having the same skill set or even not same level of experience. Traditional use case model does not estimate anything depending on the tester's prospective. So finally that drive team toward wrong estimation.

2. This is really hard for a team to determine the conversion factor (CF) correctly. Teams in agile development are totally dynamic. Though same type of project the member of the assigned team most of the time very different.

3. Team really struggle to determine an accurate conversion factor (CF) when receive a completely new type of project which they have never done before. Especially the new and young test studio meet this issue very often.

IV. OUR DEVELOPMENT

In our detailed investigation we found that present use case point estimation method is just providing the estimation depending on the properties of the project. This is not considering enough properties of the team. Whereas the team is going to do the project. Beside this is almost impossible to form a team with same skill level for all the members. After considering all the point we found, we have designed a new layer to add in the use case point estimation model to make this model more efficient.

There are six important points to implement in this proposed new layer on the use case point estimation model. They are:

- Test Cycle length and cycle weight
- Test Team member definition
- Compute efficiency factor
- Determine Risk factor for each tester
- Compute average efficiency factor
- Effort adjustment.

A. Test cycle length and cycle weight

When we are running the test cycle the length of the test cycle can be different from project to project depending on the project length, project priority, team size and project complexity. We categorized the test cycle depending on it's length from class 1 to class 5. And we have assigned the cycle weight for each class.

- Class 1: Test cycle for one or two working days length for a small product or in a situation where team has small mount of backlogs ready to be tested. Class weight for this length is 0.2.
- Class 2: Class 2 are those test cycles which have the length from 3 to 4 working days and we have assigned the cycle weight for this class is 0.4.
- Class 3: Class 3 are those test cycles which have the length from 5 to 6 working days and we have assigned the cycle weight for this class is 0.6.
- Class 4: Class 4 are those test cycles which have the length from 7 to 9 working days and we have assigned the cycle weight for this class is 0.8.
- Class 5: Class 5 are those test cycles which have the length from 10 to 12 working days and we have assigned the cycle weight for this class is 1.

If the test cycle runs more then 12 working days then we break that test cycle after 12 days and create another class 1 to 5-test cycle depending on the extra days. Such an example if one test cycle length is 15 days. Then we break that in to two cycles, one is 12 days and another one is 3 days. So there will be two classes, one is class 5 and another is class 2. But this break down is just to calculate the efficacy factor (EF). In outsourcing software development industry long test cycles are really rare. In agile software development, outsourcing industry the most test cycles are class 2 and class 3.

B. Test team member definition

The second step is to review the performance, skill level, expertise etc. of each tester of the team and define in which level they are. Most companies already have defined their Tester / quality control engineer's level. In our approach we divided the tester level in to 1 to 4 and assign the level weight 1 to 4.

- Level one: Normally a fresher and newly joined tester in the team with less than one year of experience. We assign level weight (LW) for this is 1.
- Level two: Tester with an average skill and requires help from another member of the team in test cycle. We have assigned level weight (LW) 2 for this level.
- Level three: Tester with a fair skill, fair dependability and we have assigned level weight (LW) 3 for this level.
- Level four: Tester with expert skill level, dependable, able to work independently and can mentor other

members. We have assigned the level weight (LW) 4 for this level.

In our proposed way, these levels for tester are floating. The performance of the tester will readjust the level in regular interval. Normally software development companies review the performance of the team member quarterly or in every six months. We would like to propose to do it quarterly. After reviewing the performance of each tester in a quarter the Test Manager / Project manager will readjust the level for each tester.

C. Compute efficiency factor (EF)

The third step is to compute the efficiency factor. For a project such as we have 'n' number of testers to run the test. Each of them has their own level weight and year of experience. Assuming that the last test cycle they have run is Cycle Z. We have selected to consider the last tc number of test cycles they have performed. So we will consider the test cycle Z to Z-tc test cycle's bug number they found during those test cycles.

Table 1: Tester's performance in the last tc number of test cycles

Tester no.	Level weight (LW)	Years of experience (YE)	Bug in each test cycle			
			Cycle Z	Cycle Z-1	…	Cycle Z-tc
Tester1	LW1	YE1				
…..						
Test n	LWn	YEn				

Now we will find out N number of testers efficiency factor one after one to computer the final average efficiency factor (AEF). So the formula to computer the efficiency factor for tester N,

EFn = LWn * YEn * [{ (Bug number of Cycle Z /Cycle weight) /100 +……..+ (Bug number of Z-tc/Cycle weight) /100 }/tc]

D. Define the risk factor for each tester (RF)

Now we already have the efficiency factor for our each tester. Next step is to review each tester's efficiency factor and determine how much risk the test manager / project manager would like to take on that tester. Test manager / Project manager pick one tester such as tester1 and it's efficiency factor EF1. Depending on the EF1 value, previously run test cycles build quality and over all tester's profile, Test Manager / project Manager will determine risk for tester1 by following model principles:

- If the EF is too low then test manager / Project manager will remove that tester from the team and replace that position with another tester who has a higher EF. Such an example, if Tester1 has really low score then tester1 will be replaced by 'tester n+1'. So project still has the same 'n' number of testers. Test Manager / Project manager then will investigate why efficacy factor for tester1 is low and engage tester1 in to proper training, make sure that tester1 complete them properly and assign that tester in low priority upcoming project by setting high risk. Such an example if EF1 = 0.2 then tester1 will be removed and replaced by another tester with higher EF.
- Fresher or too young team member normally have the very low efficiency factor. For them project manager / test manager need to add a high risk factor to keep the future efficiency factor (FEF) in the range of 0.55 to 0.75.
- If efficiency factor (EF) for a tester is in a moderate range then Project Manager / Test manager will set a risk factor for that tester by considering that tester's recent performance. Few factors like, tester good performance in last few test cycles, motivation for work, recent training etc. KPI will encourage Project manager / test manager to set up higher risk factor for that tester. Such an example if tester2 has EF = 0.5 and show high performance in last couple of test cycles then test manager / Project manager can take high risk on tester 2.
- If any tester's efficiency factor (EF) is in good range then project manager / test manager can take a minor risk on that. Such an example if tester3 has EF3 = 0.7 and then take 10 %- 20 % risk on that tester.
- If any tester EF more then 1 then we just round that up and keep it 0.95. That mean that tester is great and a companies asset as extra performance from that tester going to benefit the company.

Depending on our experience we would like to share the best practices to take the risk for the testers depending on their efficiency factor (EF) in the following way.

Table 2: Guideline for adding risk based on efficiency factor range

EF score range	Comment	Action	Risk %
0.1 – 0.39	Very low	Remove from the team and engage in training (if not fresher)	No risk
0.4 – 0.59	Low	Add risk by considering KPIs	20 % - 45 %
0.6– 0.89	Fair	Add risk	10 % - 20%
0.90–0.99	Good	Add no risk	0%
1+	Excellent	Round it to 0.95	0%

So the project manager / test manager will determine 'n' number for tester's risk factor from RF1 to RFn and accordingly their risk values RV1 to RVn.

E. Compute average efficicncy factor (AEF)

The next step to compute the future efficiency factor (FEF) for each tester of the team. The formula for computing the future efficiency factor for tester1

$$FEF1= (EF1 + RV1)/10$$

The formula to compute the future efficiency factor for tester n,

$$FEFn= (EFn + RFn)/10$$

So the formula to compute the average efficiency factor (AEF):

$$AEF= (FEF1+ FEF2+\ldots\ldots\ldots\ldots+FEFn) / n$$

F. Test effort adjustment and Final effort

From use case point estimation model we got the total effort (TF). So we will readjust that total effort with our average efficiency factor (AEF). The formula to compute that final effort by using average efficiency factor (AEF) is:

Final effort (FE) = total effort (FE) / average efficiency factor (AEF)

Final effort is the number of man-hours, which will be submitted to client while project bidding and project planning or company can develop it's own sells strategy around this as well.

V. PROJECT DEMONESTRATION.

Now we will show, how we had integrated our new layer in use case estimation model for a project and how it gave us the final estimated hours.

A. Project Background

We implement our model in several projects and one of them was a web application development project, which is a medium size project. The scrum team sized 17 members among them twelve were developer, three testers, one Scrum master and another one the business analyst / Product owner.

B. Compute Unadjusted use case point (UUCP)

The project we took, that had fifteen simple use cases, three average use cases and four complex use cases. And we computed that and got the total unadjusted use case point (UUCP) 34.

Table 3: Compute unadjusted use case weight (UUCW)

Use case categories	Weight	Number of use case	Result
Simple	1	15	15
Average	3	3	9
Complex	5	2	10
			Total UUCW: 34

C. Compute unadjusted action weight (UAW)

We analyzed our all actors and found out that, we had 2 simple actors, 1 average actor and 2 complex actors. The total unadjusted actor weight (UAW) we got 10.

Table 4: compute unadjusted actor weight (UAW)

Actor type	Weight	Number of actor	Result
Simple	1	2	2
Average	2	1	2
Complex	3	2	6
			Total Unadjusted Actor weight (UAW): 10

D. Compute unadjusted use case point (UUCP)

The formula to compute unadjusted use case point (UUCP) is:

$$UUCP = UUCW + UAW$$

Here, UUCW = 34 and UAW = 10

$$UUCP = 34 + 10$$

$$UUCP = 44$$

E. Determine total technical factor

We determined the entire technical factors and assigned the perceived complexities and got the total technical factor 18.

Table 5: Compute total technical factor (TF= Technical factor, PC= Perceived complexity)

TF	Description	Weight	PC	CF
T1	Test tools	2	2	4
T2	Document input	2	2	4
T3	Test-ware reuse	1	1	2
T4	Distributed system	2	2	4
T5	Performance objectives	1	1	2
T6	Security Features	1	1	1
T7	Complex interfacing	1	1	1
				Total Technical factor: 18

F. Compute technical complexity factor (TCF)

The formula to compute the technical complexity factor (TCF) is:

$$TCF = C1+ (C2 * \text{Technical Total Factor})$$

Here, C1 = 0.6 and C2 = 0.014,Technical total factor = 18

$$TCF = 0.6 + (0.014 * 18)$$

$$TCF = 0.852$$

G. Determine and compute Envirnmental total factor

For this project our team determined the perceived complexity for each environmental factor and we got the total environmental total factor 17.5

Table 6: Compute environmental total factor (CF=Calculated factor, PC= Perceived complexity)

Env. factor	Description	Weight	PC	CF
E1	Application knowledge	1	5	5
E2	Test Environment	2	2	4
E3	Test Data	1	2	2
E4	Test Lead capacity	0.5	1	0.5
E5	Motivation	1	2	2
E6	Stable Requirements	2	3	6
E7	Part time workers	-1	2	-2
				Environmental total factor: 17.5

H. Compute environmental complexity factor (ECF)

First, The formula to compute the environmental complexity factor (ECF) is
ECF = C1 + (C2 * Environmental total factor)

Here, C1 = 1.4 and C2 = 0.0362
Environmental total factor = 17.5
ECF = 1.4 + (0.0362 * 17.5)
ECF = 2.0335

I. Compute final use case point (UCP)

The formula to compute final use case point is:

$$UCP = UUCP * TCF * ECF$$

Here, UUCP = 44, TCF =0.852, ECF = 2.0335
UCP = 44 * 0.853 * 2.0335
UCP = 76.32

J. Primary total effort

We determined the conversion factor (CF) for our web application is 10. So the primary total effort (PTE) estimation

$$PTE = UCP * CP$$

Here, UCP = 76.32 and CP = 10,
PTE = 76.32 * 10
PTE = 763.2

K. Compute efficiency factor (EF)

We formed test team of three testers who run the test cycles after each sprint of the project. The test team member's profiles, we are describing in the table bellow:

Table 7: Determine testers performance in last 3 test cycles (LW = Level weight, YE = Year of experience, BN= number of bug tester found in one cycle, CW= Cycle weight)

Tester	LW	YE	Bug in each test cycle					
			Last test cycle		Last test cycle -1		Last test cycle -2	
			BN	CW	BN	CW	NB	CW
Tester1	3	3	30	0.5	50	0.6	30	0.4
Tester 2	3	2	40	0.6	55	0.6	45	0.5
Tester 3	2	3	30	0.5	45	0.5	25	0.3

We know the formula to compute the efficacy factor. We considered the last three test cycles each tester run. So for tester1 the efficiency factor (EF1),
EF1 = 3 * 3 * [{(30 / 0.5)/100 +(50 / 0.6)/100 +(30 / 0.4)/100}/3]
EF1 = 9 * [{.60 + .83 + .75}/3]
EF1= 9 * 0.72
EF1 = 6.54

For tester 2 the efficiency factor (EF2),
EF2 = 2* 3 * [{(52 / 0.6) / 100 + (55 /0.6) / 100 + (45/0.5) / 100}/3]
EF2 = 6 [{0.86 + 0.91 + .90}/3]
EF2 = 6 * 0.89
EF2 = 5.34

For tester3 the efficacy factor,
EF3 = 2 * 3 *[{(30 / 0.5)/100+(45 / 0.5)/100+(25/0.3)/100}/3]
EF3 = 6 * [{0.6 + .90 + 0.83}/3]
EF3 = 6 * 0.77
EF3 = 4.62

L. Determine and assign risk factor (RF)

Now we have the three tester's efficacy factors. Lets set up their risk factor for each of them.

- For Tester 1: in level 3, had 3 years of experience and efficiency factor 6.54. For having this high level in testing, experience and efficacy factor we took the risk factor 25%.
- For tester 2: In level, year of experience is 2 years and efficacy factor 5.34. A tester having a tester level 3 in two year experience that means a good profile. So we took the risk 35%.
- For tester 3, though this tester did not have high efficacy factor but had three year of experience. And he had a bad score due to his last test cycle's low performance. He was a bit sick that time and after that sprint he was conducted a training by his mentor. So we took the highest rick on tester 3, which was 45%.

Table 8: Tester's efficiency factor (EF), risk percentages and risk value (RV)

Tester	Efficiency factor (EF)	Risk factor (RF)	Risk value (RV)
Tester1	6.54	25%	1.63
Tester2	5.34	35%	1.86
Tester 3	4.62	45%	2.07

M. Compute average efficiency factor (AEF)

The formula to compute the future efficacy factor (FEF) is

$$FEF= (EF + RV)/10$$

Future efficiency factor (FEF1) for tester 1

$$FEF1= (EF1 + RV1)/10$$

$$FEE1 = (6.54+1.63)/10$$

$$FEF1 = 0.81$$

Future efficiency factor (FEF2) for tester 2

$$FEF2= (EF2 + RV2)/10$$

$$FEF2= (5.34+ 1.86) / 10$$

$$FEF2 = 0.72$$

Future efficiency factor (FEF3) for tester 3

$$FEF3 = (EF3 + RV3)/10$$

$$FEF3 = (4.62 + 2.07) / 10$$

$$FEF3 = 0.66$$

So the formula to compute the average efficiency factor (AEF):

$$AEF= (FEF1+ FEF2+….+ FEFn) / n$$

Here, FEF1 =0.81, FEF2=0.72, FEF3=0.66 and n=3

$$AEF = (0.81+0.72+0.66)/3$$
$$AEF = 0.73$$

N. Test effort adjustment and Final effort (FE)

The formula to calculate the final effort:

Final effort (FE) = primary total effort (PTE) / average efficiency factor (AEF)

Here primary total effort FE= 763.2, average efficiency factor (AEF) = 0.73

$$FE = 763.3/0.73$$

$$FE= 1045 Hours.$$

VI. Performance Analysis

For performance analysis of our new model we implemented four projects in expert estimation model, traditional use case model and in our updated efficacy factor based use case based use case estimation model. Expert estimation is the estimation by test expert. Then we completed that project and compared between real efforts and the estimation efforts we got from the three estimation models.

For project 1: it was a large size web application. By expert estimation we got the estimation 750 hours, by traditional use case model we got the estimation 870 hours and by our model we got the estimation 980 hours. After finishing the project, we got the actual effort 1090 hours. So on this project prospective we found, expert estimation gave worst estimation, as the deviation with the real effort was 32% less. Our model gave 11% less than the real effort.

For project 2: It was a medium size mobile application for a social network service. By expert estimation we got the estimated effort 484 hours, by traditional use case model we got the estimated effort 420 hours and by our model we got the estimated effort 530 hours. After finishing the project, we got the actual effort 520 hours. So on this project perspective we found expert estimation gave worst estimation, as the deviation with the real effort was -30%. Our model gave 2% more than the real effort.

For project 3: It was medium size web application. By expert estimation we got the estimated effort 910 hours, by traditional use case model we got the estimated effort 780 hours and by our model, we got the estimated effort 1015 hours. After finishing the project, we got the actual effort 985 hours. So on this project perspective we found the traditional use case estimation have worst estimation, as the deviation with the real effort was -21%. Our model gave 3% more than the real effort.

For project 4: It was a small size mobile game application. By expert estimation we got the estimated effort 320 hours, by traditional use case model we got the estimated effort 230 hours and by our model we got the estimated effort 270 hours. After finishing the project, we got the actual effort 254 hours. So on this project perspective we found expert estimation was worst, as the deviation with the real effort was 21%. Our model gave 6% more than the real effort.

Table 9: Performance comparison between expert estimation (Expert), Use case point estimation (UC), our developed model and actual effort.

Type	Expert (Hour)	UC (Hour)	Our model (Hour)	Actual effort (Hour)	Deviation with expert	Deviation with UC	Deviation with our model
Web	750	870	980	1090	-32%	-21%	-11%
Mobile	484	420	530	520	-30%	-20%	+2%
Web	910	780	1015	985	-8%	-21%	+3%
Game	320	230	270	254	+21%	-10%	+6%

From our above four-projects study we can come to a conclusion that our efficacy factor and risk factor base use case model is performing significantly better then other estimation methods. Especially for project 2, 3 and 4 the deviation was only around 2% - 6%. That is really helpful for outsourcing project as it can give a really helpful space for handling any other additional risks during test process.

VII. Conclusion

On this paper we investigated on software test effort estimation and designed a new layer inside the use case point effort estimation model. In experimental four projects our model gave most deviated result for project1 than the real estimation but still much better than other two estimation methods. For project two, three and four, we got the deviation with the real estimation was around +5%, which is outstanding in term of accuracy. In any type of software development including software outsourcing industry, estimation is always an important and tough task. So our motivation was to develop a better model which going to help more effectively and efficiently to the software companies in test effort estimation. In future it would be great if someone work with the conversion factor (CF).

References

[1] Sharma A. , Kushwaha D.S, "A metric suite for early estimation of software testing effort using requirement engineering document and its validation," Computer and Communication Technology (ICCCT) 2nd International Conference , 2012 , pp. 373–378

[2] Li-Xin Jiang, Wan-Jiang HAN, Chen-Chen YAN , Bo-Ying SHI, "Research on Size Estimation Model for Software system Test based on testing steps and Its Application," International Conference on Computer Science and Information Processing (CSIP), 2012, pp. 1245–1248.

[3] Nassif, A.B. , Capretz L.F. , Ho D., "Estimating Software Effort Using an ANN Model Based on Use Case Points," Machine Learning and Applications (ICMLA), 2013, pp. 42–47.

[4] Ramasubbu N. , Balan, R.K, "Overcoming the challenges in cost estimation for distributed software projects," Software Engineering (ICSE) conference, 2012, pp. 91–201.

[5] K.Krishna Mohan, A.K.Verma and A. Srividya, "Software Reliability Estimation Through Black Box and White Box Testing at Prototype Level," Reliability, Safety and Hazard (ICRESH), 2010, pp. 517–522.

[6] Bruce Benton, "Model-Based Time and Cost Estimation in a Software Testing Environment," Information Technology: New Generations, (ITNG), 2009, pp. 801-816

Extended Baker map using Scan patterns for image encryption

Pham Anh Hung
International College
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
anhhungsp89@gmail.com

Pitikhate Sooraksa, Kitdakorn Klomkarn
Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
pitikhate@hotmail.com, kkkitdak@kmitl.ac.th

***Abstract*—In this paper, an improvement on free-size image encryption scheme using Baker map combining Scan patterns is proposed. The generalized discrete chaotic Baker map uses sequences of parameters generated by chaotic Gauss map to divide image matrix into boxes. Scrambling method is done by using Scan patterns in each boxes of chaotic Baker map. After scrambling method, chaotic 3-D Chen system is selected to do the pixel transformation. Through the experiments, the security of proposed scheme is tested by introducing keys analysis, entropy analysis, statistical analysis, and processing time so that the scheme achieves high encryption efficiency**

Keywords— Image Encryption, Baker map, Gauss map, Chen system, Chaotic system, Scan patterns

I. INTRODUCTION

With the proliferation of the Internet, network infrastructure and computer technology are developed rapidly so that the use of communication applications has increased in recent years. More and more information are transmitted through internet. The security of information has been recognized to be vitally necessary. Information must be encrypted to reach the requirements – confidentiality, integrity, availability and controllability – before stored in database or transmitted.

In cryptography science, image encryption is more different and difficult than text encryption. Image has its own special characteristics – high redundancy, high bulk capacity, strong correlations among adjacent pixels, not sensitive as text [1]. Therefore, traditional encryption schema – DES, AES [2] – which have strong power in text encryption are not applied in image encryption efficiently. They have taken long time, high computing power and low security.

In order to overcome the weak-points, many encryption cryptosystems have been proposed with variety of different methods – SD-AIE [3], Mirror-like image encryption [4], SCAN language [5,19], Double Random Phase Encoding [6], Visual Cryptography [7] etc. Each method has its own properties in encryption. In these methods, SCAN is known as the secure method and has compression property.

The convention of chaotic theory has strongly opened the new and unexplored area in the field of cryptography. Because their special characteristics are the same with cryptography properties – ergodicity, sensitivity to initial parameter, mixing property, deterministic dynamics, structure complexity [8] –, the study of chaotic system and their capability to cryptography has received attraction from more and more people's attention. Mao and Chen [1], Kocarev [9] have introduced and discussed the using chaos in data encryption to demonstrate that chaos methods get good advantages in applications. Therefore, more and more chaotic-based schemes from researcher are proposed: 1-D chaotic systems [10,11], 2-D chaotic systems [12,13], high dimensional chaotic systems [14,15] and compounded chaotic systems [16,17]. Standing out in relief is Baker map [1,18] which gets high speed in computation encryption and efficiency.

Because the permutation way in each rectangle of Baker map looks like scan pattern in SCAN, this paper proposed an improvement encryption scheme using Baker map combining Scan patterns for free-size image. The generalized discrete chaotic 2D-Baker map uses sequences of parameters generated by chaotic Gauss map to divide image matrix into boxes. Scrambling method is done by using Scan patterns in each boxes of chaotic Baker map. After scrambling method, chaotic 3-D Chen system is selected to do the pixel transformation. Through the experiments, the scheme achieves high encryption efficiency – large cipher key space, resistant to a variety of attacks such as known plaintext, statistical analysis…etc.

The remainder of the paper is organized as follows: Section 2 introduces the chaotic Baker map. Section 3 introduces Scan patterns. Section 4 proposed encryption scheme is presented. Experiment and security analysis of the scheme are conducted in Section 5. Conclusions are shown in Section 6.

II. CHAOTIC 2-D BAKER MAP

The generalized discrete 2D-Baker map is applied to square NxN image [1] described as follows:

$$B_d(x,y) = \left(\frac{N}{p_i}(r - F_i) + s \bmod \frac{N}{p_i}, \frac{p_i}{N}\left(s - s \bmod \frac{N}{p_i}\right) + F_i\right) \quad (1)$$

In (1), the pixel (x,y) is in with $F_i \le r < F_i + n_i$, $0 \le s < N$. The sequence of k integers $\{p_1, p_2, \dots, p_k\}$ is chosen each integer p_i devides N, $F_i = p_1 + p_2 + \cdots + p_i$, $N = p_1 + p_2 +$

$\cdots + p_k$. The map stretches each rectangle in horizontal by a factor of $\frac{N}{p_i}$ while contract in vertical by a factor of p_i.

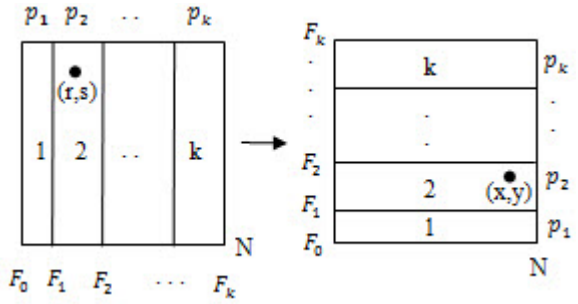


Fig. 1. Generalized discrete 2D-Baker map: all p_i devides N

In the case that not all of the integers $\{p_1, p_2, \dots, p_k\}$ not devide N, algorithm is more complex and discribed as follows:

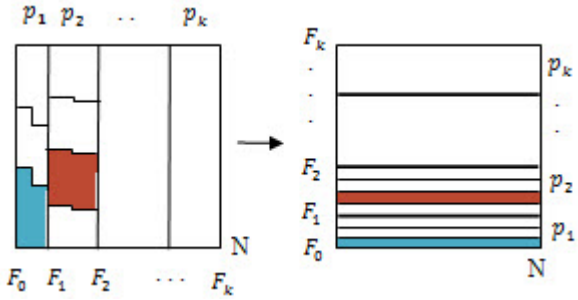


Fig. 2. Generalized discrete 2D-Baker map: not all of the integers $\{p_1, p_2, \dots, p_k\}$ not devide N

III. SCAN PATTERNS

SCAN is a class of formal languages based two dimensional spatial accessing methodology which can be applied to compression, encryption, data hiding or combinations. It can present and generate a large number of a wide variety of scanning paths or space filling curves. The SCAN family of formal languages includes several versions such as Simple SCAN, Extended SCAN, and Generalized SCAN, each of which can represent and generate a specific set of scanning paths.

Each pattern has eight transformations numbered from 0 to 7. For each basic scan pattern, the transformations 1, 3, 5, 7 are reverses of transformations 0, 2, 4, 6.

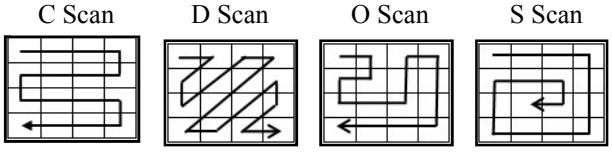


Fig. 3. Basic scan pattern

IV. PROPOSED SCHEME

A. *Square grayscale image (size NxN)*

1) Step 1:

a) Choose a chaotic map to generate 2 set of parameters for Baker map and Scan pattern. In this paper, chaotic Gauss map (also known as Gaussian map or mouse map) is applied and have function described below:

$$x_{n+1} = e^{(-\alpha x_n^2)} + \beta \quad (2)$$

With $\propto = [4.5,20]$ and $\beta = [-0.2, -0.9]$ this map gets chaotic. Comparing with almost 1-D and 2-D map, it has larger range parameter and more complex

b) Select k_1, k_2 and k_3 as three initial parameters /conditions to perform the Gauss map with 1000 iterations. Because Gauss map works in real space, the values from mapping have to be transformed to integer space. After that, select two subsequences $m = \{m_1, m_2, \dots, m_k\}$ and $n = \{n_1, n_2, \dots, n_k\}$ such that $m_1 \neq m_2 \neq \cdots \neq m_k$, $1 < m_1, m_2, \dots, m_k \leq N$.

2) Step 2: Perform permutation process (change position of image pixels)

a) Rearrange $m = \{m_1, m_2, \dots, m_k\}$ to get new subsequences $f = \{f_0, f_1, f_2, \dots, f_k\}$ in increase order where $f_1 = 1, f_k = N$ such that $N = p_1 + p_2 + \cdots + p_k$ where $f_i = p_i - p_{i-1}$.

b) Each value in $n = \{n_1, n_2, \dots, n_k\}$ is selected to match to the key path for Scan pattern which is applied in each rectangle of Baker map. Then perform 2D-Baker map combining Scan patterns:

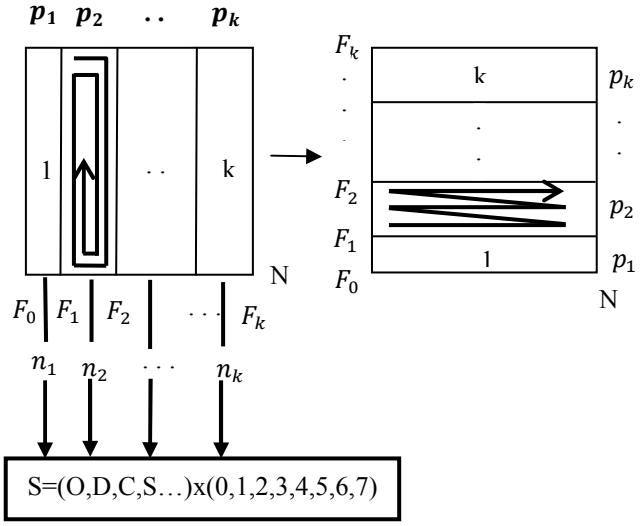


Fig. 4. 2-D Baker map combines Scan patterns. Example n_2 match to S0 and scan path S0 is done on the second rectangle

3) Step 3: Perform transformation process

a) Choose another chaotic map to generate random sequence for permutation process. In this process, we prefer hyper-chaotic 3-D Chen system because there are three reasons: almost low dimension systems generate not strong secure sequence; hyper-chaotic systems have more complicated phase space than low dimensional chaos; the equation of Chen system is more complicated and has bigger lyapunov than the other hyper-chaotic Loren, Rossler, Chua,…etc. Discrete 3-D Chen system is presented using Euler algorithm as follow:

$$\begin{cases} x_{n+1} = x_n + h(a(y_n - x_n)) \\ y_{n+1} = y_n + h((c-a)x_n - x_n z_n + c y_n) \\ z_{n+1} = z_n + h(x_n y_n - b z_n) \end{cases} \quad (3)$$

Where a,b,c are parameters and h is step in Euler algorithm. With $a = 35, b = 3, c \in [18,30]$ *or* $a = 35, b = [1,5], c = 28$ Chen system reaches good chaos property with maximum lyapunov $\gamma = 2.18962514$

b) Select k_4, k_5, k_6, k_7 and k_8 as four initial parameters/conditions to perform the discrete 3-D Chen system with 1000 iterations to get sequence $\{l_1, l_2, \dots, l_k\}$ with k=NxN. Because grayscale images are in integer, the sequence of mapping has to be transformed to integer by function described below:

$$l_k = \mathrm{mod}(\mathrm{round}(l_k 10^h), 256) \text{ h is integer} \quad (4)$$

Round() operation is used to get the value $l_k 10^h$ to the nearest integer. Mod(value,256) operation is used to limit the value not over 256 because maximum grayscale image are is 256.

c) After that, the pixels of the permuted image got from step 2 are arranged from 2-D matrix to 1-D matrix by the order from left to right and then top to bottom as follow:

$$B = \{(x_0, y_0), (x_1, y_1), \dots, (x_N, y_N)\} = \{B_1, B_2, \dots, B_{N\times N}\} \quad (5)$$

d) Then using below function to perform transformation:

$$\begin{aligned} C_1 &= h \oplus B_1 \oplus l_1 \\ C_2 &= C_1 \oplus B_2 \oplus l_2 \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ C_{N\times N} &= C_{N\times N-1} \oplus B_{N\times N} \oplus L_{N\times N} \end{aligned} \quad (6)$$

e) All values in function are transformed to binary to be used in XOR operation and then transformed again to decimal. The 1-D matrix $C = \{C_1, C_2, \dots, C_{N\times N}\}$ received is transformed back to 2-D matrix to get the encrypted image

$$C = \{C_1, C_2, \dots, C_{N\times N}\} = \{(x_0, y_0), (x_1, y_1), \dots, (x_N, y_N)\} \quad (7)$$

4) Step 4: Iterate step 2 and step 3 some rounds.

B. Rectangle grayscale image (size WxL)

In the case that the image is a rectangular image WxL, it can be reformed into a square NxN image where N is the integer makes NxN-WxL minimum. After the reformation, the new reformed image NxN can be encrypted by the scheme described upon.

V. EXPERIMENT AND SECURITY ANALYSIS:

The Lenna grayscale level images with square size NxN and rectangle size WxL are selected to do experiment with the proposed scheme. The set of keys $\{k_1, k_2, k_3, k_4, k_5, k_6, k_7\}$ which each of them contain 32 bit is randomly selected to use in the scheme. The scheme iterates 10 times in which each of time the set of keys is changed randomly. The result is shown in "Fig. 5" and "Fig. 6".

A. Key analysis

1) Key space:

The proposed scheme uses the set of keys $\{k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8\}$ which each of them contain 32 bit so that the key space size is $2^{256} \approx 10^{77}$. Therefore, the key space gets so large that resists the opponent try to guess the key by brute force attack infeasible. Compared with standard 2D-Baker map and the other scheme, this scheme has much larger key space.

2) Key sensitivity:

A test on Lenna grayscale level images in square size NxN and rectangle size WxL with the set of key {6.2,-0.58,0.0001,28,-1,3,14} are done. Then change a little bit in the set of key {6.2,-0.58,**0.0001**,28,-1,3,14} to {6.2,-0.58,**0.00011**,28,-1,3,14}, and use two sets of key to decrypt the encrypted image. Result shows that even though the set of key changes a little bit, decrypted image is not reversed back the original image.

B. Statistical analysis:

From the experiment in "Fig. 5" and "Fig. 6", the result shows that the histograms of the encrypted image are fairly uniform and significantly different from the histograms of the original image. It demonstrates that the scheme has good transformation properties and resists statistical attack on the encryption image.

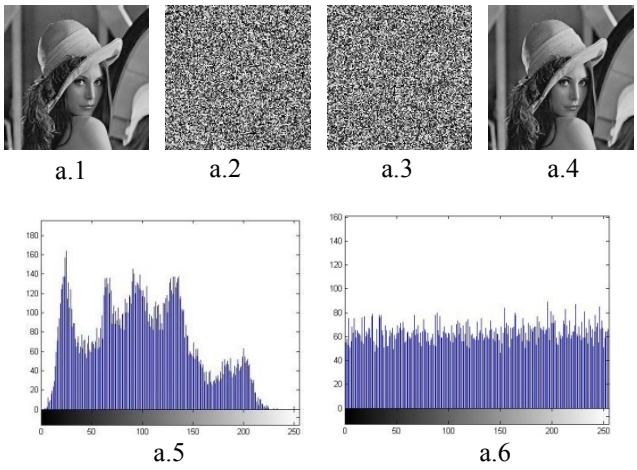
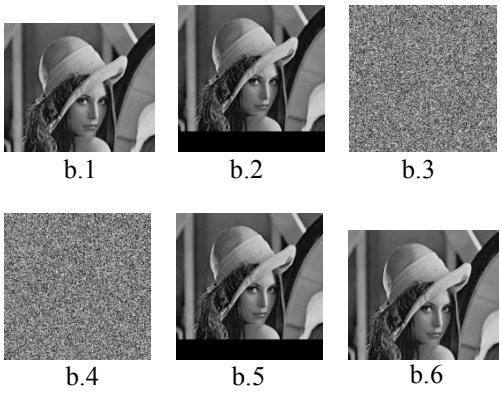


Fig. 5. a.1 128x128 lenna
a.2 Encrypted 128x128 Lenna
a.3 Decrypted 128x128 Lennan (wrong key)
a.4 Decrypted 128x128 Lennan (right key)
a.5 Histogram of 128x128 Lenna
a.6 Histogram of encrypted 128x128 Lenna



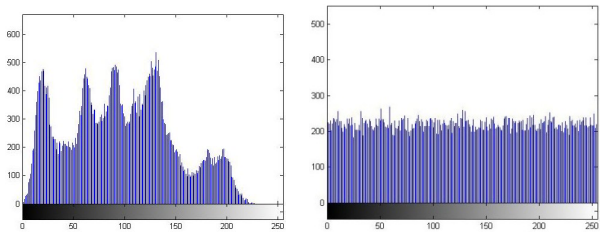


Fig. 6. b.1 220x256 lenna
b.2 220x256 lenna is reformed to size 256x256
b.3 Encrypted reformed 256x256 Lenna
b.4 Decrypted reformed 256x256 Lenna (wrong key)
b.5 Decrypted reformed 256x256 Lenna (right key)
b.6 220x256 Lenna is reformed back from decrypted reformed 256x256 Lennan
b.7 Histogram of reformed 256x256 Lenna
b.8 Histogram of reformed 256x256 Lenna

C. *Correlation of two adjacent pixels:*

This test is applied to illustrate the permutation property of scheme. There are three directions which have to be tested – vertical, horizontal and diagonal. Select 5000 pairs of adjacent pixels in each direction and then the procedure is done by using the following formulas:

$$E(x) = \frac{1}{N}\sum_{i=1}^{N} x_i$$
$$D(x) = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - E(x)\right)^2 \quad (8)$$
$$\mathrm{cov}(x,y) = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - E(x)\right)\left(y_i - E(y)\right)$$

Where x and y are grey levels of two adjacent pixels in the image, N is the number of pairs of adjacent pixels. The value of correlation coefficient shows that the pairs of adjacent pixels have strong correlation when value nearly reaches 1 and in contrast the value nearly reaches 0. The result shows in "TABLE 1".

TABLE 1. CORRELATION COFFICIENT

Size	Lenna gray image	Horizontal	Vertical	Diagonal
128x128	Original	0.8804	0.9408	0.8387
	Encrypted	0.0043	0.0049	0.0016
256x256	Original	0.9402	0.9669	0.9225
	Encrypted	0.0061	0.0088	0.0049
512x512	Original	0.9718	0.9851	0.9588
	Encrypted	0.0074	0.0063	0.0055
220x256	Original	0.939	0.9704	0.9146
	Encrypted	0.1902	0.0014	0.0051

D. *Entropy analysis:*

The entropy in grayscale image is the value that evaluates the probability distribution of each level in grayscale levels image. It is defined as follows:

$$H(x) = -\sum_{i=1}^{N} p(x_i)\log_2 p(x_i) \quad (9)$$

In (9), $p(x_i)$ states the probability of level x_i. When each of the level appears in equal probability $(x_i) = \frac{1}{N}$, the maximum of entropy is $H_{max} = log_2 N$. In situation of gray-scale image, the gray-scale levels are from 0 to 256. Hence the maximum entropy of grayscale image is $log_2 N = log_2 256 = 8$.

The entropy result in "Fig. 7" shows that encrypted image in this scheme gets the value bigger than original image and reaches nearly 8. So the information entropy improved markedly.

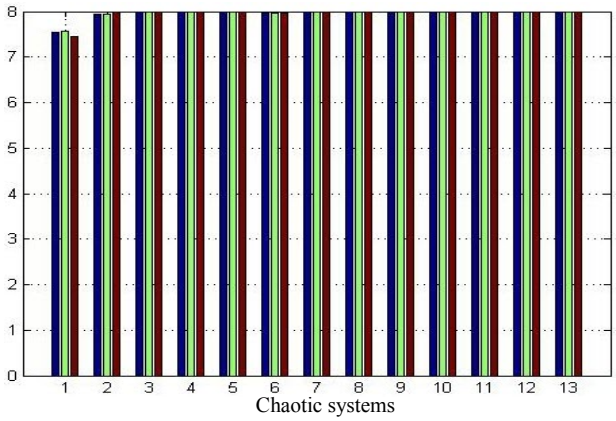


Fig. 7. Information entropy

7. Original image
8. Logistic map
9. Lorenz system
10. Henon map
11. Cross map
12. Circle map
13. Lozi map

1. Rossler system
2. Chen system
3. Kaplan map
4. Duffing map
5. Gauss map
6. Ikeda map

E. *Processing time:*

"Fig. 8" and "Fig. 9" shows the processing time of the proposed scheme compared with the other chaotic scheme. "TABLE 2" presents comparison with AES and DES. It presents that the proposed scheme has suitable processing time and better than the others.

TABLE 2. TIME PROCESS (SECONDS) AND KEY SPACE

Method	Lenna grayscale image			Key space
	128x128	256x256	512x512	
DES	≈126	≈511	≈2045	3.4×10^{38} (128 bit)
AES	≈10	≈36.5	≈142	1.1×10^{77} (256 bit)
Proposed method	≈0.3	≈1.1	≈4.2	1.1×10^{77} (256 bit)

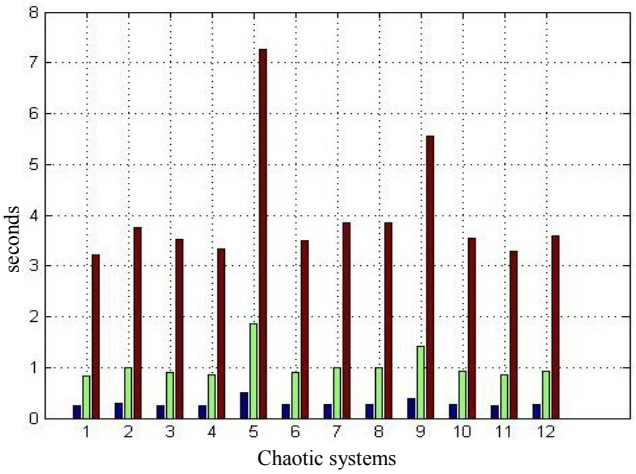


Fig. 8. Transformation processing time (seconds)

1. Logistic map
2. Lorenz system
3. Henon map
4. Cross map
5. Circle map
6. Lozi map
7. Rossler system
8. Chen system
9. Kaplan map
10. Duffing map
11. Gauss map
12. Ikeda map

Lenna 128x128
Lenna 256x256
Lenna 512x512

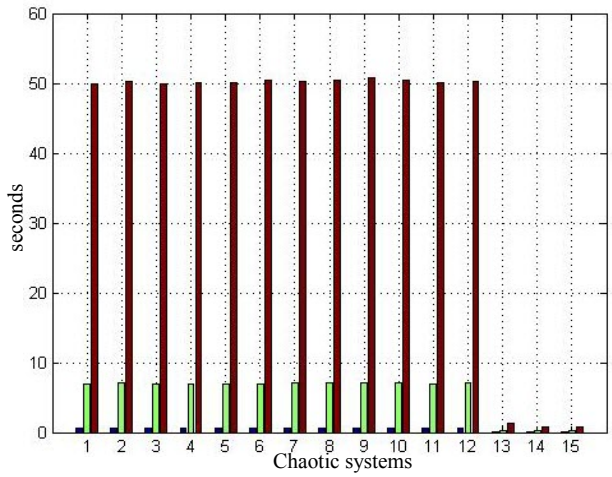


Fig. 9. Permutation processing time (seconds)

8. Logistic map
9. Lorenz system
10. Henon map
11. Cross map
12. Circle map
13. Lozi map
14. Rossler system
15. Chen system
1. Kaplan map
2. Duffing map
3. Gauss map
4. Ikeda map
5. Cat map
6. Baker map
7. Baker + Scan

Lenna 128x128
Lenna 256x256
Lenna 512x512

VI. CONCLUSION:

An improvement image encryption scheme is proposed using Baker map combining Scan patterns. The experiment is done on Lenna grayscale image in square size NxN and rectangle size WxL. The security analysis of the experiments demonstrates that the proposed scheme achieves high encryption efficiency – large cipher key space, resistant to a variety of attacks

REFERENCES

[1] Y.Mao and G.Chen, "Handbook of Computational geometry for pattern recognition, computer vision, neural computing and robotics", Springer 2003, pp.1-47.

[2] C. Paar and J. Pelzl, "Understanding Cryptography", Springer-Verlag Berlin Heidelberg 2010.

[3] S.Dey, "SD-AEI: An advanced encryption technique for images", IEEE 2012, pp.68-73.

[4] J.Guo and J.Yen, "A new mirror-like image encryption algorithm and its VLSI architecture", Department of Electronics Engineering National Lien-Ho College of Technology and Commerce.

[5] S.S.Manccam and N.G. Bourbakis, 'Lossless image compression and encryption using SCAN", Pattern Recognition34(2001), pp.1229-1245.

[6] S.Zhang and M.A.Karim, 'Color image encryption using double random phase encoding", MICROWAVE AND OPTICAL TECHNOLOGY LETTERS(1999), pp.318-322.

[7] Y.Hou, "Visual cryptography for color images", Pattern Recognition 36(2003).

[8] L. Kocarev and S. Lian (Eds.), "Chaos-Based Cryptography", SCI 354_c Springer-Verlag Berlin Heidelberg 2011.

[9] Ljupc˘o Kocarev, "ChaosBased Cryptography-A Brief Overview".

[10] M.Huang, Y.Huang and M.Wang, "Image encryption algorithm using chaotic map", IEEE 2010, pp.154-158.

[11] D.Chattopadhyay, M.K. Mandal and D.Nandi, " Symmetric key chaotic image encryption using circle map", Indian Journal of Science and Technology 2011, pp.593-599.

[12] C.W.Bin and Z.Xin, "Image encryption algorithm using Henon chaotic system", IEEE 2009.

[13] L.Wang, Q.Ye, Y.Xiao, Y.Zou and B.Zhang, "An image encryption scheme using Cross chaotic map", IEEE 2008, pp.22-26.

[14] A.Senthil Arumugam and D.Kiruba Jothi, "Image encryption using Improved 3-D choatic Cat map", IEEE 2010.

[15] J.Peng, S. Jin and X.Liao, "A novel digital image encryption algorithm using hyperchaos by controlling Lorenz system", IEEE 2009, pp.395-399.

[16] Y.Wang and J.Wang, "A new image encryption algorithm using compound chaotic sequence", IEEE 2012, pp.962-966.

[17] Y.Honglei and W.Guang-shou, "The compound chaotic sequence research in image encryption algorithm", IEEE 2009, pp.252-256.

[18] F.Han, X.Yu, S.Han, "Improved Baker map for image encryption", IEEE, pp.1273-1276.

[19] G. NAGARAJU and T. V. HYMA LAKSHMI, "Image Encryption using Secret-Key images and SCAN Patterns", NCIPA_VVIT 2012, pp.13-18.

Fermicidae swarm system

Thammarat Taengtang
Department of Electronics, Faculty of Electrical Engineer,
King Mongkut' of technology Ladkrabang,.
Ladkrabang, Bangkok 10520, THAILAND.
e-mail: ktthamma@kmitl.ac.th

Witthaya Sitthivet and Kitti Paithoonwattanakij
Department of Electronics, Faculty of Electrical Engineer,
King Mongkut' of technology, Ladkrabang.
Ladkrabang, Bangkok 10520, THAILAND.
e-mail: kpkitti@kmitl.ac.th

***Abstract*— improved an ant colony by firefly algorithm, in this paper is proposed the method that is interwoven between ant colony optimization and firefly algorithm to increase efficiency of solving the traveling salesmen problem it is called that Fermicidae swarm system (FSS). It uses relationship between pheromone and distance which is attractiveness and absorption coefficient. This method is based on ant colony optimization which state transition rule of ant colony is improved by adding detection, which is a condition of distances. The performance of FSS is divided into two parts: the speed and tour length of a result. A speedy result of FSS is faster than ACS and tour length of FSS is near the best result.**

Keywords—ant colony system, particle swarm, firefly algorthm

I. INTRODUCTION

The behavior models of an animal, class insects, have been presented such as ant colony system (ACS) [1] and firefly algorithm (FA) etc [2]. Nowadays we know that ants use only pheromones for a communication. ACS is the foraging model of ants, which can find the shortest path between the nest and the food source. This ability of an ant arises from an effective communication via the pheromone trails of it. There are a lot of pheromone codes to communicate ant colony which nobody in the globe cannot is completely interprets of the messages.

FA is based on particle swarm optimization (PSO) [3], is increased a capability to find partial solutions by the brightness. Actually, it is inappropriate to solve NP-hard problem: Traveling salesmen problem (TSP), Quadratic assignment (QAP) [4] etc. Natural models are modified to increase efficiency by many strategies that are two popular tactics: first, the parameters are improved of natural models by many ways [5], second, hybrid methods are created by natural models or the other mathematics [6]. Improved an ant colony by firefly algorithm is based on ant colony system (IAF) but global and local updating rule are unchanged. Specifically, this method is updated the transition state rule of ACS with attractiveness and absorption coefficient of FA method. The result of IAF were compared to the standard TSP which algorithm can be converge near the best solution

II. ANT COLONY SYSTEM

ACO, the biology model of class insects, is applied to solve NP-hard problem. It is very popular to solve the problems such as, TSP, QAP. ACS is appropriate to solve TSP. Only one ant per a city it can rest before start foraging. Ants will create partial solutions of relocation from *r* to *s* city with transition state rule, while travel to next city they will release pheromones on the ground. Relation between pheromone and distance are showed in an equation (1). It is rule used for decision to select the next city. Ants have to go to every city and then loop returned to r city (the first city). A length of tour is the partial solutions or distance that ants walked around of one cycle.

$$s=\begin{cases}\underset{u\in Jk(r)}{\operatorname{argmax}}[\tau(r,u)]^{\alpha}[\eta(r,u)]^{\beta} & if \quad q\le q_0 \ (\text{exp}loitation)\\ S & otherwise \ (\text{exp}loration)\end{cases} \tag{1}$$

where τ is the pheromone, η is the inverse distance, β is a correlative parameter of a pheromone versus distance ($\beta > 0$), $J_k(r)$ is the set of cities that unvisited by the *k-th* ant, S is the probability distribution, it is showed by an equation (2), q is a random number in [0,1], and q_0 is a parameter in [0,1].

$$p_k(r,s)=\frac{[\tau(r,s)]\cdot[\eta(r,s)]^{\beta}}{\sum_{u\in Jk(r)}[\tau(r,s)]\cdot[\eta(r,s)]^{\beta}} \tag{2}$$

While the ants are creating partial solutions, are release pheromones on the ground. The density of pheromone is relative with a length and a number of ants. Pheromones will be always updated all the calculations, which are divided as global updating and local updating rule.

A. Global updating rule of ACO

Global updating rule is used when all ants have visited every city. There is the best result of the partial solution chose to update the pheromone.

$$\tau(r,s)\leftarrow(1-\alpha)\cdot\tau(r,s)+\alpha\cdot\Delta\tau(r,s) \tag{3}$$

$$\Delta\tau(r,s)=\begin{cases}(L_{gb})^{-1} & if\,(r,s)\in global-best-tour\\ 0 & otherwise\end{cases}$$

where $0 < \alpha < 1$ is adjustable rate of a global pheromone, L_{gb} is the global best of solution.

B. Local updating rule of ACO

The *k-th* ant at *r* city lays pheromone on the ground while it is going to *s* city. Suddenly, the valued pheromone will be updated by an equation (4)

$$\tau(r,s) \leftarrow (1-\rho)\cdot\tau(r,s)+\rho\cdot\Delta\tau(r,s) \quad (4)$$

where $0 < \rho < 1$ is adjustable rate of a local pheromone, $\Delta\tau(r,s) = \tau_0$ is a factor of an adjustable pheromone [2]

III. Firefly Algorithm

The natural behavior of firefly, they use brightness for attracting other ones to them. Each firefly can attract other ones unalike depend on distance and light intensity. Firefly algorithm (FA) [8] can solve a nonlinear problem. Therefore, the motion of fireflies can be calculated by an equation (5).

$$x_i(t+1) = x_i(t)+\beta_0 e^{-\gamma r_{ij}^2}(x_i - x_j)+\alpha\left(rand-\frac{1}{2}\right) \quad (5)$$

where $x_i(t)$ is a location of firefly, $x_i(t+1)$ is a next location, r_{ij} is the distance between *i* and *j* firefly, β_0 is the attractiveness at $r_{ij} = 0$, γ is absorption coefficient, *rand* is a random number distributed in [0, 1] and α is a parameter in [0, 1].

IV. Fermicidae Swarm System

Improved ant colony optimization, ACO that is modified a transition state rule for increase ability by FA. The transition state rule of ACO is probability of pheromone density that relate an inverse distance. The pheromone densities consider as attractiveness of an ant (firefly) and a distance among the city consider as reduction of brightness. The transition state rule of ACO is changed to exponential equation of absorption coefficient. It gives by equation (6). FSS will modify only transition state but does not modify of local updating and global updating rule.

$$s = \begin{cases} \arg\max_{u\in J_i(r)}\left[\tau(r,u)\right]\left[\exp(-\gamma r_{ru})\right] & if \quad q \le q_0 \quad (\text{exploitation}) \\ S & otherwise \quad (\text{exploration}) \end{cases} \quad (6)$$

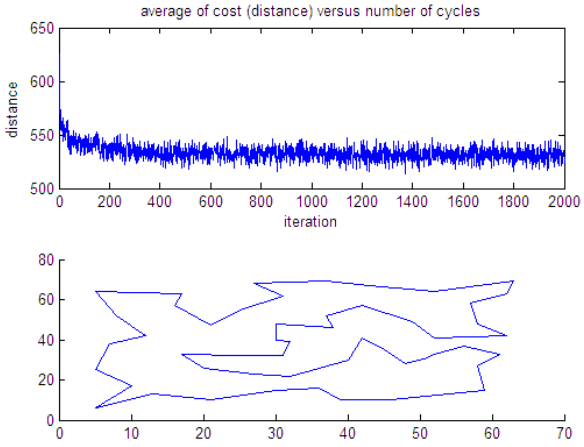
where $\tau(r,u)$ is the attractiveness (pheromone), γ is decimal constant of absorption coefficient, $J_k(r)$ is the set of cities that unvisited by the *k*-th ant, *q* is a random number in [0, 1], and q_0 is a parameter [0, 1], *S* is the probability distribution, it is showed by an equation (7).

$$p_k(r,s) = \frac{\left[\tau(r,s)\right]\cdot\exp(-\gamma r_{rs})}{\sum_{u\in Jk(r)}\left[\tau(r,s)\right]\cdot\exp(-\gamma r_{rs})} \quad (7)$$

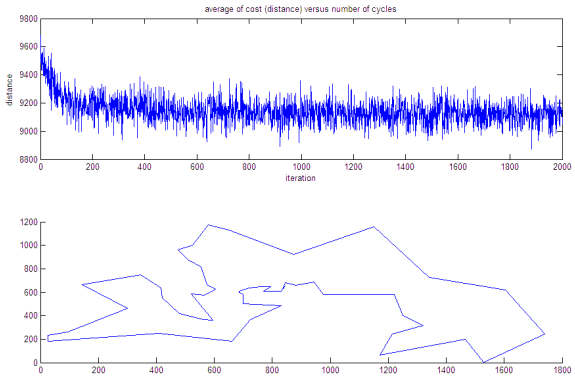
Now that FSS is based on ACO, then all partial solution will be updated by the global updating rule and the local updating rule of ACO.

V. Experimental and Result

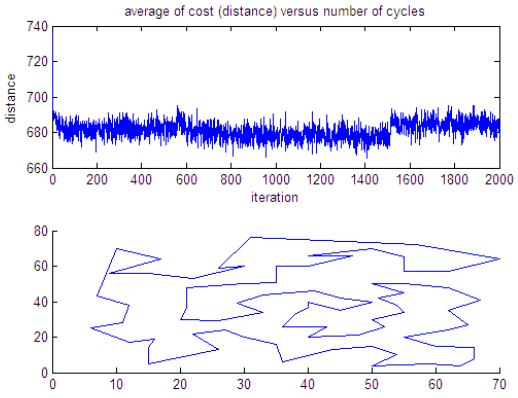
Testing the efficient algorithm, FSS is divided as two sections. The tour length and speech of finding result are compared between ACS and FSS. All of the initial pheromones on trail of the experiment are very low decimal constant and an absorption coefficient is 0.05. In this experiment are used three standard of TSP to test the ability and the speed of this algorithm. For fig. 1, it is showed average distance all of the ants and the best tour about ten times.



(a). eil51, tour length of result is 426



(b). berlin51, tour length of result is 7544



(c). eil76, the tour length of result is 555

Fig. 1. The best solution of this method

Each of the 2000 loops of an experiment are used. The tour length of results, this method is compared with standard of the best length showed in table 1. From 10 experiments per standard TSP, the results of FSS are near the best length.

TABLE I. Comparison of tour length results

TSP	Best lengths	Results
Eil51	426	426
Berlin52	7542	7544
Eil76	538	555

Speedy efficiency of algorithm is tested by Big O which the number of input is 8 to 500 cities. The speedy results are showed in fig. 2 which blue and red curve is speedy result of ACO and FSS. Each of time results calculated of iteration is asynchronous because computation with ACO is divided as exploitation and exploration to find partial solution. From fig. 2, blue curve is average of three times which is estimated of the cube function and red curve is estimated of the square function.

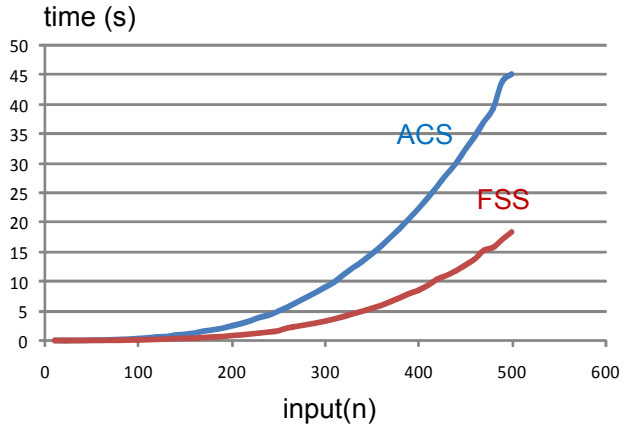


Fig. 2 speedy efficiency

VI. CONCLUDSION

This paper intends to lead the two methods, ant colony system and firefly algorithm are fusion to a new algorithm. Pheromones of an ant are appeal to other ant for finding the best solution. Distances between r and s city are absorption coefficient of exponential attenuation. From experimental results are considered of speed and efficiency by Big O and comparison of tour length result. The speedy processing of FSS is square function of which FSS is fast than ACS. Algorithm can be fined the tour length result closed the best length of standard TSP. Significantly, absorption coefficients have effect the length result and the number of input that uses experiment.

Acknowledgment

I would like to express my sincere thanks to computer services center of KMITL that supporting software (Mathlab).

REFERENCES

[1] M. Dorigo et L.M. Gambardella, "Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem," IEEE Transactions on Evolutionary Computation, vol 1, pp 53-66, 1997.

[2] Xin-She Yang, "Firefly algorithms for multimodal optimization," Stochastic Algorithms: Foundations and Applications, SAGA Lecture Notes in Computer Sciences, pp. 169–178, 2009..

[3] Yuhui Shi and Russell Eberhart, "A modified particle swarm optimizer," Proceedings of IEEE International Conference on Evolutionary Computation, pp. 69–73, 1998.

[4] E. N. Caceres, H. Fingler, H. Mongelli, "Ant colony system based solutions to the quadratic assignment problem on GPGPU," 41st International Conference on Parallel Processing Workshops, 2012.

[5] Wanjun Liu, Juan Zhang, Junli Liu, "Ant Colony Optimization Algorithm Based on Immunity Vaccine and Dynamic Pheromone Updating," IEEE International Conference on Data Mining Workshops, 2010.

[6] Junzhong Ji, Renbing Hu, Hongxun Zhang, Chunnian Liu, "A hybrind methdo for leanning bayesian networks based on ant colony optimization," Applied soft computing, vol. 11, pp. 3373-3384, June 2011.

Generating Customized Web Search Result Through Community Driven Search Engine

Bimo Sunarfri Hantono, Guntur Dharma Putra
Department of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
bhe@ugm.ac.id, guntur.dharma@mail.ugm.ac.id

***Abstract*— These days, the growth of web has led it to a big source of information. Web search engine plays an important role of searching desired information from this enormous web. However, search engine provides the same result independently to the user while actually each user has different preference. In this paper, we present a novel method of customized web search result generation to provide a better result according to community's preference. We benefit from proxy servers, which are widely used in a community network to reduce bandwidth needs. Proxy servers are, actually, providing the user preference within its access log that contains accessed URLs. Instead of web crawler, we will use this logs, which is always updated as users browse the web through this proxy. This would be the base of our customized web search. As the proxy log only covers URL list, we still need to crawl the information contained in an URL. When the crawling method has completed, document vector is created to make those data to be more machine friendly. Eventually, searching process is carried out by utilizing the vector space model.**

Keywords—search engine; information retrieval; customized web search; proxy log; tf-idf

I. INTRODUCTION

To seek preferred information in the web is a challenging task. This is simply because the exponential growth of the web has led it to a huge number of websites for more than 156 million websites [1]. The increasing Internet penetration has been boosting the Internet user day by day. Thus, in this part, search engines play a necessary role.

Typical search engine provides the same result for a query independently to the user [2]. This means, search engine provides less satisfactory result for the specific user need. For example, search engine would provide the same result for "lion" as the search query, no matter what the preference of the user is. In fact, the query "lion" itself would probably be animal, a computer operating system, or even an airline company in Indonesia.

Proxy servers are widely used to reduce Internet bandwidth usage [2]. By using this proxy, there is no need to fetch the same information while that information is still kept on the cache. Furthermore, a proxy server actually stores the list of websites and resources, which were accessed by the users. This log presents the community's preference if we could gather appropriate information from the access log.

We benefit from this condition. We will use the widely used proxy server to provide user preferences that are stored within its access log. Instead of web crawler that usually wanders across the web to gather the information, this proxy log would also act as information gatherer. The proxy log would be always renewed as the users access the Internet through this proxy so that the information stored in the log would remain updated.

The rest of this paper is organized as follows: related works and literature are reviewed in chapter 2. Proposed system model is presented in chapter 3, while the proxy server log as the data is explained in chapter 4. Then the methodology is defined in chapter 5. Finally, conclusion is drawn in chapter 6.

II. RELATED WORKS

There are several researches that dealt with this problem [3, 4,5]. Meta data for customization is also used to deal with this [8].

A comprehensive large-scale log files have carried out in China [3]. The Chinese search engine, Sogou, logs were used in this research to optimize and to personalize web search. The logs then used to analyze the search engine's performance. This research is just a comparison between personalized searches that are offered by Sogou.

Another method of creating customized search is to employ social relationship to produce a novel search engine [4]. As time goes by, the links between people in social media will form the so-called "Strong Links". For those users, information provided by the friends with strong link is considered as more interesting and useful. This link would later be used as the base of search engine rather than based on only measuring the similarity between keywords and articles. Although this method offers a robust solution, the links between people needs more complex processing than our method proposed in this paper.

While social relationship is used to generate search engine, another research is exploring the semantic web as their base of experiment [5]. That research presents WebOWL, an approach to make use of the latest technology to create a Semantic Web

Search Engine. It contains a group of intelligent agents that have several responsibilities: to do what a crawler does, to store data in semantic database, to support semantic queries with typical query mechanism, and to determine the order of search result with a ranking algorithm. Basically, this research entirely employs Web Ontology Language (OWL) to develop Semantic Web search engine. Then, OWL query is used to get the search engine result. Although this search engine prototype offers more semantic result, OWL query has more structure that is not human-friendly.

CYBER is a Community based search engine, for information retrieval utilizing community feedback information in a Distributed Hash Network (DHT) network [6]. This research goal is almost the same with our goal, to create customized web search for community, but we differs in approach and methodology. Furthermore, this research has to be implemented specifically in a DHT network, which is not familiar and widely used, while our research is using TCP/IP network that is more common in implementation.

Our previous work was carried out by using only the meta data of an URL [1], in this work; we will be using entire content of an URL in order to improve our search engine accuracy.

III. PROPOSED SYSTEM MODEL

The diagram of our system model is described as follows:

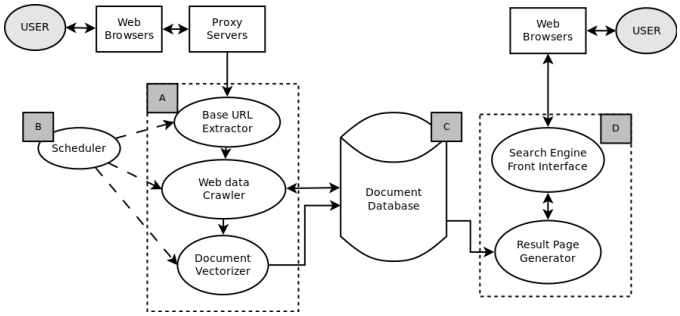


Fig. 1. The proposed system model overview.

A. The preprocessing Unit (A)

This unit does an important role in processing the raw proxy logs to be ready for the search engine. At first, this unit would extract the URLs from proxy logs. Then, the web data crawler will fetch the information stored in each URL and store it in document database for later process. Finally, to make this textual information more machine friendly, document vectorizer would transform it into numerical data (document vector). All the result from this unit, are stored in document database (C).

B. Scheduler (B)

Scheduler is responsible to trigger the preprocessing unit to work periodically for updating the document database. We make use of cron jobs in Linux to update the document database every 24-hour. Thus, this search engine would always remain updated.

C. Generating search result

Search result generation is all carried out in local search engine (D). The process starts when user requests a web search for a specific query. Then local search engine would calculate similarity between user's query and every document in the document database. To rank the URL in the search result page shown as the final result to the user, local search engine uses calculated vector similarity and URL's TF-IDF value as described in the following graph:

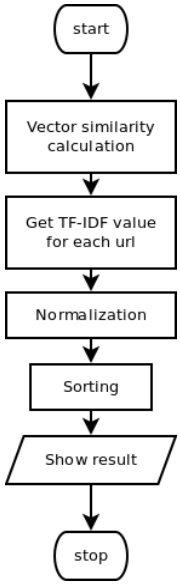


Fig. 2. Generating search result simple flowchart.

IV. THE PROXY SERVER LOG

This work was carried out by utilizing Squid proxy server's log files, which were collected from Department of Electrical Engineering and Information Technology's (EE & IT) local network, Indonesia. This data simply represent the community of Department of EE & IT, Indonesia, since it's storing the list of URLs that were accessed by users.

The log files used for this work were derived from Squid's access.log file. Those logs contain 10 fields, time, duration, client address, result code, size (in byte), request method, URL, requesting client name, hierarchy code, and type.

In this research, we are only using the URL field from Squid's access log file fields in order to extract users' preferences. The URL (Uniform Resource Locator) is a global address for specifying the location of a resource or a transaction. In this work we only parsed the URL up to the TLD (Top-level domain).

V. METHODOLOGY

To achieve this search engine, we use proxy server logs from Department of Electrical Engineering and Information Technology to fetch the unique URLs. Next step is to crawl the information (HTML) from all of the unique URLs. Then, to convert textual data into machine friendly numerical data we create document vector. Those results would be our base to

create the search engine and we use vector space model to generate search results from user's query.

A. Extracting URL

Before we move to further process, we need to extract the URLs from proxy server logs. In this extracting process, there are lots of URL that are actually pointing to a same website. That is just simply because lots of components contained within a web page are coming from multiple sources [1]. For example, inside a particular website there are lots of images displayed. Those images are actually coming from multiple sources, and Squid also list those images' location in the log files. If we crawl these URLs, it gives no textual result. Since crawling process takes time, it is such a wasting time to try to crawl something that would not give any results.

To overcome this circumstance, we developed a tricky step called whitelist-blacklist filtering mechanism:

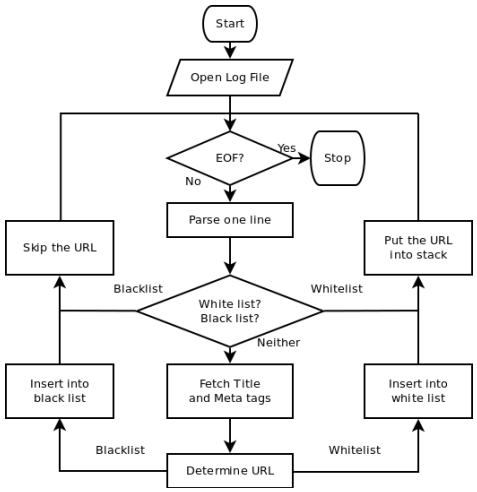


Fig. 3. Whitelist-blacklist mechanism flowchart.

Then we have blacklisted URLs that have no information in it and whitelisted URLs that have useful information in it for our customized web search.

B. The weighting scheme

The information gathered by our crawler is purely textual data. Later on, to process this information we need to transform it into a machine friendly data or document vector. The figure below describes our step to convert gathered information to be document vector.

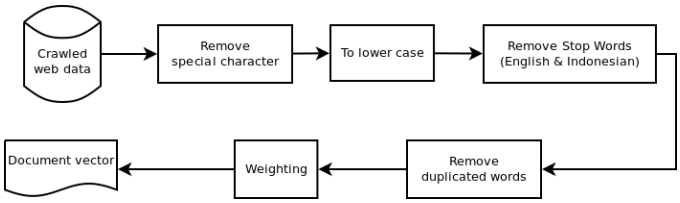


Fig. 4. Creating document vector flowchart.

There is two factors that are used in common information processing system [9]: TF, the frequency of the term in a text segment, and IDF, which is used to indicate the distinction of the term.

Here in this work, in order to transform textual data into numerical data, we use our previous formula used in our previous work [1].

$$weight(i,j) = \begin{cases} (1+\log(tf_{i,j}))\log\dfrac{N}{uf_{i,j}}, & if\ tf_{i,j} > 0 \\ 0, & if\ tf_{i,j} = 0 \end{cases} \quad (1)$$

$tf_{i,j}$ is the number of occurrences of word w_i in URL u_j, while $uf_{i,j}$ is the number of URLs in the collection in which w_i occurs in.

C. Vector Space Model

The Vector Space Model is commonly used structured form for text data in which individual text documents are represented as a set of vectors [9]. Later the matrix M would be converted into single vector V (word) so that those collections of words could be clustered using k-means algorithm. Converting process involves algebraic model of Vector Space Model. Then the vector is calculated as such:

$$V_i = \sqrt{\sum_{j=1}^{n} M_{i,j}} \quad (2)$$

Actually, this would work will be giving more satisfactory result if we implement text clustering [10].

D. Retrieving the Desired Information

To do retrieval in the vector space model, documents are ranked according to similarity with the query as measured by the cosine measure or normalized correlation coefficient. Thus, every document in this collection of documents would be measured the similarity with the search query requested by users.

Firstly, user's query would be treated as a single document and weighted using the same TF-IDF formula (1) as the URLs were computed. Getting the result, those 2 vectors (query Q and each URLs D) measured by using vector similarity:

$$sim(Q,D) = \frac{\sum_{i=1}^{n} W_{qi} * W_{di}}{\sqrt{\sum_{i=1}^{n} (W_{qi})^2 * \sum_{i=1}^{n} (W_{di})^2}} \quad (3)$$

W_{qi} is the weight of word W_i in the query Q, while W_{di} is the weight of word W_i in the URL D (meta information).

The most similar vector, in this case is URL, would have the highest value of formula 3. Finally, the search engine result would be derived from formula 3 written above.

VI. EXPERIMENTAL RESULT

We used a small sized proxy log that was collected in 24 hours to carry out the testing process. Then, the top seven of search engine result for query "internet" is as follow:

TABLE I. EXPERIMENTAL RESULTS

Rank	URL
1	http://www.safesquid.com
2	http://1jutarupiah.com
3	http://yuleko.com
4	http://vip.net.id
5	http://belajarbisnis.net
6	http://mitra-haji.com
7	http://www.browserdefender.com

Most URLs listed in the table above are Indonesian website that contain localized content.

VII. CONCLUSION AND FUTURE WORKS

In this paper, we presented a novel method to generate a customized web search with the help of proxy server log. So far, we have not made any comprehensive comparison and benchmarking of our custom search engine and the widely used search engine. However, we showed that it is possible to generate a web search through proxy server log.

In order to improve this search engine capabilities, there are some works that could be carried out particularly:

- Including a deeper user's query treatment, for instance, query recommendations, spelling and grammar checking.
- Benchmarking this search engine with the widely used search engine would be important to examine this search engine performance.
- Using different proxy server logs provided from different communities and same ambiguous terms to check whether this search engine could give more satisfactory result.
- Using other approaches to model the user preference from the communities.

ACKNOWLEDGMENT

The authors would like to express gratitude to Mr. Jafilun as the Network Administrator of Electrical Engineering and Information Technology Department UGM for giving the access to the proxy log files.

REFERENCES

[1] Putra, G. D., Saputra, F. A., & Watanabe, K. (2012). Community Driven Search Engine Based on Community's Proxy Server Log. *The 26th Annual Conference of the Japanese Society for Artificial Intelligence Proceeding*.2012.

[2] Taghavi, M., Patel, A., Schmidt, N., Wills, C., & Tew, Y. (2012). An analysis of web proxy logs with query distribution pattern approach for search engines. *Computer Standards & Interfaces*, *34*(1), 162–170. doi:10.1016/j.csi.2011.07.001

[3] Gao, J., Yang, M., Li, S., Zhao, T., & Qi, H. (2010). Improve Search by Optimization or Personalization, A Case Study in Sogou Log. *2010 International Conference on Asian Language Processing*, 231–234. doi:10.1109/IALP.2010.15

[4] Lu, H.-H., Ting, I.-H., & Wang, S.-L. (2012). A Novel Search Engine Based on Social Relationships in Online Social Networking Website. *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 1111–1116. doi:10.1109/ASONAM.2012.191

[5] Batzios, A., & Mitkas, P. a. (2012). WebOWL: A Semantic Web search engine development experiment. *Expert Systems with Applications*, *39*(5), 5052–5060. doi:10.1016/j.eswa.2011.11.034

[6] Li, Y., Shou, L., & Tan, K.-L. (2008). CYBER: A CommunitY-Based sEaRch engine. *2008 Eighth International Conference on Peer-to-Peer Computing*, 215–224. doi:10.1109/P2P.2008.24

[7] Shafiq, O., Alhajj, R., & Rokne, J. G. (2010). Community Aware Personalized Web Search. *2010 International Conference on Advances in Social Networks Analysis and Mining*, 351–355. doi:10.1109/ASONAM.2010.85

[8] Hassanpour, H., & Zahmatkesh, F. (2012). An adaptive meta-search engine considering the user's field of interest. *Journal of King Saud University - Computer and Information Sciences*, *24*(1), 71–81. doi:10.1016/j.jksuci.2011.10.004

[9] Xu, M., He, L., & Lin, X. (2010). A Refined TF-IDF Algorithm Based on Channel Distribution Information for Web News Feature Extraction. *2010 Second International Workshop on Education Technology and Computer Science*, *15001*(10), 15–19. doi:10.1109/ETCS.2010.130

[10] Tang, X., & Dang, J. (2012). An Exploratory Study of Enhancing Text Clustering with Auto-Generated Semantic Tags. *2012 Eighth International Conference on Semantics, Knowledge and Grids*, 104–111. doi:10.1109/SKG.2012.17

Hand Gesture Recognition Using Optimized Neural Network Shape Fitting On ARM11

Heri Setiawan[1], Iwan Setyawan[2], Saptadi Nugroho[3]
Electronics and Computer Engineering Department
Satya Wacana Christian University
Salatiga, Indonesia
[1]heri33s@gmail.com, [2]iwan.setyawan@ieee.org, [3]saptadi_nugroho@yahoo.com

Abstract— **Various methods of hand gesture recognition have been proposed in the literature, with high recognition rate. But implementing these methods in embedded system is still challenging since image processing applications needs a high-performance processor. In this paper, a hand gesture recognition system is implemented on a system with an OK6410B board. This board has a processor that runs at 532 MHz, which is relatively high for a small processor. The hand gesture recognition method proposed in this paper is based on the Neural Network Shape Fitting. In this paper we propose some modifications to this method. The modifications were pixel randomizing during the initialization step, addition of several neurons in the iterations, using lookup table for distance measurement and simplification of the finger detection. These modifications yielded a faster processing time (0.95s on the OK6410B) and a higher recognition rate (94.44% using still images as input and 84.53% using live input from a webcam).**

Keywords-hand gesture recognition; finger detection; image processing; Neural Network Shape Fitting

I. INTRODUCTION

Various hand gesture recognition methods have been proposed in the literature. Example: shape-based methods [1], graph-matching methods [2], and edge detection methods [3], [4], [5]. The main aim of these methods is achieving high recognition rate. However, these methods are not yet widely implemented as the main user interface to devices other than the Personal Computer which is big. One of the reasons is due to the fact that hand gesture recognition involves image processing which needs a high-performance processor, which is commonly found only in a Personal Computer.

Some robots need to recognize hand gestures so the interaction between human and robots is more natural like interaction between people. The processor of the robots are designed as small as possible so it only needs a small space. Thus, a small system embedded is required. The system embedded has disadvantages i.e., processor is than a Personal Computer and the resources like memory are limited. The advantages of small system embedded compared to a Personal Computer are small and low in power consumption.

In this paper, we implement a hand gesture recognition system on an ARM11 (Advanced RISC Machine) development board OK6410B. This development board contains a Samsung S3C6410 processor running at 532 MHz, which is relatively high for a small processor. The board also has 256 MB of RAM and 2 GB flash memory. The operating system is Linux 3.0.1. The software used to recognize the hand gesture is implemented using C cross-compiler GCC version 4.2.2 to run on embedded Linux.

The hand gesture recognition method proposed in this paper is based on the Neural Network Shape Fitting method developed by Stergiopoulou and Papamarkos (2009) [6]. In this approach, the authors proposed a novel Self-Growing and Self-Organizing Neural Gas (SGONG) which combines the advantages of both the Kohonen Self-Organizing Feature Map (SOFM) [7] and the Growing Neural Gas (GNG) [8]. The GNG was used in vector quantization, clustering, and interpolation [8] as in [9], [10], and [11]. The method developed by Stergiopoulou and Papamarkos had a 90.49% recognition rate [6] but it was very computationally intensive. The software implementation proposed in our paper is optimized to be more computationally efficient so that it will run faster than the original method. Furthermore, we have also implemented some customization to make the software run properly with the OK6410B board and a USB web camera.

The optimization and customization implemented in our approach include:

1. Changing the input of the system. In [6], the authors used still images as input, while in this paper the input is continuously captured using a web camera.
2. Parameter detected. The hand gesture to be recognized is only the number of finger raised.
3. Optimization of the calculation. We propose lookup table for used to measure the distance.
4. Reduction of the number of iterations. Some neurons are added at once in a loop.

In this paper, we define the hand gesture as the number of fingers the user has raised. Therefore, there are 5 possible gesture classes i.e., 1, 2, 3, 4, 5 with 31 possible combinations. The hand gesture recognition module send the output i.e., gesture class converted into the associated character (‘1’, ‘2’, ‘3’, ‘4’, and ‘5’) through the serial port. Therefore, we can control other devices that can communicate and receive commands through the serial port.

II. MATERIALS AND METHODS

In this section, we will discuss in more details the implementation of the proposed method. The overall flow chart of the system is presented in Fig. 1. The hand gesture given by the user is captured using an M-TECH USB web camera (shown in Fig. 2).

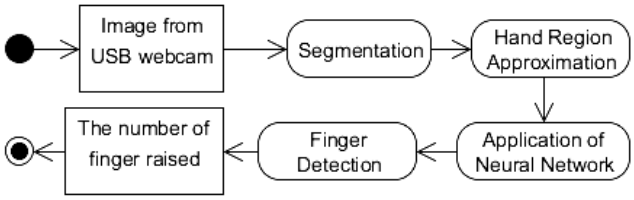


Figure 1. Hand gesture recognition flow chart.

Figure 2. M-TECH USB web camera.

The web camera is connected to the OK6410B development board (shown in Fig. 3). The captured image has a resolution of 320 × 240 pixels. This input image is then processed through the segmentation and hand region approximation steps, resulting in a binary segmented image. After performing this process, the Neural Network computation is applied. The result of this step is used to detect the fingers in the input image resulting in the number of fingers raised (and therefore the hand gesture class). These steps are discussed in the following sub-sections. Furthermore, in the final part of this section we shall discuss the details of adaptations and optimizations which are performed to make the proposed method run well on the OK6410B development board.

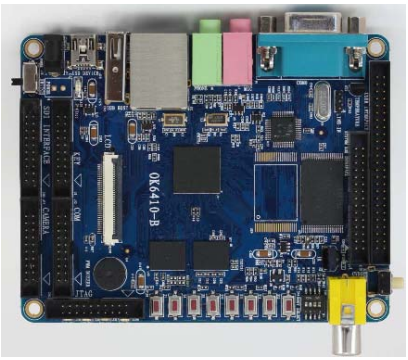


Figure 3. OK6410 development board.

A. Segmentation

The segmentation step is performed to separate the hand regions from the background in the input image. In our proposed method, the segmentation is based on the color of the hand in YUV color space. We use a thresholding method on the image chroma components (i.e., U and V components) to perform the segmentation. The upper and lower thresholds for the U and V components are respectively {80, 130} and {130, 165}. These thresholds are determined empirically. A sample of input image used in the segmentation process and the result of the segmentation step are shown in Fig. 4.

Figure 4. (a) A sample of input image in the segmentation process and (b) hand segmentation result by the hand gesture recognizer.

B. Hand Region Approximation

The regions of interest in our proposed hand gesture recognition system are the palm and fingers area. Therefore, this step of the hand region approximation process is performed to find the position of these areas for further processing.

The hand region approximation is performed by identifying the position of the wrist. Since when the arm is raised then any finger(s) will be the highest points in the image (i.e., closest to the upper border of the image), the position of the wrist will become the lower boundary of the region of interest. The wrist position is detected using the following steps.

1. Calculate the horizontal projection of the binary segmented image.
2. Find the global maximum value of the projection. The global maximum value of the projection represents the middle of the palm region.
3. Find the local minimum value of the projection below this global maximum value of the projection. The position of this local minimum value of the projection is the wrist position.

These steps of wrist position detection are illustrated in Fig. 5, with wrist detected vertically at red line position.

Figure 5. (a) Segmented image, (b) horisontal projection, and (c) wrist detection.

C. *Application of Neural Network*

This process of Neural Network Shape Fitting is performed in order to form the neuron map on the hand region found by the previous process. This process is adapted from the method proposed by Stergiopoulou and Papamarkos [6]. The output of this process is a mapping of neurons with edges connecting the neurons. The output of this process is shown in Fig. 6.

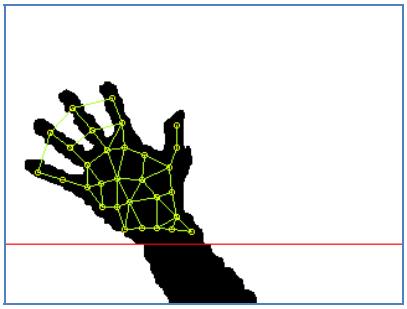


Figure 6. Neural Network Shape Fitting applied to the hand region.

D. *Finger Detection*

The result of neuron mapping from the previous process contains two kinds of edges, namely the edges that lie completely within the hand region and those that pass through the background region. The fingertips in the image can be detected using the following steps:

1. Remove the edges that go through the background.
2. Find the neurons that are connected to their neighbor by exactly one edge. These neurons must also be spatially above the neighbor to which it is connected. These neurons are the fingertips.

In order to determine whether or not an edge goes through the background, we do not exhaustively check every pixel on the edge. Instead we only check the pixel that lies in the middle of an edge connecting two neurons. Fig. 7 shows a part of image with the middle of edges. The result of finger detection process is shown in Fig. 8.

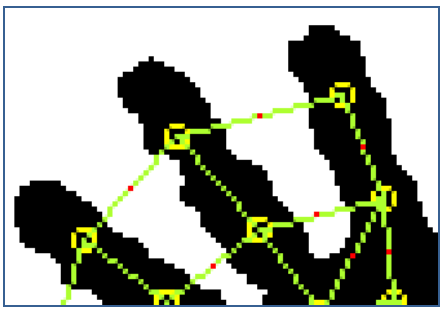


Figure 7. The middle of edges (marked as red dots)

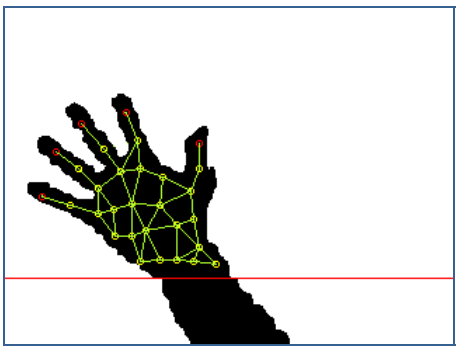


Figure 8. Finger Detection (Fingertips neurons are marked as red nodes)

E. *Adaptation and Optimization*

In [6], the authors reported that their implementation of SGONG method, running on a computer with a 3 GHz CPU, requires average 1.5s to process each input image. In this paper, the method is implemented on an ARM11 development board with a 532 MHz CPU. Therefore, in order to achieve a reasonable processing time, the method must be optimized and adapted to this hardware configuration.

The method which are adapted from [6] were optimized mainly at the Application of Neural Network step. Because the Application of Neural Network step are the most complex step of the whole method and spent the most processing time.

The first modification is in the randomization of pixels. In the original implementation [6], a random pixel is picked in each iteration of neural network computation. This is done in order to make sure that the neurons spread evenly in the desired area. This randomization process is replaced in our implementation by a pixel shuffling process performed during the initialization step of the program. This way, the program does not need to randomize the pixels in each iteration and therefore speeding up the processing time.

The second modification is in how we increase the number of neurons. The number of neurons at the beginning of the process is 21. The number of neurons is increased one at a time in the original implementation [6]. In our implementation, we add three neurons in one iteration. This reduces the number of iterations needed and speeds up the processing time. One loop is performed at the end of the process for error correction to keep the neurons evenly spread.

The third modification is done to the distance calculations. The distance calculation is performed in each iteration to calculate the distance of two different points in the image. This process needs floating point operations which are slowing down the overall process. Therefore, in our implementation we perform the calculation during the initialization process and construct a look up table. To construct the look up table, all of possible distances in an image are calculated using the difference of the x– and y–axis. During the iterations, the program only needs to access the index using the difference of the x– and y–axis to find the value i.e. distance between two points.

The final modification is in the manner with which we eliminate edges that go through the background. The details of this modification are explained in the previous sub-section.

Finally, another difference from our implementation and that of [6] is in the classification of the hand gestures. The implementation in [6] detects which finger is raised and all the possible combinations thereof. In our implementation, we simplify this such that the system only detects the number of fingers raised, regardless of which finger(s) or combinations of fingers are raised. Therefore, our system also does not differentiate between the left or right hand or whether the palm faces the camera or not. The only requirement is that the fingers are visible to the camera.

III. RESULTS AND ANALYSIS

In this project, the hand gestures are classified by the number of raised fingers, i.e., 1, 2, 3, 4, and 5. The inputs of the hand gesture recognizer are video frames or still images, i.e. hand gesture images taken by using an USB web camera. The lighting conditions affect the hand segmentation process. The hand segmentation process works in the intensity of light between 20 - 200 lux. Fig. 9 shows the results of the hand segmentation process in different light intensities.

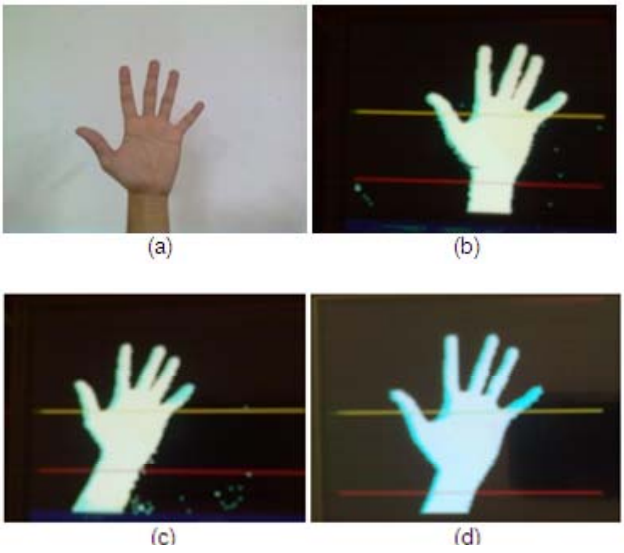


Figure 9. (a) An input image and hand segmentation process results at (b) 20lx, (c)100lx, and (d) 200lx.

The distance between hand and camera affects the hand gesture recognition result. The testers had different size of hands. The length of hand was measured from the wrist to the middle finger tip. The shortest hand is 17 cm length and the longest hand is 20 cm length. The hand gesture recognition rate is above 80% at distance between 40 cm and 60 cm. It will be worse at farther distance. Fig. 10 shows the hand gesture recognition rate versus the distance between hand and camera. The hand gesture recognizer does not work at the distances of 80 cm and 90 cm for the shortest and longest hand respectively.

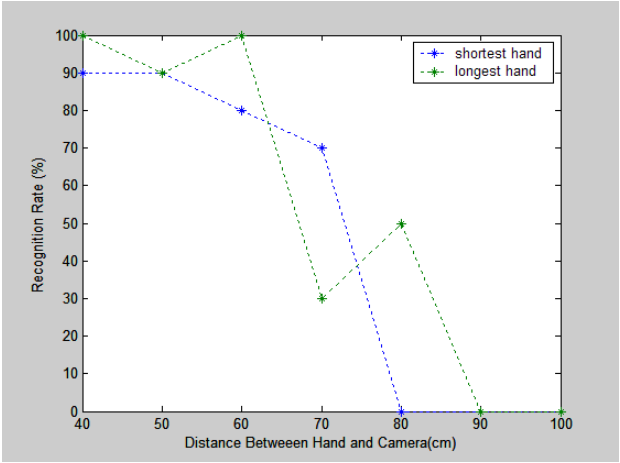


Figure 10. The hand gesture recognition rate versus the distance between hand and camera.

The first experiment was done using 180 hand gesture images given by 10 persons which are taken using a web camera under controlled condition and a uniform white background. The hand gestures given by testers had no fingers which are stick together. The gesture were counted as recognized-correctly if the system resulted the same number as the number of fingers raised. To calculate the recognition rate of the first experiment (R_1), the number of gesture images recognized correctly (r_1) was divided by the total number of hand gesture images (t) as in (1).

$$R_1 = (r_1 / t) \times 100\% \quad (1)$$

The hand gesture recognition rate from the first experiment approaches 94.44%.

In the second experiment, the distance between hand and camera was varied between 40 cm and 60 cm. This experiment was done using 750 hand gesture images given by 30 persons who have different size of hands. Each person gave 25 hand gestures. The second experiment was also conducted in the same condition as the first experiment. For the second experiment, the output of the system were taken from the same results of 3 consecutive images received from camera. If there were not same results of 3 consecutive images until the result of the 10^{th} image then the result for the gesture given at the time were counted as not recognized. The output were taken in this manner so the gestures changes should not counted as output. The gesture were counted as recognized-correctly if the system resulted the same number as the number of fingers raised. To calculate the recognition rate of the second

experiment (R_2), the number of hand gestures recognized correctly (r_2) was divided by the total number of hand gestures given, i.e. 30×25 as in (2).

$$R_2 = (r_2 / (30\times25)) \times 100\% \quad (2)$$

In the second experiment, the hand gesture recognizer can recognize the hand gesture at a rate of about 84.53% correctly.

The processing time of the hand gesture recognizer to process each frame depends on the number of pixels classified as skin color at the hand segmentation process. Too many pixels can slow down the next process i.e., Neural Network Shape Fitting process. So some pixels are taken to be the samples. If too few pixels are sampled, then the hand shape will not represented by the pixels. The maximum number of pixels that are sampled to become input in the Neural Network Shape Fitting process is 10,000, as determined empirically.

The average processing time of Optimized Neural Network Shape Fitting method is 0.95s. The result of the segmentation and Neural Network Shape Fitting processing time measurement is shown in Fig. 11.

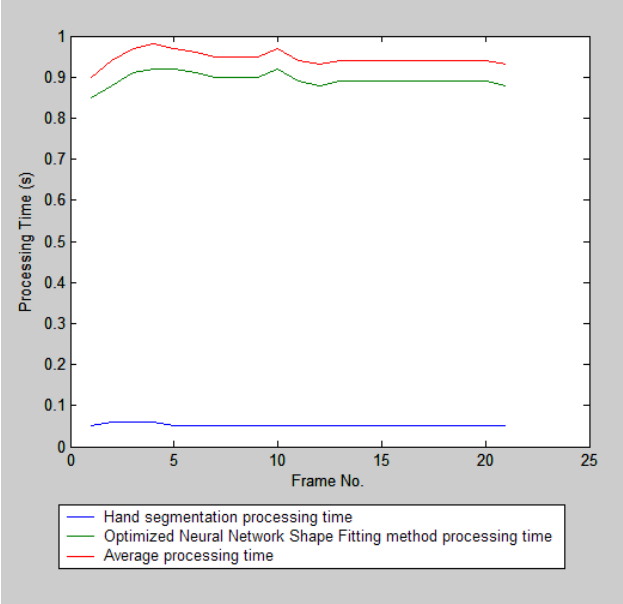


Figure 11. Processing time of hand segmentation an and Optimized Neural Network Shape Fitting of each frame of images.

The processing time of Optimized Neural Network measurement results shows the improvement of this method in terms of time. With the optimization of calculation and iterations reduction, the processing time of Optimized Neural Network reduced from the result in [4], i.e. 1.5 s using 3GHz CPU to 0.95 s using 532 MHz ARM11.

The hand area could not be retrieved correctly in the image with various backgrounds during the hand segmentation process because there were any objects which its color is similar to the skin color at the background of image. Fig. 12 shows the input images and the results of segmentation process at various backgrounds. The more other objects were detected as hands at the background of image, the more hands gesture recognizer could not form the hand morphology correctly.

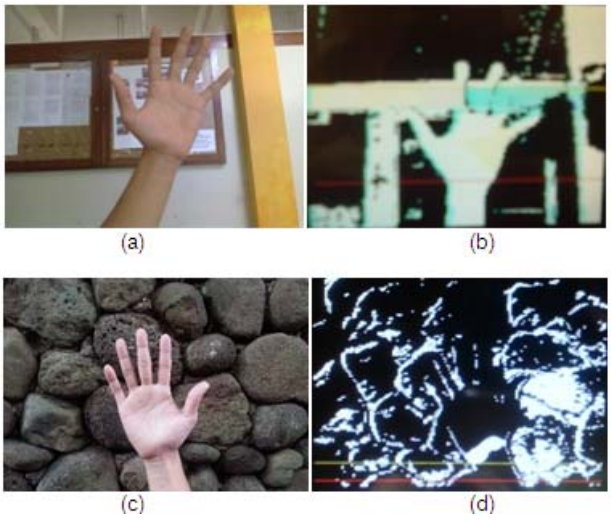


Figure 12. The input images (a), (c) and the results of segmentation process at various backgrounds (b), (d).

IV. CONCLUSION

The hand gestures recognizer which implements an Optimized Neural Network Shape Fitting method runs on an ARM11 development board OK6410B which has a speed of 532 MHz. To recognize the hand gestures, any gloves or special markers are not needed during the hand gesture recognition process. The Neural Network Shape Fitting method is optimized by randomizing pixel coordinates at initialization, adding some neuron in an iteration, distance calculation using lookup table, and simplifying finger detection. The execution time of the hand gesture recognizer running on OK6410B is 0.95 seconds on average with recognition rate 94.44% using still images input and 84.53% using live input from a webcam.

The difference between the processing time of previous works of [4] and the processing time of Optimized Neural Network Shape Fitting is significant. The processing time of Optimized Neural Network Shape Fitting is 0.95 s using 532 MHZ ARM11. It is faster than the processing time of [4] which is only 1.5 s using 3 GHz CPU.

For the future of work, the hand segmentation method could be improved in a such way that it will become more robust to skin-colored object at background.

REFERENCES

[1] E. Yoruk, E. Konukoglu, B. Sankur, and J. Darbon, "Shape-based hand recognition," IEEE Transactions on Image Processing 15, no. 7, pp. 1803–1815, 2006.

[2] J. Triesch, C. von der Malsburg, "Classification of hand postures against complex backgrounds using elastic graph matching," Image and Vision Computing 20, no. 13, pp. 937–943, 2002.

[3] Dr. Jane J. Stephan and Sana'a Khudayer, "Gesture recognition for Human-Computer Interaction (HCI)," International Journal of Advancements in Computing Technology, vol. 2, no. 4, pp. 30–35, October 2010.

[4] Ravikiran J, Kavi Mahesh, Suhas Mahishi, Dheeraj R, Sudheender S, Nitin V Pujari, "Finger detection for sign language recognition," Proceedings of the International Multi Conference of Engineers and Computer Scientists 2009, vol I, pp. 489–493, March 2009.

[5] R. Lionnie, I. K. Timotius, and I. Setyawan, "An analysis of edge detection as a feature extractor in a hand gesture recognition system based on nearest neighbor," Electrical Engineering and Informatics (ICEEI), 2011 International Conference on. IEEE, pp.1–4, 2011.

[6] E. Stergiopoulou, N. Papamarkos, "Hand gesture recognition using a neural network shape fitting technique," Engineering Applications of Artificial Intelligence, vol. 22, Issue 8, pp. 1141–1158, Dec 2009.

[7] T. Kohonen, "The self-organizing map," Proc. IEEE, vol. 78, no. 9, pp.1464 –1480, 1990.

[8] B. Fritzke, "A growing neural gas network learns topologies," Advances in neural information processing systems 7, pp. 625–632, 1995.

[9] K. Garcia, and C. H. Q. Forster, "Supervised growing neural gas," Intelligent Data Engineering and Automated Learning-IDEAL 2012. Springer Berlin Heidelberg, pp. 502–507, 2012.

[10] H. Sasaki, N. Kubota, K. Sekiyama, and T. Fukuda, "Multiple object detection for intelligent robot vision by using growing neural gas," Micro-NanoMechatronics and Human Science, IEEE International Symposium, pp. 80–85, 2009.

[11] F. Flentge, "Locally weighted interpolating growing neural gas," Neural Networks, IEEE Transactions on 17, no. 6, pp. 1382–1393, 2006.

Improved Thai Text Detection from Natural Scenes

Kuntpong Woraratpanya[1]
Faculty of Information Technology,
King Mongkut's Institute of Technology Ladkrabang,
Bangkok, Thailand 10520
Email: kuntpong@it.kmitl.ac.th[1]

Pimlak Boonchukusol[2]
Faculty of Information Technology,
King Mongkut's Institute of Technology Ladkrabang,
Bangkok, Thailand 10520
Email: pimlak.b@gmail.com[2]

Yoshimitsu Kuroki[3]
Kurume National College of Technology,
Fukuoka, Japan 835–8555
Email: kuroki@kurume-nct.ac.jp[3]

Yasushi Kato[4]
Tsuruoka National College of Technology,
Tsuruoka, Yamagata, Japan 997–8511
Email: president@tsuruoka-nct.ac.jp[4]

***Abstract*—Thai text detection from natural scenes is still a challenging task for language translation applications, since there are many unsolved issues. Furthermore, the existing related works cannot completely detect Thai text. The main reason is that Thai text layout has vowels and tonal marks that differ from other languages. This paper proposes an approach to detect Thai text from natural scenes. The approach consists of two main procedures. (i) Fast boundary clustering algorithm decomposes scene features into multilayers, so that it is faster and easier to analyze Thai text characters. (ii) Modified connected component analysis method is applied to such scene features in order to detect Thai text boundaries. Based on 150 test images with 4,920 characters, the experimental results demonstrate that the proposed approach achieves the high average precision and recall, 0.80 and 0.90.**

Keywords—Thai text detection; natural scene; fast boundary clustering; modified connected component analysis

I. INTRODUCTION

Text labels are everywhere in our lives, such as traffic signs, place name signs, warning signs, and description signs. These texts in natural scenes are information that makes our live easier; therefore, Thai text detection is a very important procedure for applying to language translation applications for smart phone to help foreign tourists understand Thai language signs.

Automatic text detection from natural scenes is still a challenging task, since there are many unsolved issues, which are not only variety of text: font style, size, color, and orientation but also text intensity affected by lighting conditions: low light, highlight, and shadow. Thus, error rate of text detection from natural scenes is higher than that of OCR systems which work on scanned documents under controlled environment [1].

In natural scene images, text is usually printed as a string which contains more than one character. The string most likely appears similar in size, character color, background color, and character alignment. By nature of these features, the text detection without a character recognition can be performed by using adjacent character grouping and connected component labeling proposed by C. Yi et al. [2], and R. Chang [3], respectively. However, these approaches usually fail to remove other objects which resemble text characters. To solve this problem, the multilayer separation method introduced by M. Kumar et al. [4] was employed to split text characters from other objects. In the same way, C. Yi et al. [5] proposed boundary clustering to group the boundary pixels with bigram color uniformity on the border of text and attachment surface. These algorithms take more computational time in data preparation for clustering, since all neighborhood pixels of all edge pixels of each boundary are computed.

In this paper, we propose an approach to detect Thai text with multiple sizes and colors from natural scene images. This approach is composed of two main procedures: (i) fast boundary clustering and (ii) modified connected component analysis. In the former procedure, the features based on colors and positions are extracted. Then k-means algorithm is applied to such features in order to decompose character boundaries from other objects. In the latter procedure, the decomposed features are analyzed by using a structure of text characters, such as position, size, stroke width, distance between two components, and character alignment, in order to detect Thai text boundary. The proposed method improves the efficiency of Thai text detection from natural scene images.

The remainder of this paper is organized as follows. Fast boundary clustering and modified connected component analysis, which are two principal procedures of the proposed approach, are described in section II and section III, respectively. Section IV presents and discusses results of experiments. Finally, the conclusion is given in section V.

II. FAST BOUNDARY CLUSTERING

Boundary clustering is a procedure to decompose boundary objects into multilayers, which is very useful to accurately detect the target objects. C. Yi et al. [5] proposed boundary clustering to group the boundary pixels with bigram color uniformity on the border of text and attachment surface. This method effectively decomposes boundary objects, but it consumes a lot of computing time, especially complex back-

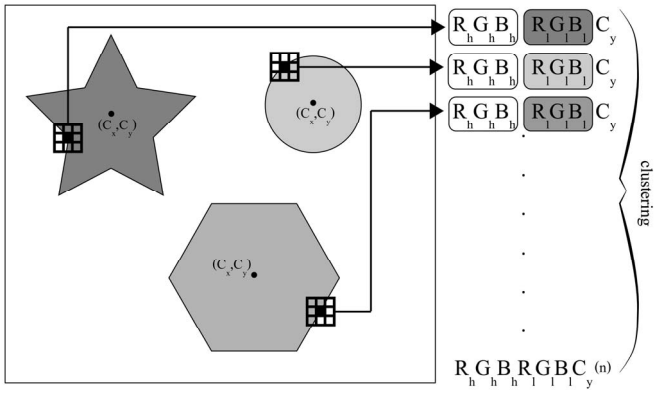


Fig. 1. Feature vector extraction based on the proposed assumptions.

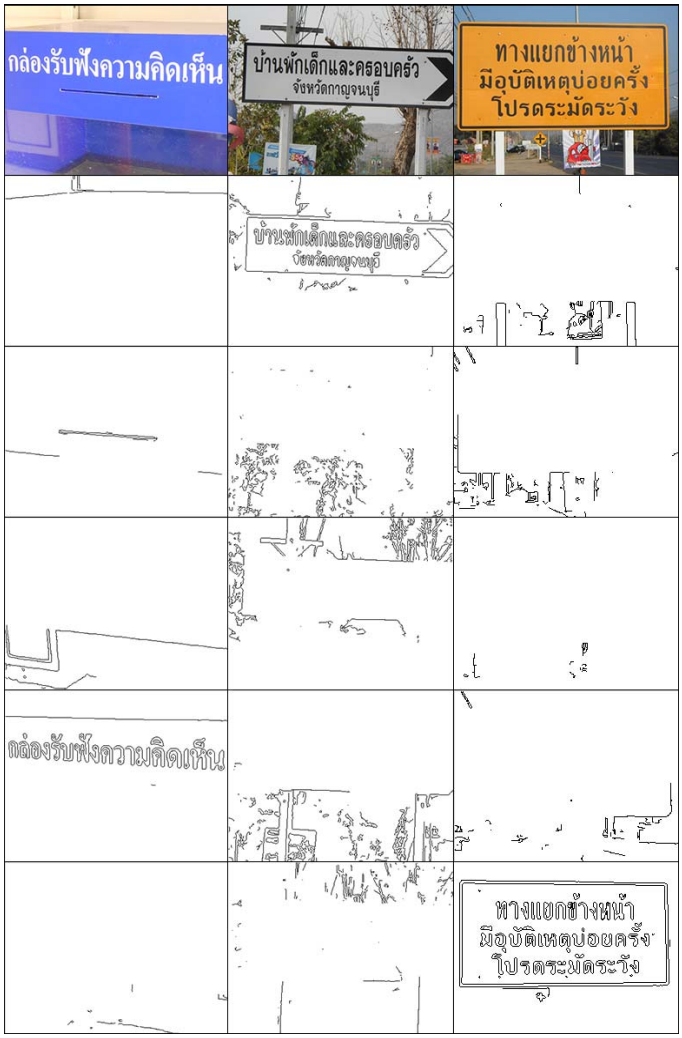


Fig. 2. Boundary clustering, the first row is three orginal images and other rows are boundary layers after clustering.

grounds. Therefore, in order to effectively separate Thai text strings from the complex backgrounds, this paper proposes the fast boundary clustering algorithm based on the essential assumptions described as follows:

- Characters in a same scene have almost the same foreground and background colors.
- Text foreground and background colors are normally a uniform color.

Based on these assumptions, the proposed approach can speed up by using the following two steps:

In the first step, Canny edge detector is used to obtain object edge sets. As mentioned by the above assumptions, the feature of each object edge is extracted by determining the maximum color difference between two opposite pixels on a 3×3 mask as demonstrated in Fig. 1. In this way, the feature vector, V, of each object edge is formed from three components, the highest color intensity, the lowest color intensity, and the spatial position of boundary, i.e., y-coordinate of the centroid of boundary. The vector can be defined as (1). The highest and lowest color intensities are the sum of z-scores of RGB values. The z-score equation is defined as (2).

$$V_i = [R_h, G_h, B_h, R_l, G_l, B_l, C_y]_i, \quad (1)$$

where $[R_h, G_h, B_h]$, $[R_l, G_l, B_l]$, and $[C_y]$ imply pixel color values of the highest intensity, pixel color values of the lowest intensity, and the spatial position, respectively.

$$z = \frac{x - \mu}{\sigma}, \quad (2)$$

where z, x, μ, and σ denotes standardized z-score, values of R, G, or B, mean of each color value, and standard deviation of each color value, respectively.

In the second step, the feature vectors are decomposed into multilayers by means of k-means clustering algorithm. The number of clusters, K, is an important factor. If K is too small, text boundaries are not effectively extracted from complex backgrounds. On the other hand, if K is too large, the text boundaries are probably broken into several fragments and assigned into different boundary layers. Fig. 2 illustrates three examples of boundary layers after clustering. Edge pixels with similar color pairs and spatial positions are perfectly grouped into the same layer. This shows the best results of applying the clustering algorithm proposed in [5], when the number of clusters is equal to 5.

III. Modified Connected Component Analysis

Connected component analysis is a procedure to detect text characters and filter unwanted objects out. C. Yi et al. [2] proposed adjacent character grouping to decide whether three connected components are character sets and merge the character sets together. This method effectively detects characters, but its accuracy probably reduces after boundary clustering, when the connected characters are less than three characters on each boundary layers. R. Chang [3] proposed connected component labeling to filter unwanted objects based on image sizes, component sizes, and the number of edge pixels. This method is a preprocessing step for a recognition

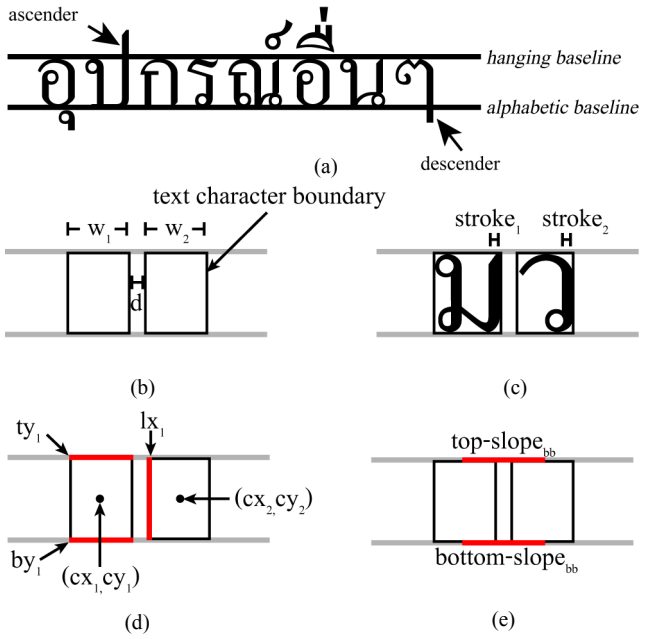


Fig. 3. Two-connected component detection model, (a) Thai word typography line, (b) width of text character boundary and distance between two boundaries, (c) stroke width of character, (d) borders and centroids of boundary, and (e) slopes between two boundaries.

phase to decide whether detected connected components is a character, but it cannot remove unwanted connected objects similar to connected characters. Moreover, these methods were designed to detect English characters only, thus they cannot completely detect Thai vowels and tonal marks.

Therefore, in order to effectively detect Thai characters on multilayers, this paper proposes the modified connected component analysis algorithm. Thai text strings in natural scene images usually appear in alignment, namely, each text character in a text string have to connect to another text character. The structure features among all components can be used to determine whether the connected components belong to text characters or unexpected noises. In the same way, most Thai typography characters, including 44 consonants, 19 vowels, and 4 tonal marks, are printed as a text string, which contains two or more characters. Therefore, to identify two connected components and locate its alphabetic and hanging baselines as shown in Fig. 3(a), the essential constraints are defined and described below; the constraint parameters are described in Table I; and a threshold parameter T_n with the suitable value is defined in Table II.

- Thai characters commonly align on the alphabetic baseline, such that $1/T_1 \le w_1/w_2 \le T_1$ as shown in Fig. 3(b).
- A gap of two connected characters is not far from each other character, such that $d/w_1<1/T_2$ and $d/w_2<1/T_2$ as shown in Fig. 3(b).
- Two connected characters usually appear in the same font size, such that $(|stroke_1–stroke_2|/stroke_1) < 1/T_2$ and $(|stroke_1 – stroke_2|/stroke_2) < 1/T_2$ as shown in Fig. 3(c).

TABLE I. TERMINOLOGIES OF TWO-CONNECTED DETECTION MODEL

Parameter	Description
alphabetic baseline	A line which most characters align on
hanging baseline	A line is above alphabetic baseline with character height
ascender	A part of characters that extend above the hanging baseline
descender	A part of characters that extend below the alphabetic baseline
ty	y-coordinate of top-border of boundary
by	y-coordinate of bottom-border of boundary
lx	x-coordinate of left-border of boundary
Ty	y-coordinate of top-border of text string boundary
By	y-coordinate of bottom-border of text string boundary
(cx,cy)	centroid of boundary
(Cx,Cy)	centroid of text string boundary
w	width of boundary
h	height of boundary
W	width of two connected boundaries
H	height of two connected boundaries
stroke	stroke width of a character
d	distance between two connected boundaries
$FColor_1$	average RGB values of foreground of text string boundary
$FColor_2$	average RGB values of background of text string boundary
$slope_{bb}$	slope between two boundaries
$slope_{cb}$	slope between two-connected boundary and another boundary
$slope_B$	slope of text string boundary
$slope_{BB}$	slope between two text string boundaries

- Two-connected character area can be slightly overlapped, such that $cx_1 < lx_2$ as shown in Fig. 3(d).
- Two connected characters are usually aligned in a horizontal direction, such that $ty_1<cy_2<by_1$ as shown in Fig. 3(d).
- Two connected characters without ascender or descender usually align between the alphabetic baseline and the hanging baseline, such that $|top\text{-}slope_{bb} - bottom\text{-}slope_{bb}| < T_3$ as shown in Fig. 3(e).

After alphabetic baseline and hanging baseline of each text string are located, left and right projections are used to determine whether the remaining boundaries in the same layer belong to characters in the same text string as depicted in Fig. 4(a). The starting point is the minimum x-coordinate of the detected text string boundary for the left projection and the maximum x-coordinate of the detected text string boundary for the right projection. Text characters and its background usually have distinct intensity values, thus the end point is the x-coordinate of the pixel whose the intensity value differs

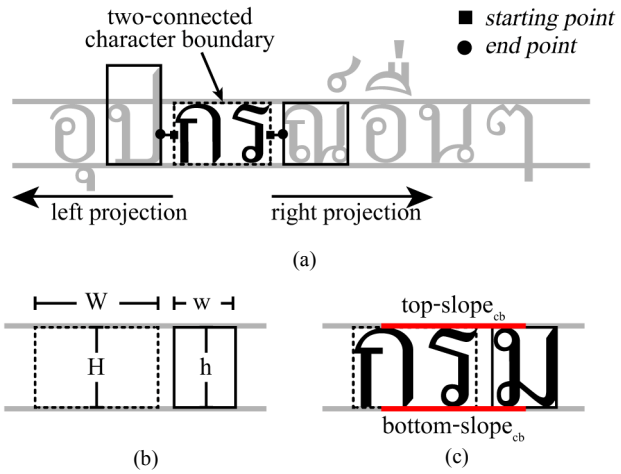


Fig. 4. Thai character boundary detection using projection model, (a) direction of projection and position of starting and end points, (b) size of two-connected boundary and other boundaries, and (c) slopes between two-connected boundary and another boundary.

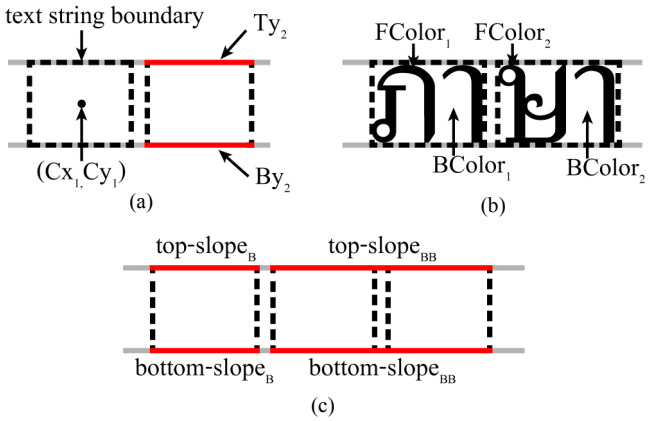


Fig. 5. Text string boundary grouping model, (a) borders and centroid of text string boundary, (b) foreground and background colors of text string boundary, and (c) slopes of text string boundary and slopes between two text string boundaries.

from previous pixel greater than T_4. The starting point and end point of the left projection are y-coordinate of the center between the hanging baseline and the alphabetic baseline. If the end point is overlapped in a boundary, this boundary is determined whether it is a character or an unwanted object by using the following constraints.

- $1/T_1 \le w/W \le T_1$ as shown in Fig. 4 (b).
- $1/T_5 \le h/H \le T_2$ as shown in Fig. 4 (b).
- Either top border or bottom border of each Thai character boundary must be connected to the alphabetic baseline or the hanging baseline, such that $|\text{top-slope}_{cb} - \text{top-slope}_{bb}| < T_3$ or $|\text{bottom-slope}_{cb} - \text{bottom-slope}_{bb}| < T_3$ as shown in Fig. 4(c).

The projection procedure continuously performs until no more characters in the same text string is detected. However, each text string is probably assigned into different boundary layers, especially natural scene images with low background

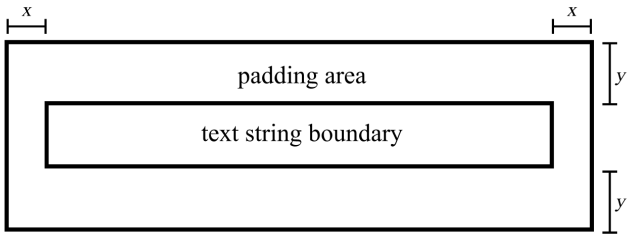


Fig. 6. The padding area used to detect Thai vowels and tonal marks.

complexity. Therefore, the constraints to group these text strings together are defined as follows:

- A text string is usually aligned in horizontal direction, such that $Ty_2<Cy_1<By_2$ as shown in Fig. 5(a).
- A text string usually appears in the same foreground and background colors, such that $|FColor_1 - FColor_2| \le T_4$ and $|BColor_1 - BColor_2| \le T_4$ as shown in Fig. 5(b).
- A text string has only one alphabetic baseline and one hanging line, such that $|\text{top-slope}_B - \text{top-slope}_{BB}| < T_3$ and $|\text{bottom-slope}_B - \text{bottom-slope}_{BB}| < T_3$ as shown in Fig. 5(c).

In order to increase the accuracy of detecting characters, which align on the alphabetic baseline, the left and right projections are used again. Furthermore, remaining boundaries that appear between left and right borders of each text string boundary are also detected.

However, in some cases, Thai characters do not align on the alphabetic baseline, since some vowels are above the hanging baseline or below the alphabetic baseline. Furthermore, all tonal marks always are above the hanging baseline. We determine whether the remaining boundaries belong to Thai characters that align above or lower the detected text string by using text string boundary surrounding padding. If a component is within the padding area, this component should be a character. The padding area size is specified as shown in Fig. 6, where *x* represents the sum of average boundary width and average distance between two connected boundaries and *y* represents average boundary height.

IV. Experiments

In order to evaluate the efficiency of the proposed method in terms of precision, recall, and computing time, the experimental methods are setup as follows.

A. Dataset

There are two experimental datasets. The first dataset consists of 100 test images which contain 1 Thai character per each image. The pixel resolutions of test images are 64×64. The second dataset comprises of 150 scene images with 4,920 character boundaries. Each image consists of one to four text strings and all of text strings are horizontal. This dataset is classified into three groups based on background complexity. The pixel resolution of test images is 480×360.

TABLE II. PARAMETER VALUES FOR EXPERIMENTS

Parameter	T_1	T_2	T_3	T_4	T_5
Threshold Value	7	2	0.05	50	4

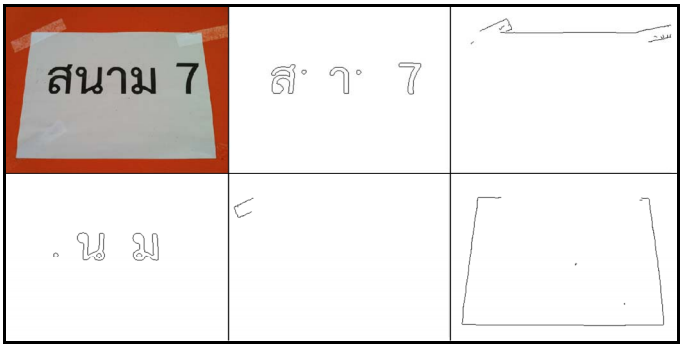


Fig. 7. A failure result from FBC method affected to accuracy of Thai text detection.

B. Performance Evaluation

The performance is evaluated by two metrics, *precision* and *recall*, defined as (3) and (4), respectively.

$$precision = \frac{tp}{tp + fp} \tag{3}$$

and

$$recall = \frac{tp}{tp + fn}, \tag{4}$$

where *tp*, *fp*, and *fn* represent a number of successful detected characters, a number of detected noises, and a number of unsuccessful detected characters, respectively. In other words, precision is the ratio of the number of successful detected text characters to the total number of detected characters, whereas recall is the ratio of the number of successful detected characters to the total number of characters. The low precision implies overestimation while the low recall implies underestimation. The significant parameters of experiments are set up in Table II.

C. Results and Discussions

In order to evaluate the efficiency of the proposed method, three experiments are setup. The first experiments aim to test the fast boundary clustering (FBC) algorithm in terms of computing time and an error rate of the feature extraction, respectively. The second experiment aims to test Thai text detection. This experiment compares the results of a connected component analysis (CCA) with a modified connected component analysis (MCCA) method in terms of precision and recall. Finally, the last experiment aims to test Thai text detection without character recognition by using a combination of FBC and MCCA methods in terms of precision and recall. The first dataset is used for testing an error rate of the feature extraction and the second dataset is used for the others.

TABLE III. COMPARISON OF COMPUTING TIME OF BC AND FBC ALGORITHMS

Background Complexity	Average Processing Time (Sec/Image)	
	BC	*FBC*
Low	98.09	0.98
Medium	184.13	2.46
High	241.76	3.57

TABLE IV. COMPARISON OF EVALUATION RESULTS CCA AND MCCA ALGORITHMS

Background Complexity	Performance			
	CCA		*MCCA*	
	Precision	*Recall*	*Precision*	*Recall*
Low	0.77	0.93	**0.89**	**0.94**
Medium	0.42	0.90	**0.77**	**0.92**
High	0.33	0.81	**0.66**	**0.85**

TABLE V. EVALUATION RESULTS OF A COMBINATION OF FBC AND MCCA ALGORITHMS

Background Complexity	Performance	
	Precision	*Recall*
Low	**0.90**	0.93
Medium	**0.81**	0.92
High	**0.70**	0.84

In the first experiments, the boundary clustering (BC) algorithm proposed in [5] is implemented as a baseline. This algorithm extracts color features by determining all neighborhood pixels of all edges. In comparison of average processing time, the experimental results show that the BC algorithm takes an average 174.66 seconds/image whereas the FBC takes an average 2.34 seconds/image as presented in Table III. Furthermore, the error rate of color feature extraction is 2.42% when compared to the baseline algorithm. This shows that the FBC algorithm is not only faster but also more effective to extract color features.

In the second experiment, CCA algorithm proposed in [6] is implemented as a baseline. Table IV presents a comparison of Thai text detection with CCA and MCCA algorithms. It is clearly that MCCA achieves better precision and recall than CCA algorithm. Particularly, in medium and high background complexities, the precisions of the MCCA algorithm are greatly improved.

In the last experiment, a combination of FBC and MCCA methods is evaluated. Table V shows the experimental results. It is evident that the proposed method achieves a better precision. An important reason to support this achievement is that the FBC algorithm effectively decomposes character boundaries from other unwanted boundaries. This makes it easier to detect two connected character boundaries. Moreover, the FBC algorithm help improve the precision affected

Fig. 8. A part of experimental results of Thai text detection marked in cyan.

by background complexity, since scene images with high background complexity have more objects that resemble Thai text characters. In terms of recall, both MCCA algorithm and MCCA with FBC algorithm yields almost the same results. However, as text boundaries are probably broken into several fragments and assigned into different boundary layers, it is easy to fail to detect text characters. Fig. 7 shows an example of a failure result from FBC algorithm. In this case, the objects are decomposed into five layers and each layer does not contain two-connected component under the defined constraints in section III, thus MCCA algorithm fails to detect Thai text. Fig. 8 shows a part of Thai text detection results.

V. CONCLUSION

In this paper, we have proposed an approach to detect Thai text from natural scenes. This approach consists of two main procedures: fast boundary clustering (FBC) and modified connected component analysis (MCCA). The FBC procedure uses for separating text characters from other objects and the MCCA procedure uses for determining whether which object is Thai text character. The combination of FBC and MCCA procedures improves the precision of medium and high background complexities. Based on experimental results, the proposed approach achieves the high average precision and recall.

VI. REFERENCES

[1] X. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic Detection and Recognition of Signs From Natural Scenes," IEEE Trans. on Image Process., vol. 13, no. 1, pp. 87-99, January 2004.

[2] C. Yi and Y. Tian, "Text String Detection From Natural Scenes by Structure-Based Partition and Grouping," IEEE Trans. on Image Process., vol. 20, no. 9, pp. 2594-2605, September 2011.

[3] R. Chang, "Intelligent Text Detection and Extraction from Natural Scene Images," Proc. 15th North-East Asia Symposium on Nano, Information Technologt and Reliability, pp. 23-28, 2011.

[4] M. Kumar, Y. C. Kim, and G. S. Lee, "Text Detection using Multilayer Separation in Real Scene Images," Proc. 10th IEEE International Conference on Computer and Information Technology, pp. 1413-1417, 2010.

[5] C. Yi and Y. Tian, "Localizing Text in Scene Images by Boundary Clustering, Stroke Segmentation, and String Fragment Classification," IEEE Trans. on Image Process., vol. 21, no. 9, pp. 4256-4268, September 2012.

[6] K. Woraratpanya and P. Boonchukusol, "Thai Text Detection from Natural Scenes by using Boundary Clustering and Connected Component Analysis," Proc. 9th National Conference on Computing and Information Technology, pp. 665-672, May 2013.

Integration of DEMATEL and ANP Methods for Calculate The Weight of Characteristics Software Quality Based Model ISO 9126

Sugiyanto
Department of Informatics
Institut Teknologi Adhi Tama Surabaya (ITATS)
Surabaya, Indonesia
sugiyanto.lecture@gmail.com

Siti Rochimah
Department of Informatics
Institut Teknologi Sepuluh Nopember (ITS)
Surabaya, Indonesia
siti@its-sby.edu

***Abstract*— One of the difficulties that occur in the model is to decide the weights of quality characteristics. This is due to the interrelations existence among the quality factors based model ISO 9126. Each of these characteristics can influence or even contradict each other. The interrelations existence among the factors affects the weight of characteristics software quality, and will affect the software quality calculation. Therefore, researchers will integrate DEMATEL and ANP methods for calculate the weight of characteristics software quality based model ISO 9126. DEMATEL method used to calculate sum of influences for each characteristics model ISO 9126, while the ANP method used to calculate local weights and global weight for each sub characteristics model ISO 9126. Results from this study is the value of local weights for each of the characteristics of ISO 9126, and global weights for each sub characteristics ISO 9126 which represent the level of importance of the characteristics and sub characteristics ISO 9126.**

Keywords— software quality, ISO 9126, DEMATEL, ANP

I. INTRODUCTION

The number of software products from abroad benefit users because of the large choice of products and prices. It is a challenge for the domestic software products to enhance their international competitiveness. Developers and software houses in Indonesia need to pay attention to software quality issues in order not to fall behind from abroad.

Measuring the quality of software is not an easy job. When someone gives very good value for a software, others not necessarily say the same thing. One's viewpoint can be oriented to one side of the problem, such as about the reliability and efficiency of the software, while others gives bad value for a software because it uses another viewpoint like aspect of the design and usability.

Developers can view software quality from the perspective of the software development process (process point of view) and the perspective of the resulting product (product point of view). This assessment is affects the orientation to how the software developed as expected by the user. From the product point of view, practitioners and software developers usually measured software quality using model ISO 9126. From the process point of view, practitioners and software developers usually measured software quality using standard model ISO 9001.

Reference [1] has used the AHP method to calculate software quality model ISO 9126. One of the difficulties that occur in the model is to decide the weights of quality characteristics. Reference [2] found how the interrelations existence among the quality factors based model ISO 9126 (fig. 1). Each of these characteristics can influence or even contradict each other, such as reliability between maintainability, or reliability between efficiency.

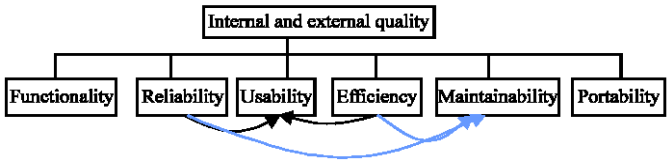


Fig.1. Interrelationship among the quality factors based model ISO 9126 [2]

The interrelations existence among the factors affects the weight of characteristics software quality, and will affect the software quality calculation. Therefore, researchers will integrate DEMATEL and ANP methods for calculate the weight of characteristics software quality based model ISO 9126. Researches chose this method and refer [3] and [4], which have successfully integrated the methods for determining the weights of criteria and sub-criteria in many cases.

II. LITERATURE REVIEW

A. ISO 9126

Standard ISO 9126 was first introduced in 1991 through the definition of software quality [1]. The original 13-page document is designed as a foundation for further, more detailed, and have a standard that can processed. ISO 9126 introducing quality type quality where the following elements are known:

1. Effectiveness is the ability to make the user goal through accuracy and completeness.
2. Productivity is an effort to avoid excess use of resources, such as staff costs in the user goal.

3. Safety is trying to avoid the risk level for the crime and other entities such as business, software, property and the environment
4. Satisfaction is user satisfaction in using the software.

ISO 9126 identifies six main characteristics of software quality are:

1. Functionality: the ability to cover the function of the software product that provides the user needs.
2. Reliability: the ability to care software with performance level.
3. Usability: skills related to the use of the software.
4. Efficiency: the ability to relate to the physical resources that are used when the software running.
5. Maintainability: the skills needed to make changes to software
6. Portability: the ability to relate the software that is sent to a different environment.

B. *DEMATEL (Decision Making Trial And Evaluation Laboratory)*

DEMATEL can be used to find and analyze the dominant characteristics of a system [5]. Through DEMATEL method, it can be seen that the dominant characteristic of the model of software quality, and most importantly will also note the relationship between characteristics that can later be used as input and decision-making models for ANP method.

Input in this method is a questionnaire that respondents completed. Data results of the questionnaire is qualitative data converted into quantitative. So we get the impact relations map (IRM) at the result, which is the basis for modeling the ANP.

In this process, there are several stages, start from :

1. Calculate the average first matrix using score. Sample of experts was asked to show a direct effect based on their perception that every element i influence every element other j, using a scale range from 0 to 4. High influence represented by 4, and no influence represented by 0.
2. Calculate the first influence matrix.
3. Creating relationship among the characteristic matrix in total.

To create a matrix of values that have been normalized using formula (1), (2), (3):

$$[d_{ij}]nxn = s[a_{ij}]nxn' \; s > 0, \qquad i,j \in \{1,2,\dots,n\} \tag{1}$$

$$s = Min\left[\frac{1}{\max\ 1 \le i \le n \sum_{i=1}^{n}|a_{ij}|}, \frac{1}{\max\ 1 \le i \le n \sum_{i=1}^{n}|a_{ij}|}\right] \tag{2}$$

$$\lim_{m\to\infty} D^m = [0]_{nxn} \quad where\ D = [d_{ij}]_{nxn} \tag{3}$$

$$0 \le d_{ij} < 1$$

After getting a relationship matrix which has been normalized, then proceed with processing to obtain the total relationship matrix using formula (4), (5), (6), (7):

$$T = D + D^2 + \dots + D^m = D\,(I\text{-}D)^{-1} \text{ when } m \to \infty \tag{4}$$

$$T = [t_{ij}], \quad i,j = 1,2,\dots,n, \tag{5}$$

$$r = [r_i]_{nx1} = \left(\sum_{j=1}^{n} t_{ij}\right)_{nx1} \tag{6}$$

$$c = [c_i]'_{1xn} = \left(\sum_{i=1}^{n} t_{ij}\right)_{1xn} \tag{7}$$

4. Set the threshold value and generate impact relations map. This value is generated by taking into account the experts opinions are presented in the matrix element T.

C. *ANP (Analytic Network Process)*

ANP (Analytic Network Process) is an approach that allows the problem interdependence completion among the characteristics of alternative solutions [6]. The first stage of ANP is to compare the measurement criteria in the system to form a super matrix [7]. Here is the general form of the super matrix (8):

$$W = \begin{array}{c} \\ C_1 \\ C_2 \\ \vdots \\ C_n \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \cdots & C_n \\ e_{11}\dots e_{1m_1} & e_{21}\dots e_{2m_2} & & e_{n1}\dots e_{nm_n} \end{array} \\ \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1n} \\ W_{12} & W_{22} & \cdots & W_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ W_{n1} & W_{n2} & \cdots & W_{nn} \end{bmatrix} \end{array} \tag{8}$$

The second stage of ANP is to generate weighted super matrix using formula (9):

$$\lim_{k\to\infty} W^k \tag{9}$$

The third stage of ANP is raised limiting super matrix to calculate global weights using formula (10):

$$\lim_{k\to\infty} \left(\frac{1}{N}\right) \sum_{k=1}^{N} W^k \tag{10}$$

III. RESEARCH FRAMEWORK

A. *Flowchart System*

Flowchart of this system show in fig. 2. Users enter the matrix influence the characteristics model ISO 9126 (for the DEMATEL method). Users also include the value of pair-wise comparison sub characteristics model ISO 9126 (for the ANP method). The system will calculate the first influence matrix, and create linkages between the characteristic matrix. Furthermore, the system will generate a weighted super matrix, and generate limiting super matrix. Finally, the system will display the output value of the local weights and global weights of each characteristics ISO 9126.

The method used in this study include:

1. DEMATEL (Decision Making Trial And Evaluation Laboratory) used to calculate sum of influences for each characteristics model ISO 9126.
2. ANP (Analytic Network Process) used to calculate local weights and global weight for each sub characteristics model ISO 9126.

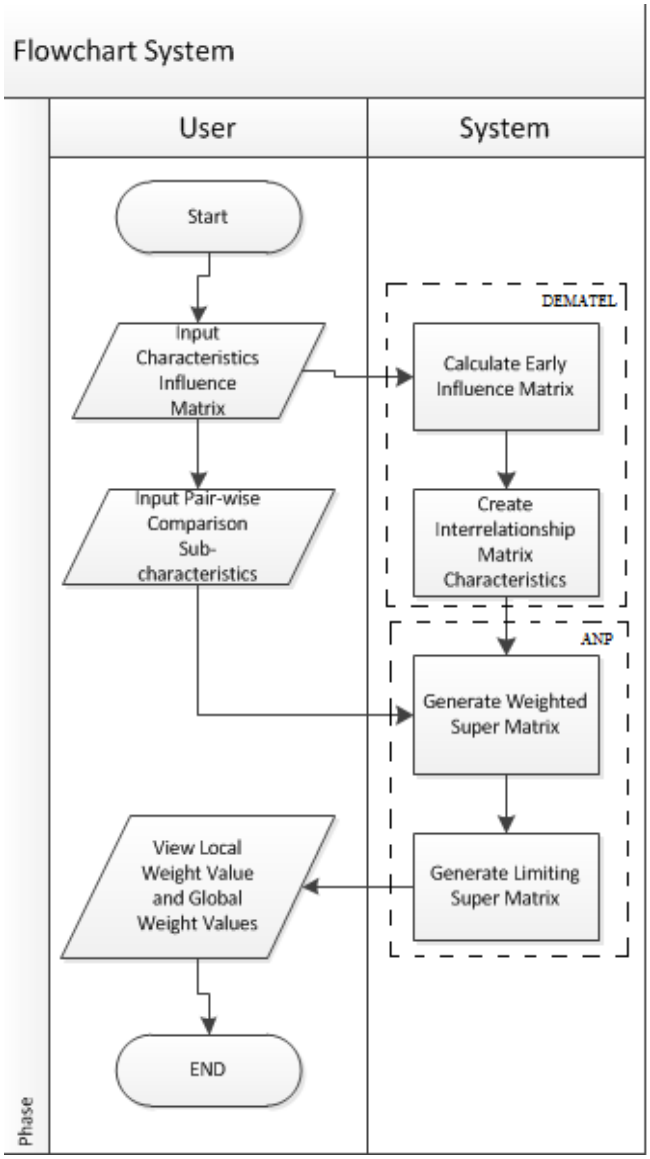


Fig.2. Flowchart system

B. Characteristics and sub characteristics model ISO 9126

Characteristics that considered in model ISO 9126 is 6 characteristics and 22 sub characteristics as shown in Table I.

TABLE I
CHARACTERISTICS AND SUB CHARACTERISTICS ISO 9126

Characteristic	No	Sub Characteristic
Functionality	1	Suitability
	2	Accurateness
	3	Interoperability
	4	Compliance
	5	Security
Reliability	6	Maturity
	7	Fault tolerance
	8	Recoverability
Usability	9	Understandability
	10	Learn ability
	11	Operability
	12	Attractiveness
Efficiency	13	Time Behavior
	14	Resource utilization
Maintainability	15	Analyzability
	16	Changeability
	17	Stability
	18	Testability
Portability	19	Adaptability
	20	Install ability
	21	Conformance
	22	Replace ability

C. Implementation

The first phase, users enter the matrix influences (fig. 3) based on their perception that every characteristic A effect on the other characteristic, using a scale ranging from 0 to 4. E.g. influences Reliability on Portability characteristics.

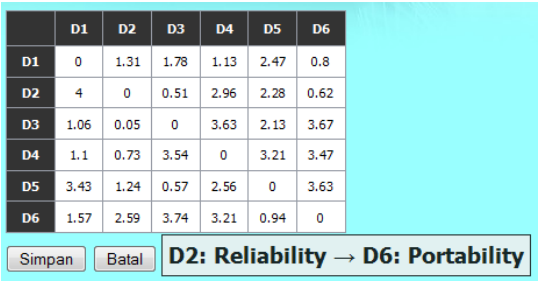


Fig. 3. Matrix influences form

Furthermore, the user enters the value of pair wise comparison of each sub characteristics contained in model ISO 9126 (fig. 4). E.g. influences Accurateness on Security.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22
C1	1	3	1/3	7	7	1/7	1/5	7	3	5	7	5	1/3	3	5	1/9	1/5	3	1/3	5	7	7
C2	1/3	1	1/9	7	1/9	9	1/5	3	1/7	3	1/5	9	1/7	7	5	5	1/7	1/7	5	5	1/5	3
C3	3	9	1	1/3	9	1/9	3	1/3	5	7	1/9	5	1/9	1/5	1/5	1/7	3	1/9	1/5	7	1/7	3
C4	1/7	1/7	3	1	7	9	3	7	3	5	7	1/9	1/3	1/7	1/7	5	1/7	9	3	3	1/5	1/5
C5	1/7	9	1/9	1/7	1	7	5	3	1/3	1/9	7	1/7	7	1/9	7	1/7	1/3	5	3	9	1/9	9
C6	7	1/9											1/7	1/5	1/7	5	3	7	9	3	3	3
C7	5	5											5	5	5	7	3	3	7	7	5	5
C8	1/7	1/3	3	1/7	1/3	3	7	1	3	9	1/7	1/7	1/5	5	1/5	1/9	5	5	1/9	1/9	1/7	1/3
C9	1/3	7	1/5	1/3	3	9	9	1/3	1	1/5	5	3	3	3	1/9	5	1/5	5	5	1/5	1/3	1/5
C10	1/5	1/3	1/7	1/5	9	1/5	5	1/9	5	1	7	1/7	7	5	1/5	1/9	5	7	3	3	3	5
C11	1/7	5	9	1/7	1/7	1/9	1/5	7	1/5	1/7	1	1/9	3	1/9	1/7	5	5	1/9	1/5	1/3	1/7	1/5
C12	1/5	1/9	1/5	9	7	1/3	3	7	1/3	7	9	1	1/9	1/3	1/9	1/9	1/3	3	7	1/5	9	1/7
C13	3	7	9	3	1/7	7	1/5	5	1/3	1/7	1/3	9	1	1/9	3	3	9	5	1/5	3	3	1/5
C14	1/3	1/7	5	7	9	5	1/5	1/5	1/3	1/5	9	3	9	1	7	9	3	1/5	7	3	1/3	1/5

C2: Accurateness → C5: Security

Simpan Batal

Fig. 4. Pair wise Comparison Sub characteristics ISO 9126

Click on a column filled with the value of pair wise comparison. Then, the questionnaire pair wise comparison form will appear (fig. 5). Users only need to enter the value of the interrelations of the ISO 9126 sub characteristic selected. E.g. influences Accurateness on Security. To help users, this form using the scroll bar that has a 2 way relationship that would represent the 2 sub characteristics model ISO 9126 interrelations existence.

Fig. 5. Pair wise Comparison Questionnaire

The questionnaire using membership function (Table II) as used by [3] and [4] so that the user can enter data pair wise comparison qualitatively.

TABLE II
MEMBERSHIP FUNCTION

Value	Linguistic Variabel
1	Equally important / preferred
3	Moderately important / preferred
5	Strongly important / preferred
7	Very strongly important / preferred
9	Extremely important / preferred

D. Output System

Output from this application are as follows (fig. 6):

Matrix A

	D1	D2	D3	D4	D5	D6
D1	0	1.31	1.78	1.13	2.47	0.8
D2	4	0	0.51	2.96	2.28	0.62
D3	1.06	0.05	0	3.63	2.13	3.67
D4	1.1	0.73	3.54	0	3.21	3.47
D5	3.43	1.24	0.57	2.56	0	3.63
D6	1.57	2.59	3.74	3.21	0.94	0

$s = \min(1 / 12.05, 1 / 13.49) = 1 / 12.05 = 0.0830$

Matrix $D = s \times A$

	D1	D2	D3	D4	D5	D6
D1	0	0.1087	0.1477	0.0938	0.205	0.0664
D2	0.332	0	0.0423	0.2456	0.1892	0.0515
D3	0.088	0.0041	0	0.3012	0.1768	0.3046
D4	0.0913	0.0606	0.2938	0	0.2664	0.288
D5	0.2846	0.1029	0.0473	0.2124	0	0.3012
D6	0.1303	0.2149	0.3104	0.2664	0.078	0

Matrix (I – D)

	D1	D2	D3	D4	D5	D6
D1	1	-0.108	-0.1477	-0.0938	-0.205	-0.0664
D2	-0.332	1	-0.0423	-0.2456	-0.1892	-0.0515
D3	-0.088	-0.004	1	-0.3012	-0.1768	-0.3046
D4	-0.0913	-0.060	-0.2938	1	-0.2664	-0.288
D5	-0.2846	-0.1029	-0.0473	-0.2124	1	-0.3012
D6	-0.1303	-0.2149	-0.3104	-0.2664	-0.078	1

Matrix $T = D(I - D)^{-1}$

	D1	D2	D3	D4	D5	D6
D1	0.8929	0.6511	1.0808	1.2131	1.117	1.1742
D2	1.4062	0.7083	1.2674	1.6112	1.3775	1.4462
D3	1.3294	0.8236	1.432	1.8756	1.5033	1.8643
D4	1.4723	0.9443	1.7855	1.8044	1.6999	2.0098
D5	1.5023	0.9111	1.462	1.7901	1.3574	1.8175
D6	1.4709	1.0303	1.7577	1.9732	1.5449	1.7196

Sum of Influences

	$r_i + c_i$	$r_i - c_i$
D1	14.2030	-1.9447
D2	12.8855	2.7482
D3	17.6136	0.0428
D4	19.9839	-0.5514
D5	17.4403	0.2401
D6	19.5283	-0.5350

Pair-wise Comparison Sub-characteristic ISO 9126

	C1	C2	C3	C4	....	C21	C22
C1	1.6667	3	0.5111	7	....	7	7
C2	0.5111	1.6667	0.1217	7	....	0.2254	3
C3	3	8.3333	1.6667	0.5111	....	0.1513	3
C4	0.1513	0.1513	3	1.6667	....	0.2254	0.2254
....	....	....	....	....	....	....	....
C21	0.1513	5	7	5	....	1.6667	0.1217
C22	0.1513	0.5111	0.5111	5	....	8.3333	1.6667

Local Weight D1: Functionality

	C1	C2	C3	C4	C5	Local Weight
C1	1.6667	3	0.5111	7	7	0.3485
C2	0.5111	1.6667	0.1217	7	0.1217	0.0816
C3	3	8.3333	1.6667	0.5111	8.3333	0.3737
C4	0.1513	0.1513	3	1.6667	7	0.1269
C5	0.1513	8.3333	0.1217	0.1513	1.6667	0.0692

Local Weight D2: Reliability

	C6	C7	C8	Local Weight
C6	1.6667	0.1217	0.5111	0.0935
C7	8.3333	1.6667	0.1513	0.2551
C8	3	7	1.6667	0.6514

Local Weight D3: Usability

	C9	C10	C11	C12	Local Weight
C9	1.6667	0.2254	5	3	0.2482
C10	5	1.6667	7	0.1513	0.2777
C11	0.2254	0.1513	1.6667	0.1217	0.0465
C12	0.5111	7	8.3333	1.6667	0.4277

Local Weight D4: Efficiency

	C13	C14	Local Weight
C13	1.6667	0.1217	0.1078
C14	8.3333	1.6667	0.8922

Local Weight D5: Maintainability

	C15	C16	C17	C18	Local Weight
C15	1.6667	0.2254	7	0.2254	0.1523
C16	5	1.6667	8.3333	0.5111	0.4238
C17	0.1513	0.1217	1.6667	3	0.0956
C18	5	3	0.5111	1.6667	0.3283

Local Weight D6: Portability

	C19	C20	C21	C22	Local Weight
C19	1.6667	0.2254	0.5111	0.1513	0.0550
C20	5	1.6667	8.3333	0.5111	0.3251
C21	3	0.1217	1.6667	0.1217	0.0695
C22	7	3	8.3333	1.6667	0.5504

Global Weight

	Sub characteristic	Global Weight
C1	Suitability	0.0740
C2	Accurateness	0.0418
C3	Interoperability	0.0361
C4	Compliance	0.0471
C5	Security	0.0443
C6	Maturity	0.0453
C7	Fault tolerance	0.0758
C8	Recoverability	0.0271
C9	Understandability	0.0497
C10	Learn ability	0.0495
C11	Operability	0.0190
C12	Attractiveness	0.0380
C13	Time Behavior	0.0585
C14	Resource utilization	0.0616
C15	Analyzability	0.0426
C16	Changeability	0.0426
C17	Stability	0.0369
C18	Testability	0.0314
C19	Adaptability	0.0374
C20	Install ability	0.0373
C21	Conformance	0.0424
C22	Replace ability	0.0616

Fig. 6. Output System

Output of this application is the value of local weights for each characteristics ISO 9126, and global weights for each sub-characteristics ISO 9126. Fig.6 shows that sub characteristic Fault Tolerance is the most important sub characteristic and has the most impact in software quality assessment based model of ISO 9126. Sub characteristic Fault Tolerance has the highest weight is 0.0758. Operability is sub characteristic that has the least influence software quality assessment based on model ISO 9126. Sub characteristic Operability has the lowest weight is 0.019.

E. Discussions and Evaluation

Researchers evaluated the system in the same way with the research [9] and [10]. Research [9] and [10] evaluated the system by comparing weights calculated using the proposed method and traditional methods. Traditional method using the

same input values with the proposed method, but ignore the interrelations existence among the quality factors. Comparisons of global weights of each characteristic between our proposed method and the traditional methods can be seen in Fig.7.

Global Weight

	Sub characteristic	The Purposed Method	Traditional Method	Difference
C1	Suitability	0.0740	0.0256	0.0484
C2	Accurateness	0.0418	0.0085	0.0333
C3	Interoperability	0.0361	0.0769	-0.0408
C4	Compliance	0.0471	0.0037	0.0434
C5	Security	0.0443	0.0037	0.0406
C6	Maturity	0.0453	0.1795	-0.1342
C7	Fault tolerance	0.0758	0.1282	-0.0524
C8	Recoverability	0.0271	0.0037	0.0234
C9	Understandability	0.0497	0.0085	0.0412
C10	Learnability	0.0495	0.0051	0.0444
C11	Operability	0.0190	0.0037	0.0153
C12	Attractiveness	0.0380	0.0051	0.0329
C13	Time Behaviour	0.0585	0.0769	-0.0184
C14	Resource utilization	0.0616	0.0085	0.0531
C15	Analyzability	0.0426	0.0051	0.0375
C16	Changeability	0.0426	0.2308	-0.1882
C17	Stability	0.0369	0.1282	-0.0913
C18	Testability	0.0314	0.0085	0.0229
C19	Adaptability	0.0374	0.0769	-0.0395
C20	Installability	0.0373	0.0051	0.0322
C21	Conformance	0.0424	0.0037	0.0387
C22	Replaceability	0.0616	0.0037	0.0579

Fig. 7. Comparisons of global weights of each characteristic between our proposed method and the traditional method

Then, researchers show global difference weights of each characteristic between our proposed method and the traditional method in the graph (fig.8). Integration DEMATEL and ANP methods produce a more normal weight values, and proportionate. This is due to the interrelationship among the quality factors of ISO 9126 based models. Researchers have tested this application, and the results are appropriate with research [3] and [4].

IV. CONCLUSION

Based on the results, it can concluded that the weight of characteristics software quality model ISO 9126 can calculated by using the integration DEMATEL and ANP methods. The result of this calculation is the value of local weights for each characteristics ISO 9126, and global weight for each sub characteristics ISO 9126. This integration methods can used to represent interrelationship among the quality factors based model ISO 9126.

For the next study, subsequent research can combined fuzzy DEMATEL, and fuzzy ANP [8] methods to represent the important of characteristics and level of sub characteristic model ISO 9126. This study only discusses the weights of characteristics model ISO 9126, subsequent research can be done by adding the methods used to measure the quality of software e.g. integration method DEMATEL, ANP, and TOPSIS.

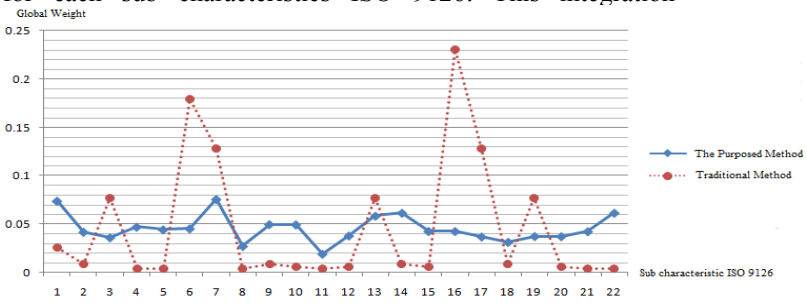


Fig. 8. Difference global weights of each characteristic in graph

REFERENCES

[1] Behkamal, B.; Kahani, M.; Akbari, M.;, "Customizing ISO 9126 quality model for evaluation of B2B applications." *Elsevier Journal Information and Software Technology*, 2009.

[2] Zulzalil, H.; Ghani, A.; Selamat, M.; Mahmod, R.;, "A Case Study to Identify QUALITY Attribute Relationships for Web Based Applications." *Int. J. Comput. Sci. Network Security*, vol. 8, pp. 215-220, 2008.

[3] Chaghooshi, A.; Fathi, M.; Faghih, A; Zarchi, M.;, "Applying a New Integration of MCDM Techniques for Supplier Selection (Case Study: Pars Tire Company)." *Australian Journal of Basic and Applied Sciences*, vol. 6(2), pp. 9-19, 2012.

[4] Chen, J. K.; I, S. C.;, "Using a novel conjunctive MCDM approach based on DEMATEL, fuzzy ANP, and TOPSIS as an innovation support system for Taiwanese higher education", *Expert Systems with Applications*, vol. 37, pp. 1981-1990, 2010.

[5] Tzeng, G.H., C.H. Chiang and C.W. Li;, "Evaluating intertwined effects in elearning programs: A novel hybrid MCDM model based on factor analysis and DEMATEL", *Expert Systems with Applications*, 32: 1028-1044, 2007.

[6] Ong, C.; Huang, J.; Tzeng, J;, "Multidimensional data in multidimensional scaling using the analytic network process", *Pattern Recognition Letters*, vol. 26(6), pp. 755-767, 2004.

[7] Liou, J.; Tzeng, G.; H.C;, "Airline safety measurement using a hybrid model", *Journal of Air Transport Management*, vol. 13, pp. 243-249, 2007.

[8] Yazgan, H. R.;, "Selection of Dispatching Rules with Fuzzy ANP Approach," *The International Journal of Advanced Manufacturing Technology*, vol. 52, pp. 651-667, 2010.

[9] Yang, J. L.; Tzeng, G. H.;, "An integrated MCDM technique combined with DEMATEL for a novel cluster-weighted with ANP method," *Expert Systems with Applications*, 38: 1417-1424, 2011.

[10] Yang, Y. P. O.; Shieh, H. M.; Leu, J. D.; Tzreng, G. H.;, "A Novel Hybrid MCDM Model Combined with DEMATEL and ANP with Applications," *International Journal of Operations Research*, vol. 5, no. 3, pp. 160-168, 2008.

Likelihood Calculation Classification for Indonesian Language News Documents

Aini Rachmania[1], Jafreezal Jaafar[2],Norshuhani Zamin[3]
Department of Computer & Information Sciences
Universiti Teknologi PETRONAS
Tronoh, MALAYSIA
[1]aini.rachmania@gmail.com, [2]jafreez@petronas.com.my, [3]norshuhani@petronas.com.my

***Abstract*-Text categorization has been an important research area that seeks to classify textual documents into a group of predetermined categories. Unfortunately, the interest towards Indonesian news classification has been very little. In this paper, we propose a text categorization algorithm based on Bracewell method that uses the likelihood calculation between the article and the category's keywords. Through experiments, the algorithm succeeded in classifying Indonesian news corpus with accuracy as high as 93,84% in offline environment, 93,82% in online environment, and 80% benchmarking against human evaluation.**

***Keywords*—text categorization, Indonesian documents, information retrieval, news domain, likelihood calculation**

I. INTRODUCTION

Text classification copes with the task of assigning a text record from a collection of documents with a predetermined label denoting its content in general [1]. In the past couple of decades, it has been one of the most vibrant research field in the area of information retrieval. It has progressed massively due to the immense flow of data, especially ones available in the internet [2], and has become an important approach in managing and organizing the prodigious dimension of data [3, 4]. This phenomenon has created a high demand for automated textual document classification.

The implementation of text classification may dealt with various types of corpora as testing material. These corpora can be in the form of text dialogues, scientific writings, and any textual data existing online [5]. Web pages are among the most readily available source for data collections. These data include the online news articles. They are free, accessable and easy to retrieve. However, the large data volume is synonymous with vast variety in terms of data attributes such as article length [6] and the writing style. Classification on the online news domain, has been a more challenging situation as compared to those on general essays. News in one label tend to have sporadic theme, resulting in subtle visible similarity between articles [7]. Moreover, with the help of the internet technology and internet connection, online news articles can be updated in sync with the real world event. Consequently, this requires the text classifier to be easily updated and be efficient.

The focus on text classification has always been leaning more towards English and European language corpora classifitaion [8, 9]. Approaches for Asian language text documents are relatively more limited, as compared to the number of methods and research using English documents. Asian text classification has gained prominent importance [10], considering the large number of speakers. In the Indonesian language context, problems emerged are distinct from those in the Japanese and Chinese language. The difference in alphabet letters may lead to a substantial difference in the preprocessing step.

From a series of extensive study, we have found limited number of methods that deal with Indonesian news. Several works have implemented a rather common method such as Naive Bayes [11, 12] and Support Vector Machine [13]. As such, in this paper we propose an adaptation of the Bracewell's method [7] where the classifier was originally meant to categorize English and Japanese news corpora. The aim of this research is to classify Indonesian news documents using Bracewell's algorithm and to reduce the large data characteristic challenges mentioned above. We have adopted the steps in the category classification algorithm, detached the topic classification part, and employed a simpler training process.

II. REVIEW OF THE RELATED METHODS

In this section, we review several Indonesian text classification techniques in this area. One approach using the k-Nearest Neighbor (kNN) method was developed in 2012 by Arifin*et al* [14]. Even though the end product is emotion classification, the methodology bears similar overall structure with the algorithm described in this paper. It first preprocesses the training data, stems and then weighs the data collection. Once the data is weighed, the kNN phase begins. This algorithm classifies an object by first selecting *k* other nearest objects ("neighbors") surrounding it, and assign the object to the most frequent neighbor. The fundamental problem with kNN is the dependency it has towards the majority of members ofthe class, hence unsuitable for collection with highly diversed data. Another issue is the difficulty in determining where the exact number for *k* should start, given that there has never been a certain rule for it except from experiment. As seen in [14], the *k* used in the tests may varied from 2 to 60, while the optimum *k* yielded is 40. An improvement of the method was soon introduced by [15]. A combination with k-means clustering algorithm was aimed on reducing the high calculation complexity, and helping the classifier to be less dependent on the size of the training data. This is done by clustering the training documents into *k* number of clusters,

while each cluster is represented by a chosen centroid. Thus, instead of calculating similarity between the complete set of training data with the sample, the classification now only deals with respective centroids, consequently producing more efficient computation. The shortcoming of this method lies in the nature of most clustering methods, where the best starting number for clusterhas always been vague. The algorithm yielded 85% F-measure.

In 2009, Asy'arie *et al* [12] conducted a study on news classifier using the Naive Bayes classification method. In this paper, the method employed identical initial phase with [14]. For each document, terms are preprocessed, stemmed, and weighed. Classification is done by calculating the probability between corpus collection and the label. The label assigned to the document is category with maximum posteriori. The classifier managed to reach more than 90% F-measure average with recall rate up to 93,75%. The weak point in this effort is that the method hardly processes news with more than 1,000 words length, and simply produce null as its value. Asy'arie mentioned this [14] but the problem is that news articles can be 2 to 6 paragraphs long, thus providing difficulty to classify news corpus. Another implementation of a Bayesian classifier was recently studied early this year [11]. The unique approach on Naive Bayes method in this paper was its usage to classify personality in a social website by studying the personality paragraph written by the user in the page. Of course with user-defined materials, it is even more difficult to uniform the collection due to the style of writing that may differ significantly among users. Although the method satisfied a 92,5% accuracy, it relies excessively on the number of training documents it studies with [11].

One of the most commonly used technique in text classification is Support Vector Machine (SVM). On 2011, [13] utilized this algorithm on Indonesian news collections and achieved an average accuracy of 85%. The SVM classifier uses an Optimum Separating Hyperplane which is set to separate objects into their fitting class. If an object is located on the separator line, it becomes the support vector. Sharing similar issue with [14], the parameter *gamma SVM* in this method is defined by user without specific criteria.

In this paper, we propose the use of Bracewell's method to classify Indonesian news documents. Although the method was not tested on Indonesian corpora, the high precision and recall values on both English and Japanese domains justify the versatility that this algorithm has in dealing with classification for different languages. Apart from that, because the calculation technique does not require complex elements nor user dependent variable, we assumed this method is suitable for online news corpus.

III. PROPOSED METHOD

This paper adopted Bracewell's Algorithm for category classification without the use of its topic discovery component [7]. The proposed method is carried out in two main phases: training phase and classification phase. Figure 1demonstrates the workflow of Bracewell's Algorithm for category classification.

Before training, the news documents are collected from the news web site. They are digitized and stored in the same format of corpus. We call this step data storing. Data storing step takes important articles and store them to the set of predefined category. In this study, we use a compilation from several news categories from several online newspaper companies :*www.kompas.com*, *www.antaranews.com* and *www.tempointeraktif.com* in order to get the optimum set. Once the data storing step is done, the training phase begins.

When the training phase is complete, the classifier is ready to classify the testing samples. The classification algorithm proposed in this paper is relatively simple, since it does not need a lot of components to be calculated. It compares the likelihood between the keywords of the testing sample with the keywords of the documents [7].

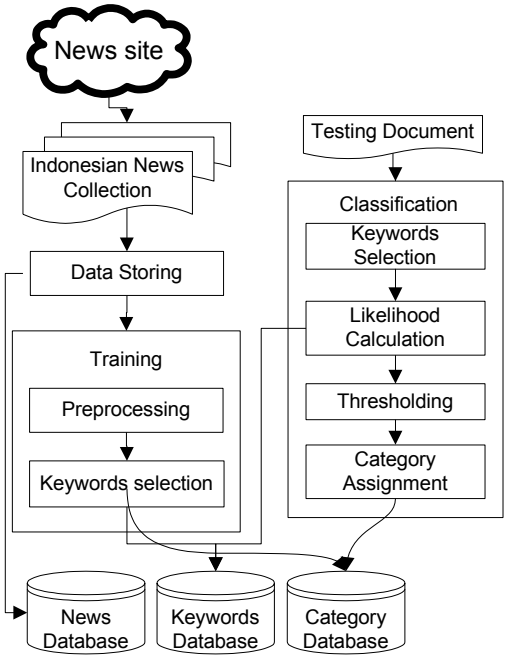


Figure 1. Workflow Process of the Proposed Method

A. *Training Phase*

Training phase is first done by going to each predetermined category, and uniform its member articles. For each category, the articles supporting it are preprocessed. The preprocessing step yields a set of weighted words. This set is considered as the representatives of each articles, and then used in the keywords selection process.

a. Preprocessing

Preprocessing process is important in text processing as well as text classification, because it helps to sum up an article effectively by eliminating unnecessary words and conforming each word to its root form, helping the classifier to understand the article more efficiently. The outline of preprocessing step can be seen in Figure 2 .

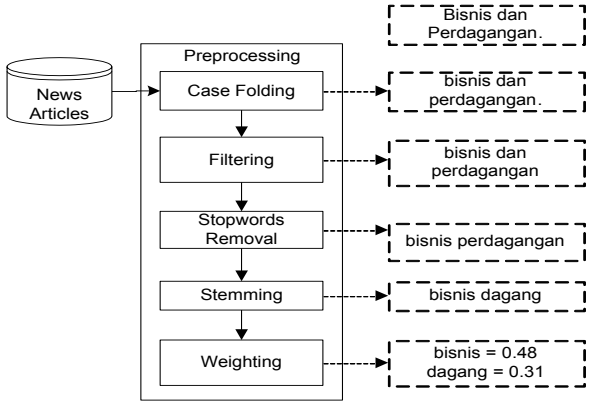


Figure 2. Preprocessing Process

Shown in Figure 2, the Preprocessing process commonly covers 5 steps: case folding, filtering, stopwords removal and weighting. The example described is the phrase "*Bisnis dan Perdagangan.*" (Business and Trade). On the first level, the phrases are both transformed into small case letters, producing "*bisnis dan perdagangan.*". This transformation is done during the case folding step. Next, the folded phrase is filtered in the filtering step, producing just "*bisnis dan perdagangan*". Filtering is an attempt to eliminate characters that are non-alphabetic, such as symbols, hyphens, and numbers. Once it is filtered, the phrases are then stemmed. Stemming is required to turn the phrases back into its root form. Since only "*perdagangan*" that has affixes, the word is stemmed into the term "*dagang*", resulting in the phrase "*bisnis dan dagang"*. In this paper, we apply the Enhanced Confix Stripping Stemmer approach to stem the articles [16]. After that, the last step is weighing. We use TF-IDF [17-19] weighing scheme because not only that it is common among NLP researchers, but it is also efficient, and has been used many times in Indonesian news classification studies [12-14, 20].

The TF-IDF weighting system is a development from IDF where it recognizes the importance of a term not only in a document, but also in the collection of documents [21]. A term is deemed as significant if it frequently appears in a document (IDF), but rarely seen in other documents in the whole collection. The calculation for a term's TF-IDF weight is explained in Equation 1.

$$w_{ij} = tf_{ij}.\log_2\left(\frac{N}{df_i}\right), \quad (1)$$

w_{ij} is the weight of term i in document j, tf_{ij} is term i occurences frequency in document j, N is the total number of documents, and d_{fi} is term i occurences frequency in documents in the collection.

b. Keywords Selection

The original keywords selection method for this algorithm is based on Bracewell's Multilingual Single Document Keywords Extraction method which was proposed on 2005 [22]. Despite proven effective when tested on English and Japanese corpora, the method requires the use of taggers and morphological analyzer to work with those specific languages. As working with Indonesian language is relatively easier to simplify since it does not use different alphabets, we select keywords using the top-*n* method [12, 20].

The method works by going through each category in the database and each article in the category. For example, we have article *A*. This article has undergone the preprocessing step, thus possesses a set of weighed terms. These terms are then ranked descendingly according to their weight. We then select top *n* number of terms, and keep these terms as the keywords of the category in the database along with the frequency of the term inside the document. We then proceed to article *B* and repeat the process.

The number of *n* may varied between 0 to the maximum number of terms of an article. With respect to the length of the articles, we set the *n* be: 5, 10, 15 and 20. These numbers are tailored to fit news articles which mostly range from only 2 to 6 paragraphs long. To investigate the effect of these numbers on the classifier performance, we have defined the difference among *n*s to be 5 so that each *n* would have significant difference between each other. We assume that with significant difference, the result produced from each *n* would be substantially different in that respect.

B. Classification Stage

The main idea of Bracewell's Category Classification algorithm is the computation of likelihood between article that is going to be classified, with the category [7]. When the likelihood between the articles and all the categories has been computed, the classifier then creates a dynamic threshold. This threshold is an effort to decide which category to be assigned to the article. When this step is complete, the categories are assigned to the article.

Since the likelihood calculation is done between the sample article's keywords and the category's keywords, it is necessary for the sample to first undergo the keywords selection process. The selection process applied in this step is similar to the one used in the training process. The article is first preprocessed, and then *n* number of keywords are selected.

When the keywords are retrieved, the likelihood calculation takes place. The likelihood between the sample article and the categories are calculated using Equation 2.

$$\text{Likelihood}\ (c_j|\ A=\{k_1, k_2, \dots, k_n\}) = -\sum_{i=1}^{n} P(k_i|c_j)\log\left(P(k_i|c_j)\right) \quad (2)$$

$P(k_i|c_j)$ is the probability of keyword j in category c_j, while A is the sample article represented by a group of keywords $\{k_1, k_2, \dots, k_n\}$. If a classifier has j category, then there will be likelihood calculation between A and c_1, until A and c_j. To decide which category should be assigned to the article, a threshold is calculated using Equation 3.

$$\text{Threshold} = \frac{\sum_1^{|L|} l_i}{|L|} + \sqrt{\frac{\sum (l_i - \frac{\sum_1^{|L|} l_i}{|L|})^2}{|L|}} \quad (3)$$

The threshold is a combination between the average of likelihood L_i and 1 standard deviation of all the likelihood of the categories. When a category is greater than the

threshold, it is assigned to the article. This step ensures that the category chosen truly represents the article.

IV. EXPERIMENT SETUP

We have tested the classifier on two types of experimentations. On the first type, the accuracy is a comparison between the output of the classifier with the ground truth from *www.kompas.com*. On the second type, we evaluated the classifier using human annotators, and compare the result. In the implementation phase, we have used Java programming language and run on Microsoft Windows Vista Business Edition operating system.

a. Datasets

The dataset that we used in the experiment is retrieved from KOMPAS website. We have downloaded and digitized a set of training and testing data from different categories, ranging from the year 2011 until 2012. Table I explains the details. To start with the training data, we have used 100 articles for each category. We assume that the number is sufficient because in one day, the KOMPAS website posts around 50 news in one category. However, instead of downloading the news on date or time basis, we download them rather randomly. This is because we want the classifier to learn the news' characters, hence wider vocabularies.

TABLE I. DATASETS

Training	Testing	Terms	Stopwords	Dictionary
900	455	12,912	795	29,337

There are 12,912 stemmed terms, 795 stopwords, and 29,337 root words dictionary. The stopwords are the list of words used during stopwords removal function in the preprocessing phase, while the stemmed terms are the result of the stemming step, and dictionary comprises the list of root words where the stemmer look up for matching words in the article.

The training data was stored as support documents for the predetermined categories. Through observations, we found 9 important categories. These categories are the source of classification during the training stage, as well as the ground truth for the first set of experiment. The list of the categories are shown in Table II.

TABLE II. CATEGORIES LIST

Category	Training Documents	Testing Documents
Nasional (*National*)	100	50
Regional (*Regional*)	100	51
Internasional (*International*)	100	51
Metropolitan*(Metropolitan*)	100	50
BisnisdanEkonomi (*Business and Economy*)	100	51
Olahraga (*Sports*)	100	51
Sainsdan Teknologi (*Science and Technology*)	100	51
Edukasi (*Education*)	100	50
Pariwisata (*Tourism*)	100	50
Total	**900**	**455**

The testing data that we used for the human evaluation setup is different from the one for the system evaluation. We have used 45 data from 2012 and asked the annotators to classify the data based on their judgment. The dataset for this setup is described in Table III.

TABLE III. DATASET FOR HUMAN EVALUATION

Categories	Number of Articles
Bisnis dan Ekonomi (*Business and Economics)*	5
Edukasi (*Education*)	5
Internasional (*International*)	5
Metropolitan (*Metropolitan*)	5
Nasional (*National*)	5
Olahraga (*Sports*)	5
Regional (*Regional*)	5
Pariwisata (*Tourism*)	5
Sains dan Teknologi (*Science and Technology*)	5
Total	**45**

b. Evaluation

We have tested the classifier on two environments: offline and online. The purpose of these setups is to see whether or not the classifier work in online classification, without having to always update the training process everytime it classifies. To evaluate the classifier, we use *Accuracy*. The elements used to calculate accuracy is shown in Table IV.

TABLE IV. ACCURACY MATRIX

	Relevant	Not Relevant
Retrieved	TP	FP
Not Retrieved	FN	TN

TP stands for True Positive, which represents documents that are relevant and indeed retrieved by the classifier. *TN* is True Negative, representing documents that are not relevant and rejected by the classifier. *FN* and *FP* both represent the misclassification, where *FP* denotes the documents that are not relevant yet retrieved, and *FN* are documents that are relevant but not classified by the classifier. From those components, we then calculate the accuracy using Equation 4.

$$Accuracy\ (A) = \frac{(TP + TN)}{(TP + FP + FN + TN)} \tag{4}$$

c. Results

Each experiment is conducted with a small range of retrieved keywords: 5, 10, 15 and 20 keywords. From a series of experiment, we found that these numbers of selected keywords improve the classifier most optimally. Selecting less than 5 keywords would yield a poorly described document, whilst more than 20 keywords would result in more noise. The number and interval of top keywords chosen is in a small range because they are only used to test whether the algorithm can classify documents and whether the classifier also improves as the keywords

selected get bigger. The result of each classification using each number of keywords is then compared.

Table V shows the result of the proposed method on online environment, tested on each 5, 10, 15 and 20 top-*n* selected keywords. The offline scenario testing is conducted to test the classifier using corpus available in the system that we have created. These corpus are already downloaded and digitized into corpus format in the similar manner used in the data storing process before the training phase.

TABLE V. CLASSIFICATION RESULT IN OFFLINE ENVIRONMENT

Category	**Accuracy**			
n	**5**	**10**	**15**	**20**
Bisnis & Ekonomi	92,59%	93,13%	95,38%	96,15%
Edukasi	94,12%	93,13%	94,62%	96,15%
Internasional	92,19%	94,66%	93,85%	94,62%
Metropolitan	88,89%	89,31%	90,00%	90,91%
Nasional	**98,52%**	98,47%	98,46%	99,24%
Olahraga	**98,52%**	98,52%	**100,00%**	**100,00%**
Pariwisata	84,44%	**99,29%**	86,36%	88,89%
Regional	77,78%	82,58%	80,77%	80,77%
Sains & Teknologi	94,81%	80,92%	96,92%	97,69%
Average	**91,32%**	**92,22%**	**92,93%**	**93,82%**

The result shows that the classifier can perform well on offline manner, with accuracy rates of 91,32%, 92,22%, 92,93% and 93,82% for each keywords. Albeit the high accuracy, the result also shows that the classifier performs differently as it process one category to another. Table V shows that the classifier can produce a perfect accuracy on some categories, yet relatively lower on other. This could be the result of the homogeneity of keywords which reside in a category. For example, category *Olahraga* (Sports) might have keywords which are very specific and self-explanatory such as racquet, balls, tennis and so forth, that it will not appear in any other category. On the contrary, the category *Regional*, only comprises news about events in a province or a state. Thus lacking in distinctive keywords.

The next experiment is conducted on online environment. The difference between this scenario and the offline scenario is the way the corpus is prepared. On online environment, we test the classifier tool which we have created using corpus that is downloaded directly from the website as we conduct the experiment.We have created a customized grabber and parser for KOMPAS website and used them to download the corpus during testing. The user only needs to put the news article link in the system, and the system directly downloads and categorizes the article.

For experiments on online environment, the classifier produced a 93,84% accuracy as its best result. We see that the classifier performs better on online environment, for a few possible reason. The first one is the use of grabber. The grabber that we have designed let the classifier download the news straight away from the web, thus making the data more accurate. Second, the small yet determining factor such as the computer's memory, human error during digitizing of the data, and so forth. However, we think that the difference is minor enough to be ignored.

TABLE VI. CLASSIFICATION RESULT IN ONLINE ENVIRONMENT

Category	**Accuracy**			
n	**5**	**10**	**15**	**20**
Bisnis & Ekonomi	**100,00%**	**98,58%**	99,29%	**100,00%**
Edukasi	93,66%	94,33%	95,77%	96,48%
Internasional	92,25%	92,96%	92,20%	94,33%
Metropolitan	90,85%	91,49%	92,20%	92,91%
Nasional	96,45%	96,43%	98,46%	97,87%
Olahraga	98,52%	**98,58%**	**100,00%**	**100,00%**
Pariwisata	84,51%	83,80%	86,36%	85,00%
Regional	**81,69%**	**81,56%**	**81,56%**	**81,56%**
Sains & Teknologi	95,77%	95,74%	94,33%	96,43%
Average	**92,63%**	**92,61%**	**93,35%**	**93,84%**

The second set of experiment using the human evaluation as our ground truth produced 80% matches between the system output and the evaluation. The algorithm performed in a similar manner with the first set of experiment, producing lower result on *Pariwisata* and *Regional* with only 2 document matches out of 5 each. Results are shown in Figure 3.

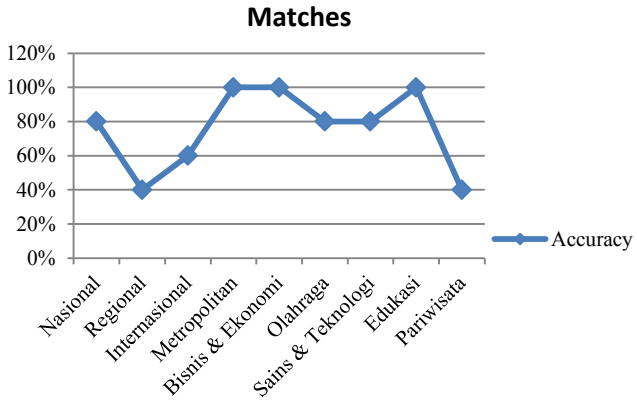


Figure 3. Accuracy Compared to Human Evaluation

The number of samples used to calculate the matches can be one of the factor why the classifier yielded results not as high as the first set of experiment. For each of the category, we only use 5 group of data, thus getting 3 out of 5 data correct will only result in 60% matches. However, with all the results compiled, the classifier is proven to be capable of clasifying Indonesian news articles with high accuracy.

V. CONCLUSION

This paper proposes method for Indonesian news classification using likelihood calculation. The method was first introduced by Bracewell *et al*, and was proven to be capable of classifying news articles. The algorithm classifies the news in two stages: training and classification. In the training phase, we have applied different stemmer and different keywords selection mechanism in our experiments. In the classification, the likelihood calculation is applied.

We have conducted two sets of experiments to test the method. The first experiment is based on the ground truth provided by the news websites. We compare the category yielded by the classifier with the category from the web. The result shows the classifier could reach up to 93,84% accuracy in online environment, and 93,82% in offline environment. On the second set of experiment, the method produced 80% matches. This proves that the algorithm is able to classify with good accuracy.

In the future work, we would like to integrate this method with a clustering method in order to further organize our data and to make the computation more efficient. Apart from that, we feel that a feature extraction method is an interesting area to be incorporated with the proposed method.

VI. REFERENCES

[1] F. Sebastiani, "Machine learning in automated text categorization," *ACM computing surveys (CSUR),* vol. 34, pp. 1-47, 2002.

[2] D. Billsus and M. J. Pazzani, "A hybrid user model for news story classification," *COURSES AND LECTURES-INTERNATIONAL CENTRE FOR MECHANICAL SCIENCES,* pp. 99-108, 1999.

[3] L. Chen*, et al.*, "Class-dependent projection based method for text categorization," *Pattern Recognition Letters,* vol. 32, pp. 1493-1501, 2011.

[4] O.-W. Kwon and J.-H. Lee, "Text categorization based on< i> k</i>-nearest neighbor approach for Web site classification," *Information processing & management,* vol. 39, pp. 25-44, 2003.

[5] A. Kilgarriff and G. Grefenstette, "Introduction to the special issue on the web as corpus," *Computational linguistics,* vol. 29, pp. 333-347, 2003.

[6] A. Selamat*, et al.*, "Web news classification using neural networks based on PCA," in *SICE 2002. Proceedings of the 41st SICE Annual Conference*, 2002, pp. 2389-2394.

[7] D. Bracewell*, et al.*, "Category Classification and Topic Discovery of Japanese and English News Articles," *Electronic Notes in Theoretical Computer Science,* pp. 51-65, 2009.

[8] H. Sawaf*, et al.*, "Statistical classification methods for Arabic news articles," *Natural Language Processing in ACL2001, Toulouse, France,* 2001.

[9] H. M. Noaman*, et al.*, "Naive Bayes Classifier based Arabic document categorization," in *Informatics and Systems (INFOS), 2010 The 7th International Conference on*, 2010, pp. 1-5.

[10] F. Peng*, et al.*, "Text classification in Asian languages without word segmentation," in *Proceedings of the sixth international workshop on Information retrieval with Asian languages-Volume 11*, 2003, pp. 41-48.

[11] N. M. A. Lestari*, et al.*, "Personality Types Classification for Indonesian Text in Partners Searching Website Using Naïve Bayes Methods," *IJCSI International Journal of Computer Science Issues,* vol. 10, pp. 1-8, 2013.

[12] A. D. Asy'arie and A. W. Pribadi, "Automatic news articles classification in indonesian language by using naive bayes classifier method," in *Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services*, 2009, pp. 658-662.

[13] D. Y. Liliana*, et al.*, "Indonesian News Classification using Support Vector Machine," *World Academy of Science, Engineering and Technology,* vol. 57, pp. 767-770, 2011.

[14] A. Firdan and K. E. Purnama, "Classification of Emotions in Indonesian Texts using K-NN Method," *International Journal of Information and Electronics Engineering,* vol. 2, pp. 899-903, 2012.

[15] P. W. Buana*, et al.*, "Combination of K-Nearest Neighbor and K-Means based on Term Re-weighting for Classify Indonesian News," *International Journal of Computer Applications (0975 – 8887),* vol. Volume 50, pp. 37-42, 2012.

[16] A. Z. Arifin*, et al.*, "Enhanced Confix Stripping Stemmer and Ants Algorithm For Classifying News Documents in Indonesian Language.," in *The 5th International Conference on Information & Communication Technology & Systems.*, Surabaya, 2008, pp. 149-158.

[17] Y. H. Li and A. K. Jain, "Classification of text documents," *The Computer Journal,* vol. 41, pp. 537-546, 1998.

[18] M. Skiba, "Text preprocessing in programmable logic," 2010.

[19] G. Salton and C. Buckley, *Term-Weighting Approaches in Automatic Text Retrieval.* : Cornell: Cornell University., 1988.

[20] A. Z. Arifin and A. N. Setiono, "Klasifikasi Dokumen Berita Kejadian Berbahasa Indonesia dengan Algoritma Single Pass Clustering," in *Proceeding of Seminar on Intelligent Technology and Its Applications (SITIA)*, Surabaya, 2002.

[21] W. Zhang*, et al.*, "A comparative study of TF* IDF, LSI and multi-words for text classification," *Expert Systems with Applications,* vol. 38, pp. 2758-2765, 2011.

[22] D. B. Bracewell*, et al.*, "Multilingual Single Document Keyword Extraction for Information Retrieval," in *NLP-KE*, 2005, pp. 517-522.

Lip Segmentation and Tracking Based on Chan-Vese Model

Aris Nasuha[1,2]
[1]Electronics Engineering Education Department
Universitas Negeri Yogyakarta
Yogyakarta, Indonesia
aris12@mhs.ee.its.ac.id

Tri Arief Sardjono[2], Mauridhi Hery Purnomo[2]
[2]Electrical Engineering Department
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia

***Abstract* — Lip reading has wide spread application, e.g. audio-visual Automatic Speech Recognition (AV-ASR), silent speech interface and person identification. Lip segmentation is one of important step in lip reading, because it provides basic information to be processed in subsequent steps. Lip tracking is a process of locating lip to associate lip in consecutive video frames. Chan-Vese model is a region-based segmentation algorithm, which also can be used as tracking method. This algorithm can detect boundary of object which not defined by gradient, while classical active contour can't be applied. This method also can detect object by any initial curve in the image, not necessarily surround the object. This paper investigates about lip segmentation and tracking based on Chan-Vese model, preceded by the color segmentation.**

Keywords—lip reading, segmentation, tracking, Chan-Vese

I. INTRODUCTION

Recently, lip reading becomes more popular research, because it has wide spread applications. The automatic speech recognition (ASR) system is widely used, for example in robotics, personal computer, and cellphone. But, in the noisy circumstances, the ASR performance will drop. By lip reading system, the problem will be overcome. Lip reading system also can be used as one of silent speech interface for laryngectomee (person who has no larynx) [1]. In specific security system, lip reading is used for person identification [2].

Lip reading system consist several steps, and lip segmentation is one of important step, because it provides basic information to be processed in next step. Several lip segmentation methods based on gray-scale image [3-5], other methods are based on color image. Some methods use color image directly [6, 7], the other method use color conversion to other color space to enlarge the color different of lip and the background, for example YCbCr [8], NTSC [9], and CIE-L*a*b* [10-12].

Another important step in lip reading is lip tracking. In this step, the key points of lip are tracked for each frame. There are several lip tracking methods, for example pattern matching snake [13] and watershed [4]. Chan-Vese model is a region-based segmentation algorithm, which can be used as segmentation method and also for tracking. The advantage of this algorithm can detect object in image with contours without gradient or weak gradient, which often presence in face image. In the case of image contour without gradient, classical active contour like Snake, can't be applied. This paper investigates about lip segmentation and tracking based on Chan-Vese model.

II. COLOR TRANSFORMATION AND IMAGE SEGMENTATION

A. CIE-L*a*b* Color Transformation

Although most of the image in our database have been stored in RGB color space format, we choose the color space CIE-L*a*b*. Because of the distance between any two points in color space is proportional to the perceived color difference, so we required a uniform space color. CIE-L*a*b* is one of the color space that satisfy uniform space color. Details of the transformation from RGB to CIE-L*a*b* are:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.813 & 0.011 \\ 0.00 & 0.010 & 0.990 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (1)$$

$$L^* = \begin{cases} 116(Y')^{\frac{1}{3}} - 16 & \text{if } Y' > 0.008856 \\ 903.3Y' & \text{otherwise} \end{cases} \quad (2)$$

$$a^* = 500\left(K_1^{\frac{1}{3}} - K_2^{\frac{1}{3}}\right) \quad (3)$$

$$b^* = 200\left(K_2^{\frac{1}{3}} - K_3^{\frac{1}{3}}\right) \quad (4)$$

$$K_i = \begin{cases} \Phi_i & \text{if } \Phi_i > 0.008856 \\ 7.787\Phi_i + \frac{16}{116} & \text{otherwise} \end{cases} \quad (5)$$

For i = 1, 2, 3

$$\Phi_1 = X' = \frac{X}{X_0}, \Phi_2 = Y' = \frac{Y}{Y_0}, \Phi_3 = Z' = \frac{Z}{Z_0} \quad (6)$$

where X_0, Y_0 and Z_0 are the values of X, Y, Z, for the reference white, respectively. The reference white in the transformation is defined as {R = G = B = 255}

B. Chan-Vese Algorithm Overview

Chan-Vese model is a segmentation method based on segmentation method by Mumford-Shah and level set method by Osher and Sethian. The detail of this model can be read in [14]. This algorithm is not based on edge function to stop evolution curve at desired edge. Chan-Vese algorithm can detect boundary of object which not defined by gradient, while classical active contour can't be applied. This method also can detect object by any initial curve in the image, not necessarily surround the object.

The objective of Chan-Vese model is to partition an input scalar image u_0: $\Omega \rightarrow$ R defined over a d-dimensional image domain $\Omega \subset R^d$ into two possibly disconnected regions Ω_1 (foreground) and Ω_2 (background) of low intra-region variance and separated by a smooth closed contour C ($\Omega = \Omega_1 \cup \Omega_2 \cup C$). Chan-Vese algorithm can be modeled in equation:

$$E_{CV}(C, c_1, c_2) = \mu|C| + \lambda_1 \int_{\Omega_1} (u_0(X) - c_1)^2 \, dx + \lambda_2 \int_{\Omega_2} (u_0(X) - c_2)^2 \, dx \quad (7)$$

where c_1 and c_2 represent the average intensity level inside Ω_1 and Ω_2, respectively, and μ, λ_1, and λ_2 are user-defined parameters. The optimal segmentation (C, c_1, c_2) corresponds to a global minimum of (7).

Chan-Vese model has been extended by using a multi-phase level set framework scheme, for piecewise constant (PC) [15] and piecewise smooth (PS) [16] optimal approximations. These models effectively solve the boundary blur or digital object segmentation and can detect interior contours in the image. However, their main drawback is the increased complexity of computing.

A lot of segmentation problems have been solved by Chan-Vese algorithm, for example in medical image segmentation: heart [17], CT bone [18], tooth [19], brain MRI [20]. This algorithm is also applied for segmenting nighttime vehicle license characters [21], and aerial photographs [22].

Chan-Vese algorithm can be applied in object tracking [23] and fluorescent cell tracking [24]. This algorithm can be described as sequential segmentation, where the final contour from one image frame is used as the initial contour for the next. The main strength of the algorithm is its ability to handle the topology changes that result from deformations in the object being tracked.

III. Implementation

In order to test our proposed method, we use image from FEI Face Database [25]. All images are colorful and taken against a white homogenous background. In this database, 20 subjects in frontal face image are randomly chosen. Each subject has two expressions, one with a neutral or non-smiling expression and the other with a smiling facial expression. Original image size is 360x260 pixels, then manually cropped to get only mouth region. The final image size is 56x114 pixels.

A. Lip Segmentation

Color image clustering is used in lip segmentation. There are many clustering algorithms, and k-means are often used in color image segmentation. To get a good clustering result, inputs of clustering should be chosen precisely. In order to enlarge the color different of lip and the background, input images are converted to CIE-L*a*b* color space. This color space consists of a luminosity layer L*, chromaticity-layer a* indicating where color falls along the red-green axis, and chromaticity-layer b* indicating where the color falls along the blue-yellow axis. Since the color information exists in the a* b* space, the clustering inputs are pixels with a* and b* values.

As a comparison, other color space, i.e. using Cb and Cr component of YCbCr color space, and also I and Q component of NTSC, is applied. The result of lip segmentation is used as input for Chan-Vese algorithm. The block diagram of implementation is sketched in Fig.1.

B. Lip Tracking

To simulate lip tracking, the first frame is taken from the neutral or non-smiling expression image, the second from the smiling expression and the third from the non-smiling expression. The result of Chan-Vese segmentation from first frame is used as initial curve for the second, and so forth. The pseudo-code of simple lip tracking algorithm using Chan-Vese model is sketched in Fig.2.

IV. Experimental Result

Original Chan-Vese algorithm is an image segmentation method which input is gray-scale image. Therefore, it can be used directly in lip image, with no color segmentation before. Input image is only converted to gray-scale image. But, the result provides that lip segmentation is failed. Examples of these images are seen in Fig.3.

Our proposed method is implemented in lip image of 20 subjects. To measure the accuracy of the segmentation, our proposed method is compared with manual segmentation. Of comparison, the accuracy is calculated by following equation:

$$\text{accuracy} = \frac{TP+TN}{TP+FP+FN+TN} = 1 - \frac{FP+FN}{TP+FP+FN+TN} \quad (8)$$

where TP is number of true positive pixel, TN is number of true negative pixel, FN is number of false negative pixel, and FP is number of false positive pixel.

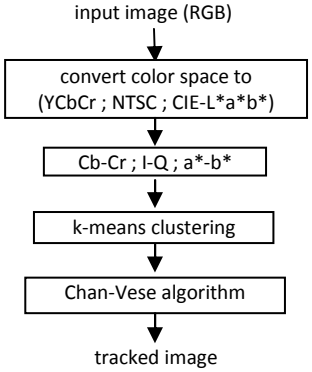


Fig.1. Block diagram of lip segmentation and tracking algorithm

```
contour 0 ← initial contour
For j = 1 to n_frames
{
   contour j ← Chan-Vese (contour j -1, image j )
   draw contour of frame-j
}
```

Fig.2. Pseudo-code of simple lip tracking

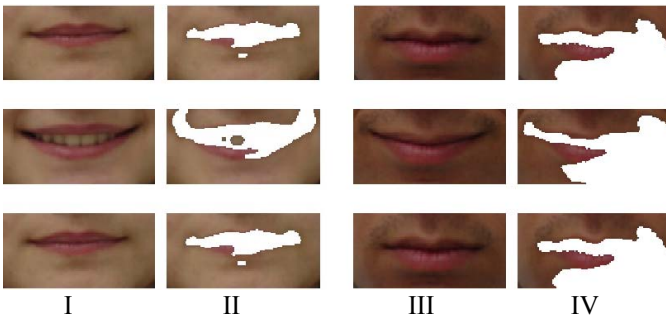


Fig.3. Lip tracking without prior color image segmentation,
(I) original image with the presence of teeth, (II) lip tracking of (I),
(III) original image without the presence of teeth, (IV) lip tracking of (III)

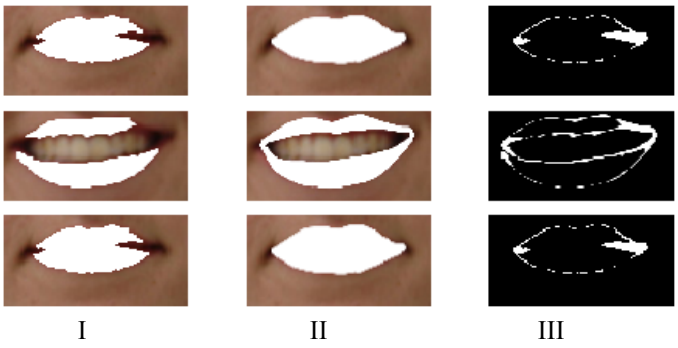


Fig.4. Example of lip segmentation and tracking using k-means
where the inputs are Cb and Cr components in YCbCr color space,
(I) automatic segmentation, (II) manual segmentation,
(III) False Positive and False Negative

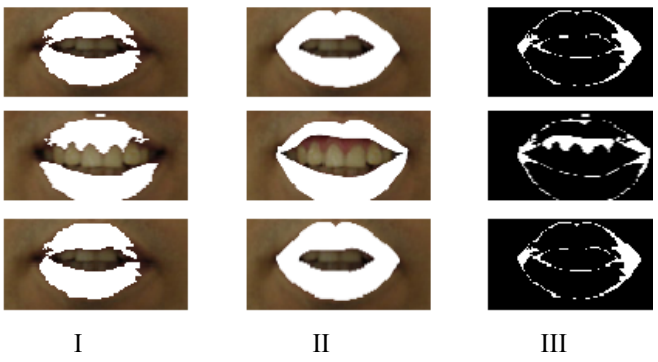


Fig.5. Example of lip segmentation and tracking using k-means
where the inputs are I and Q components in NTSC color space,
(I) automatic segmentation, (II) manual segmentation,
(III) False Positive and False Negative

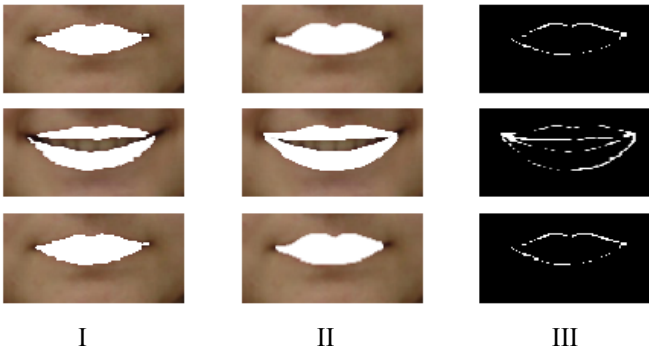


Fig.6. Example of lip segmentation and tracking using our proposed method
(k-means where the inputs are a* and b* components in CIE-L*a*b* color space), (I) automatic segmentation, (II) manual segmentation,
(III) False Positive and False Negative

The accuracy of segmentation of 60 images, (20 subjects, each of 3 images), for three kinds color image preprocessing, presented in Fig.7. The summary of lip segmentation accuracy is shown in Table 1.

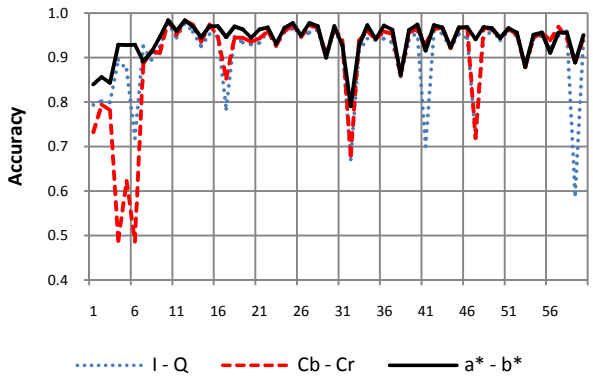


Fig.7. Curve of comparison of lip segmentation accuracy,
for three kinds color image processing

Table 1: Summary of lip segmentation accuracy

	Accuracy		
	I - Q	Cb - Cr	a* - b*
The lowest	0.592888	0.486059	0.790570
The highest	0.978383	0.983083	0.983396
Mean	0.907092	0.906386	0.940666
Standard deviation	0.081548	0.107631	0.039814

It can be seen that lip segmentation and tracking using our proposed method is better than the two other methods.

V. CONCLUSION

This paper presented a new method for lip segmentation and tracking based on Chan-Vese model, preceded by color segmentation using k-means and a*-b* component of CIE-L*a*b* color space. The result shows that our proposed method achieves 94% mean accuracy, and is better than the two other methods, i.e. using I-Q and Cb-Cr components.

As future work, we intend to explore the implementation of our proposed method in real-time application.

REFERENCES

[1] B. Denby, T. Schultz, K. Honda, T. Hueber, J.M. Gilbert, and J.S. Brumberg, "Silent speech interfaces," Speech Communication 52 (2010) 270–287

[2] U. Saeed, "Person identification using behavioral features from lip motion," *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on* , vol., no., pp.155,160, 21-25 March 2011

[3] Meng Li and Yiu-Ming Cheung, "Automatic Segmentation of Color Lip Images Based on Morphological Filter,"AMT'10 Proceedings of the 6th international conference on Active media technology Pages 296-305 Springer-Verlag Berlin, Heidelberg, 2010

[4] Siew Wen Chin, Kah Phooi Seng, Li-Minn Ang, and King Hann Lim, "New Lips Detection and Tracking System", Proceedings of the International Multi Conference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, March 18 - 20, 2009

[5] Meng Li and Yiu-ming Cheung. "Automatic lip localization under face illumination with shadow consideration," *Signal Processing,* 89, 12 (December 2009), pp. 2425-2434.

[6] H. Talea, and K. Yaghmaie, "Automatic combined lip segmentation in color images," *Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on* , vol., no., pp.109,112, 27-29 May 2011

[7] Young-Un Kim, Sun-Kyung Kang, and Sung-Tae Jung, "Design and implementation of a lip reading system in smart phone environment," *Information Reuse & Integration, 2009. IRI '09. IEEE International Conference on* , vol., no., pp.101,104, 10-12 Aug. 2009

[8] Xiaoping Wang, Yufeng Hao, Degang Fu, and Chunwei Yuan, "ROI processing for visual features extraction in lip-reading," *Neural Networks and Signal Processing, 2008 International Conference on* , vol., no., pp.178,181, 7-11 June 2008

[9] S.L. Wang, W.H. Lau, and S.H. Leung, "Automatic Lip contour extraction from color image", *Pattern Recognition*, 37, 2004, pp.2375-2387.

[10] E. Skodras, and N. Fakotakis, "An unconstrained method for lip detection in color images," *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, vol., no., pp.1013,1016, 22-27 May 2011

[11] S. L. Wang, W. H. Lau, S. H. Leung, and A. W. C. Liew, "Lip segmentation with the presence of beards," *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol.3, no., pp.iii,529-32 vol.3, 17-21 May 2004

[12] M.M. Hosseini, and S. Ghofrani, "Automatic Lip Extraction Based on Wavelet Transform," *Intelligent Systems, 2009. GCIS '09. WRI Global Congress on*, vol.4, no., pp.393,396, 19-21 May 2009

[13] Mark Barnard, Eun-Jung Holden, and Robyn Owens, "Lip tracking using pattern matching snakes," *ACCV2002: The 5th Asian Conference on Computer Vision*, 23–25 January 2002

[14] T.F. Chan, and L.A. Vese, "Active contours without edges," *Image Processing, IEEE Transactions on*, vol.10, no.2, pp.266,277, Feb 2001

[15] L.A. Vese, T.F. Chan, "A multiphase level set framework for image segmentation using the Mumford-Shah model", International Journal of Computer Vision, 50 (2002) 271-293.

[16] T.M. Le, and L.A. Vese, "Additive and multiplicative piecewise smooth segmentation models in functional minimization approach", Contemporary Mathematics, 445 (2007) 207-223.

[17] O. Rousseau, and Y. Bourgault, "Heart Segmentation With An Iterative Chan-Vese Algorithm", University of Ottawa, Ontario , 2009

[18] P. T. H. Truc, S. Lee, and T. Kim; "A Density Distance Augmented Chan-Vese Active Contour for CT Bone Segmentation", *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE* , vol., no., pp.482,485, 20-25 Aug. 2008

[19] S. Shah, A. Abaza, A. Ross and H. Ammar, "Automatic Tooth Segmentation Using Active Contour without Edges", *Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the* , vol., no., pp.1,6, Sept. 19 2006-Aug. 21 2006

[20] S. Gao, and Y. Yan, "Brain MR Image Segmentation via a Multiphase Level Set Approach" Journal of Information & Computational Science 9: 16 (2012) 4705–4711

[21] S. Gao, and L. Wang, "Local Chan-Vese Model for Segmenting Nighttime Vehicle License Characters", Journal of Information and Computing Science Vol. 6, No. 2, 2011, pp. 123-128

[22] P. Ahmadi, S. Sadri, R. Amirfattahi, and N. Gheissari, "Automatic Aerial Image Segmentation based on a Modified Chan-Vese Algorithm", *Image and Signal Processing (CISP), 2012 5th International Congress on* , vol., no., pp.643,647, 16-18 Oct. 2012

[23] M. Moelich and T. Chan. "Tracking Objects with the Chan-Vese Algorithm". UCLA CAM Reports, 03-14, 2003.

[24] M. Maska, A. Munoz-Barrutia, and C. Ortiz-de-Solorzano, "Fast tracking of fluorescent cells based on the Chan-Vese model," *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on* , vol., no., pp.1316,1319, 2-5 May 2012

[25] fei.edu.br/~cet/facedatabase.html

Word Similarity algorithm for Merging Thai Herb Information from Heterogeneous Data Sources

Phakphoom Chainapaporn, Ponrudee Netisopakul
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
chphakphoom@gmail.com, ponrudee@it.kmitl.ac.th

***Abstract*—This paper proposes two processes for merging Thai Herb information obtained from heterogeneous data sources. The objective is to combine different formats of Thai herb information into one consistent representation. The processes are implemented in a Sourcing and Merging Agent (SMA) of a Multi-Agent Thai Herb Recommendation system (MA_THR). The first process aims to find and merge the same Thai herb with different names. The second process aims to find synonyms of symptoms. Experiments give 93% accuracy of merging Thai herb information using names and 97% accuracy of finding the similarity between symptoms.**

Keywords—heterogeneous data source; merging algorithm; word similarity; Thai herb information

I. INTRODUCTION

Thai herbs information can be found publicly on various data source [1-5]. However, these sources present the information in different details and formats. For example, most of the sources contain a name of a Thai herb, but some contain only a common name (or names) without providing a scientific name. More seriously, while most sources provide names of symptoms that can be treated by each Thai herb or a part of Thai herb, these lists of information (symptom names) are different. Since each symptom can be called by different names. The differences of these lists can be reduced by finding synonym of symptom's names.

This paper proposes two processes: one for merging Thai herb names and another for finding similarities between symptoms. We implement the proposed algorithm on our Multi-Agent Thai Herb Recommendation system (MA_THR) [6, 7], which is briefly explained here.

The MA_THR system has components related to the proposed work as shown in figure 1. There are a number of WA agents (WA-Wrapper Agents) for retrieving Thai herb information from various databases. In addition, there is one WEA agent (WEA-Web Extraction Agents) for extracting Thai herb information from multiple websites. The information is sent to a SMA agent (SMA-Sourcing and Merging Agent) to merge into one knowledge-base, then store at a THMA (Thai Herb Management Agent). The CA (CA-Center Agent) is responsible for communication among these agents

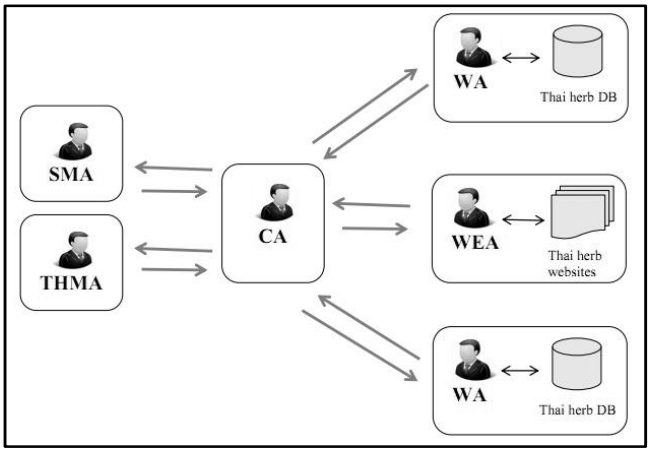


Fig. 1. MA_THR's components for merging Thai herb information

II. RELATED RESEARCH

In the past few years, there are many researches that work on word similarity. Those researches can be classified into two models. The first model is based on string matching algorithms [8-11]. The principle of this model is to find a number of characters matched and characters un-matched and normalize with a common divider. The similarity score can be ranked from 0 (no similarity) to 1 (closely similar).

The other model is based on semantic similarity. Researches based on this model often use WordNet to measure the similarity between words. The famous measure that many researchers used is an edge-counting technique [12-15]. Given words *a* and *b*, a lowest common ancestor node of them is noted. The shorter path length from either *a* or *b* to their lowest common ancestor node is used as a denomination, which is then normalized by summation of a path length from *a* to *b* (via this common ancestor node) and their depths. Hence, the score can also be ranked from 0 (no similarity) to 1 (perfect synonymy).

Our approach to find similarities between Thai herb names uses exact-matched string matching model. However, we proposed a new algorithm for finding similarities

between symptom names. There are three reasons that we cannot use the two models above for symptom names. First, we do not have Thai WordNet for symptoms. Second, similar or same symptoms can have many names. Third, similar symptom names can be totally different symptoms.

III. THE PROPOSED SOLUTION

A. Merging Thai herb name

The objective of this process is to collect Thai herb information for both official name and other names (common names). This process is necessary in order to complete the integration of Thai herb information from various data sources. The process of merging Thai herb names has 2 main steps as follows.

First step, we compare an official name of a data source to an official name of other data sources. If official names are exact-matched, both Thai herb information are merged. Second step, we compare both official name and other names of a data source to other names of other data sources. If either official name or other names of the data source is matched to other names of other data sources, both Thai herb information are merged.

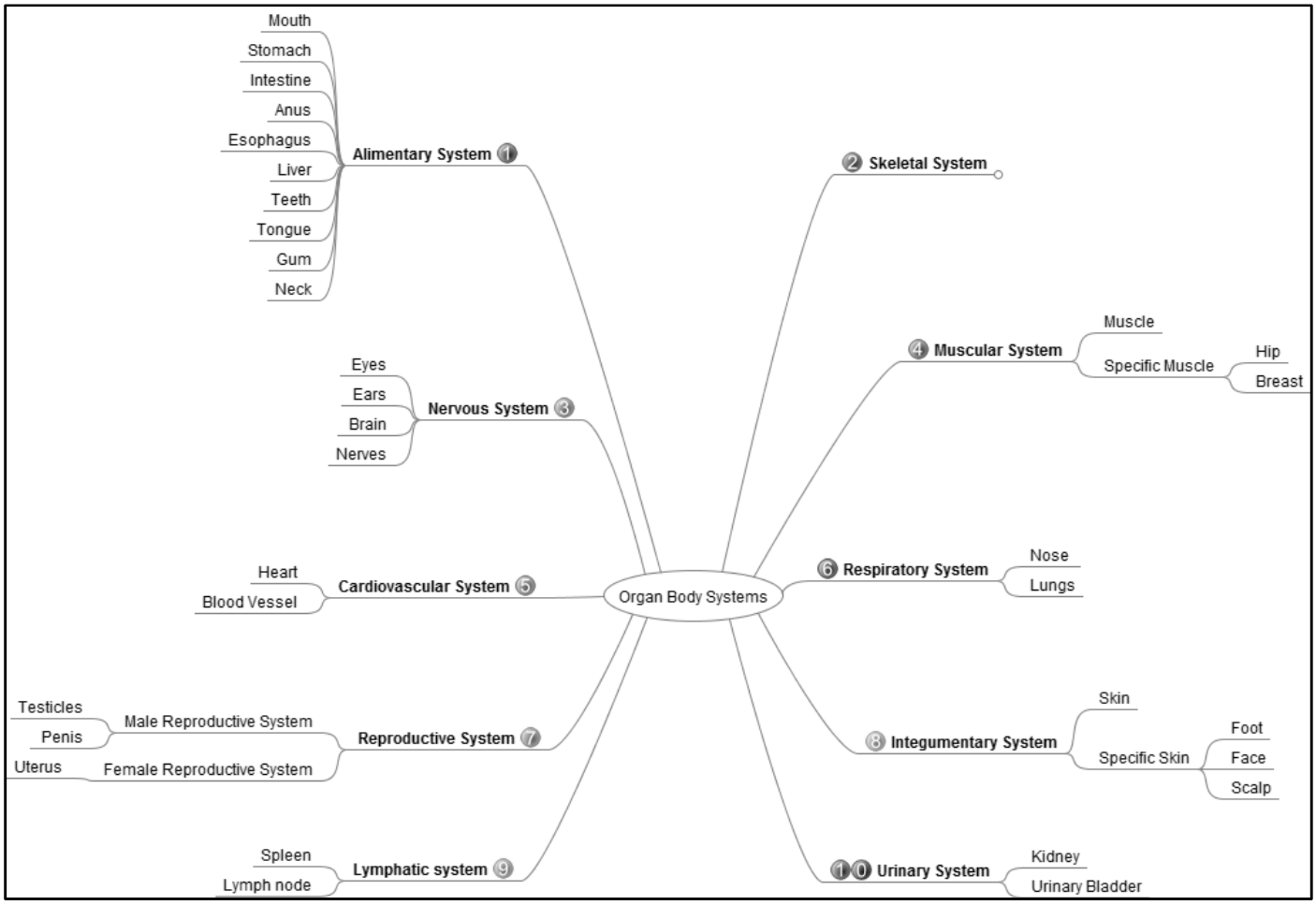


Fig. 2. The Organ Body Systems.

B. Finding similarities between symptoms

This process is the main challenge of this work because some Thai symptom names can be called by many names, such as Thongruang (ท้องร่วง) and Thongsia (ท้องเสีย). On the other hand, some totally different symptoms such as Puadhua (ปวดหัว) and Puadthong (ปวดท้อง), have high string-matching score above average. Hence, we propose an algorithm to find similarities between symptoms. The process can be divided into 4 steps as follows.

1) Step 1: Grouping body systems into sub-organ similarity tables: First, we divided groups of body organs into ten parts based on the organ body system[16] which is

shown in figure 2. Those are an alimentary system, a skeletal system, a nervous system, a muscular system, a cardiovascular system, a respiratory system, a reproductive system, an integumentary system, a lymphatic system and an urinary system. Each system contains sub-organs, such as, an alimentary system has a mouth, a stomach, an intestine, an anus and so on. Hence, ten 2-D tables are constructed; each contains similarity scores between 0 to 1 of each pair of sub-organs (based on the human expert opinion). For example, mouth and teeth has similarity score 0.7 while mouth and stomach has similarity score 0.0.

2) Step 2: Constructing a list of all symptoms organized by sub-organs: Our assumption is that the same symptom affects the same sub-organ. Hence, we construct 35 lists of symptoms which can affect each organ body part. The symptom data used in our work are extracted from five websites and two databases. Those lists are input to the next step.

3) Step 3: Constructing symptom word similarity tables: Next, we take a symptom name and separate it into words. In Thai language, one symptom is usually indicated by at least two words. One specifies a body part, the other specify the irregularities happen to the body part. These irregularity specification words (or symptom words) are grouped into tables based on sub-organs of body systems from step 1. Some of these symptom words are colic (จุกเสียด), ache (ปวด), and so on. Similar to the sub-organ similarity tables, the similarity scores are between 0 to 1. However, instead of relying on the human expert opinion, the similarity score of a symptom word pair is calculated as follows.

- For each sub-organ, if the word pair appears together at least once, count it as 1. (Hence, for totally 35 sub-organs, there can be at most 35 counts).
- Divide the count number by the number of sub-organs (35).

4) Step 4: Calculating symptom similarity score: This is the main step for calculating symptom similarity based on previous lists and tables in step 1-3. A pair of symptom names is sent to an algorithm 1 shown in figure 3. The algorithm is developed based on a standard edit distance score (distance_score) using dynamic programming from [17] with some modification. The modification is as follows. First, the input pair, symptom S1 and symptom S2, is in Thai word instead of English string. Next, if S1 and S2 are in different sub-organ lists (from step 3 above) then return distance_score 1 (mean that they are not the same symptom). Then, the pair are segmented into list of words using LexiTo [18] and sent to calculate distance_score with a new modification of diff(word1, word2) as follows:

- Step 4.1: if word1 and word2 are the same word then return distance_score 0.
- Step 4.2: if word1 and word2 are synonyms in the lexitron Thai dictionary [19] then return distance_score 0.
- Step 4.3: lookup wordscore_similarity (word1, word2) in tables from step 1 and step 2 above, and return *distance_score = 1-wordscore_similarity*.
- Finally, the distance_score is divided by a maximum length of the pair and minus by one. In sum, two symptoms S1 and S2 has similarity score defined as:

$$similarity(S1,S2) = 1 - \frac{distance\_score(S1,S2)}{Max_{Length}(S1,S2)} \quad (1)$$

This similarity score is between 0 and 1. The symptom synonym must gain similarity score above 0.6 in order to be considered the same symptom. Table I shows some result of symptom pairs.

The score 0.6 is chosen from a preliminary experiments using 20 symptoms as seeds and varied scores from 0.4, 0.5, 0.6, 0.7 and 0.8. We found that scores below 0.5 did not indicate a good pair of symptom synonyms. In addition, only few symptom synonyms can achieve scores above 0.7. This due to the nature of symptom word pairs which have length mostly between two and three.

```
Function main_algorithm(String S1, String S2){

    IF(S1 and S2 are in the same sub-organ){

        List  listWord1 = wordSegmentation(S1)
        List  listWord2 = wordSegmentation(S2)


        FOR i =0 to size of listWord1:
            distance_score [i, 0] = i

        FOR j = 1 to size of listWord2 :
            distance_score [0, j] = j

        FOR i = 1 to size of listWord1 :
            FOR j = 1 to size of listWord2 :
                distance_score [i, j] = min { distance_score [i-1,  j] + 1,
                    distance_score [i,  j-1] + 1,
                    distance_score [i-1,  j-1] +  diff(listWord1(i), listWord2(j) ) }

        return 1- ( distance_score (size of listWord1,  size of listWord2) /
                    maxLength(S1, S2))

    }ELSE{
        return  0
    }
}

Function diff(word1, word2){

    IF( word1 and word2 are the same word)
        return 0
    }ELSE IF(word1 and word2 are synonym in Lexitron dictionary ){
        return 0
    }ELSE{
        return 1 – wordscore_similarity(word1, word2)
    }
}
```

Fig. 3. The algorithm to calculate the similarity between symptoms from lists of symptoms

TABLE I. THE RESULT EXAMPLE OF SYMPTOM PAIRS.

Symptom1	Symptom2	Similarity score
Thong-uet (ท้องอืด)	Thongfoe (ท้องเฟ้อ)	1
Phaen-ron-nai-nai-pak (แผลร้อนในในปาก)	Phaen-nai-pak (แผลในปาก)	0.75
Patsawa-khun- khon (ปัสสาวะขุ่นข้น)	Patsawa-khun (ปัสสาวะขุ่น)	0.667
Batphaen-tit-chuea (บาดแผลติดเชื้อ)	Phaen Fai Mai (แผลไฟไหม้)	0.5
Wat (หวัด)	Phit-khai-semha (พิษไข้เสมหะ)	0.333
Thongsia (ท้องเสีย)	Khluensai-achian (คลื่นไส้อาเจียน)	0

IV. EXPERIMENTAL EVALUATION

We conduct experiments to measure the accuracy of the two tasks above. Task 1 is merging Thai herb name and task 2 is finding similarity between symptoms. The input collection are from five websites [1-5] and two databases constructed from two books [20, 21]. The result is manually verified. The accuracies for task 1 and task 2 are calculated from results shown in table 2 and table 3, using formula given below.

TABLE II. RESULTING MEASUREMENTS FOR CALCULATING ACCURACY OF MERGING THAI HERBS TASK (TASK 1).

Merging results	synonyms Thai herb names	non_synonyms Thai herb names
Merged	A = a number of Thai herb names that are correctly merged by the system = 274	C = a number of Thai herb names that are incorrectly merged by the system = 29
Not_merged	D = a number of Thai herb names that are (incorrectly) not merged by the system = 15	B = a number of Thai herb names that are correctly not merged by the system = 319

$$accuracy_{task1} = \frac{A+B}{A+B+C+D} \times 100 \qquad (2)$$

In table II, the accuracy of merging Thai herb names is about 93%.

There are two main reasons for incorrect merging and not merging. The first reason, some different Thai herbs have the same other names such as the other names of Khing (ขิง) are "Khing-daeng (ขิงแดง), Khing-phueak (ขิงเผือก)" and the other names of Krachai (กระชาย) are "Khing-sai (ขิงทราย), Khing-daeng (ขิงแดง)". In our algorithm, these Thai herbs are incorrectly merged because Khing-daeng (ขิงแดง) is the other name of both Thai herbs.

The second reason, some Thai herbs data sources provide only the official name which is no match to other Thai herbs in our existing data source. For example, "Thonglang-dang (ทองหลางด่าง)" is an only official name from one source and it is not matched to any Thai herbs from our data sources. However, after manually verified with other reference sources (outside our data sources), we found that the other name of this Thai herb is "Thong-ban (ทองบ้าน)", "Thonglang (ทองหลาง)". Both are presented in our database and did not get merged with "Thonglang-dang (ทองหลางด่าง)".

TABLE III. RESULTING MEASUREMENTS FOR CALCULATING ACCURACY OF FINDING SYMPTOM SYNONYMS TASK (TASK 2).

Symptom synonym calculation	True synonym symptom pairs	Not synonym symptom pairs
Similarity $\geq$ 0.6	M= a number of true synonym symptom pairs that have similarity > 0.6 = 716	N = a number of not synonym symptom pairs that have similarity $\geq$ 0.6 = 78
Similarity < 0.6	O = a number of true synonym symptom pairs that have similarity < 0.6 = 1657	P= a number of not synonym symptom pairs that have similarity < 0.6 = 68,966

$$accuracy_{task2} = \frac{M+P}{M+N+O+P} \times 100 \qquad (3)$$

From table III, we use 1139 symptom names to evaluate our proposed algorithm. We found that the accuracy of finding symptom pairs synonym is about 97%. Here are some reasons that the proposed algorithm works well. First, symptoms that affect different sub-organs are put in different sub-organ lists; hence, they receive a similarity 0. Second, we specify rather high similarity threshold (0.6); hence, only few true synonyms can pass this threshold. However, there are still rather high numbers of mis-classified results. Our investigation found the following reasons.

First, some symptom word pairs that are actually similar receive low similarity scores. For example, a similar symptom word pair such as buam (บวม) and aksep (อักเสบ), which are very closed in term of symptoms, has only a similarity score of 0.228. Hence, the default calculation favors toward "not a synonym". Second, some pairs with long symptom names have many common sub-names but the pairs are actually different symptoms. For example, Chuk-Siat Naenthong-Lae-Thongkhuen-Thongfoe (จุกเสียดแน่นท้องและท้องขึ้นท้องเฟ้อ) and Chuk-Siat Naenthong-Lae-Puatthong (จุกเสียดแน่นท้องและปวดท้อง) receive a similarity score over 0.6 but they are not the same symptom. Third, some symptom pairs with negation word insided, receive high similarity scores based on our word similarity algorithm. However, the negation indicates that they are not the same symptom. For example, Khai-Thi-Koet-Chak-Bit (ไข้ที่เกิดจากบิด) and Khai-Thi-Mai-Koet-Chak-Bit (ไข้ที่ไม่เกิดจากบิด).

Based on these observation, further improvement on the algorithm can be done in the future. First, negation words detection of a pair should reverse the value of similarity score. Next, symptom pairs with conjunction words such as *lae (และ), rue (หรือ)* should be separated into two different symptoms. Third, the basic similarity scores in the symptom

word similarity table could be verified by human experts to reflect more reliable information.

V. CONCLUSION

This paper proposes the process of merging Thai herb names and finding similarity between symptoms from heterogeneous data sources. An exact string matching algorithm is used to merge Thai herb names from different sources with accuracy of 93%. To find similar symptoms, sub-organ similarity tables and symptom words similarity tables, together with list of symptoms affecting the same sub-organ, are used as references to calculate similarity. The symptom similarity calculation is implemented based on modified edit distance dynamic programming. When applying the algorithm to calculate 1139 symptoms (71,417 pairs), it gains accuracy of 97%.

REFERENCES

[1] "Thai herb knowledge base," Dhonburi Rajabhat University, 2001. [Online]. Available: http://dit.dru.ac.th/ herb/Main.htm . [Accessed: 01-Apr-2013].

[2] "Plant Genetic Conservation Project Office," Plant Genetic Conservation Project Office. [Online]. Available: http://www.rspg.or.th/plants_data/herbs/herbs_200.htm. [Accessed: 01-Apr-2013].

[3] "Thai herb database." [Online]. Available: http://www.samunpri.com. [Accessed: 01-Apr-2013].

[4] "Sirirukkachad Nature Park," Mahidol University, 2010. [Online]. Available: http://www.pharmacy.mahidol.ac.th/ siri/index.php?page=home . [Accessed: 01-Apr-2013].

[5] "Toxic Thai herb information," Thai Herb Information Office, 2000. [Online]. Available: http://www.medplant. mahidol.ac.th/tpex/. [Accessed: 01-Apr-2013].

[6] P. Chainapaporn and P. Netisopakul, "Multi-Agent Architecture for Thai Herb Recommendation," in Proceedings of 9th Joint International Symposium on Natural Language Processing and Agricultural Ontology Service, 2011, pp. 4–9.

[7] P. Chainapaporn and P. Netisopakul, "Thai Herb Information Extraction from Multiple Websites," in Proceedings of 4th International Conference on Knowledge and Smart Technology, 2012, pp. 16–23.

[8] G. J. A. Gutha, "Surname Spellings and Computerized Record Linkage," Historical Methods Newsletter, vol. 10, no. 1, pp. 10–19, December 1976.

[9] V. I. Levenshtein, "Binary Codes Capable of Correcting Deletions, Insertions, and Reversals," Soviet Physics Doklady, vol. 10, pp. 10–19, February 1966.

[10] G. Bouchard, "The Processing of Ambiguous Links in Computerized Family Reconstruction," Quantitative and Interdisciplinary History, vol. 19, no. 1, pp. 9–19, November 1986.

[11] C. Snae and B. Diaz, "An Interface for Mining Genealogical Nominal Data using the Concept of Linkage and a Hybrid Name Matching Algorithm," Journal of 3D-Forum Society, pp. 1–9, 2002.

[12] Z. Wu and M. Palmer, "Verbs Semantics and Lexical Selection," in Proceedings of 32nd Annual Meeting of the Association for Computational Linguistics, 1994, pp. 133–138.

[13] C. Leacock and M. Chodorow, "Combining Local Context and WordNet Similarity for Word Sense Identification," in WordNet: An Electronic Lexical Database, C. Fellbaum, Ed. MIT Pres, 1998, pp. 265–283.

[14] Z. a. Bandar and D. McLean, "An approach for measuring semantic similarity between words using multiple information sources," IEEE Transactions on Knowledge and Data Engineering, vol. 15, no. 4, pp. 871–882, July 2003.

[15] X. Liu, Y. Zhou, and R. Zheng, "Measuring Semantic Similarity in Wordnet," in 2007 International Conference on Machine Learning and Cybernetics, 2007, pp. 19–22.

[16] "Organ," Wikipedia, 2012. [Online]. Available: http://th.wikipedia.org/wiki/อวัยวะ. [Accessed: 01-Apr-2013].

[17] S. Dasgupta, C. Papadimitriou, and U. Vazirani, Algorithms. McGraw-Hill, 2006.

[18] "LexTo: Thai Lexeme Tokenizer," National Electronics and Computer Technology Center. [Online]. Available: http://www.sansarn.com/lexto/. [Accessed: 01-Apr-2013].

[19] "Lexitron," National Electronics and Computer Technology Center, 1995. [Online]. Available: http://lexitron.nectec.or.th/2009_1/index_en.php?q=index.

[20] S. Matchachip, Thai Herb. Bangkok: Phraephitthaya, 2000.

[21] C. Anan, Thai Herb for Public Health. Bangkok: Duangkamon, 2008.

News Recommendation in Indonesian Language Based on User Click Behavior

Diandra Mayang Desyaputri[1], Alva Erwin[1], Maulahikmah Galinium[1], Didi Nugrahadi[2]
[1]Faculty of Engineering and Information Technology
Department of Information Technology
Swiss German University, BSD, Tangerang, Indonesia
{diandra.desyaputri@student., alva.erwin@, maulahikmah.galinium@}sgu.ac.id
[2]Beritagar
Jakarta, Indonesia
didinugrahadi@gmail.com

***Abstract*—Recommendation system has been proposed for years as the solution of information era problem. This research strives to develop an intelligent recommendation system based on user click behavior on news websites. We extracted frequent itemsets and association rules from the web server log of a news website, performed a pre-computation of similarity between news articles, and then proposed a three-level recommendation system: based on association rule discovery, news articles on the same category, and similarity between news articles. By combining collaborative filtering approach and content-based filtering, experiment results show that the technique produces reliable news recommendation.**

Keywords— web usage mining, association rules, similarity, news recommendation

I. INTRODUCTION

Supported by the fast-paced growth rate of technology that enables users to easily access worldwide news information, online news websites from around the world has become very popular. The enormous amount of information available have forced users to face information overload, such as overwhelming volumes of articles. Users have to endure tremendous volumes of information to find their desired articles. There are some cases where articles are inaccessible by users because it has been diminished behind the great load of information.

From these large sources of data and information, users like to ask for recommendations from friends, family, partners, the community, trusted people, or specialists to obtain the desired items. However, their own knowledge about particular items are limited, and yet, new articles, new websites, new blogs, new items are emerging as the clock ticks. An automatic recommendation by the intelligent system has been proposed for years as the solution for this information era problem.

Although the concept of recommendation has been introduced more than ten years ago, the technologies for support has developed greatly recently. Because of its great impact in industries as well, primarily in marketing, recommender systems have spread widely on the Internet. Not only e-commerce websites use such systems, but it has also become common for news sites or online news providers to adopt a recommendation system to generate news articles that would keep users reading.

An interesting point about user behavior in searching news websites that may affect the recommendation system is the fact that users access news websites with a 'show me something interesting' mindset, rather than knowing what they want to read beforehand. To generate recommendations for a particular user based only from his/her past news preferences, the result will be biased only to those topics. Moreover, there is a problem for new users who haven't had any transaction logs and the system can't recommend articles based on their previous activities. This problem is called 'cold start problem'. It is also a responsibility of the system to balance the wants and needs of users, and the fast-changing news everyday.

A "related news" column is often found in news websites, which links one article to others to keep users reading. This is usually done by the editor who needs to manually select interesting news from many articles that keep increasing everyday. Moreover, news is changing every second that it can be overwhelming to do this task by using manual effort.

Given the problems above, this research strives to develop an intelligent recommendation system based on user click behavior on news websites. The proposed system of this study will be implemented on an online Indonesian news provider. Therefore the web usage log will be provided by that Indonesian news provider, and the recommended articles will be generated from clickstream data as the basis.

To provide a detailed description, this paper is organized into 5 chapters. Chapter 1 serves as the introductory part of this study. This chapter describes the background, the general idea and problems of today's current state of information. Chapter 2 narrates related works of other research in this topic. Chapter 3 presents the proposed system and explains about the detailed steps. Chapter 4 presents the research findings and Chapter 5 describes the conclusion.

II. Related Work

As users browse the Internet, their actions and movements are recorded in the web server logs. From these web server logs, patterns can be extracted which can create a user model based on the articles that they have read in the past. By analyzing the navigational pattern of user behavior, for example in a news website, we can discover the preferences of that particular user and later suggest news articles to him.

Web Usage Mining [1] is a technique for extracting knowledge from patterns of users' behavior, such as user access data collected from web log files.

However, the data collected from various sources, including the web server, is not always complete and ready to be analyzed. On the contrary, raw data is usually inaccurate, erroneous, and incomplete. This may lead to misleading and incorrect information. For that reason, data preprocessing tasks are required to be performed. In recent years, researchers [2] [3] [4] are continuously finding intelligent ways to perform data preprocessing and preparation on web log file.

There are also numerous studies done by various researchers in data mining field regarding the use of association rule in recommendation system and web usage mining. A study by Peng [5] discovered association rules from web usage data by using FP-Growth algorithm. He combined the interest measure and website users to view the topology of the best access time of the user in order to optimize the interesting association rules. Mobasher et al. [6] proposed a scalable framework for Web personalization based on association rule discovery. They presented a data structure to store discovered frequent itemsets and used a fixed- size sliding window to capture the current session of a user. As a result, this framework is able to produce recommendations efficiently in real-time. Another recommendation system based on association rule mining is proposed by Lin et al. [7], which requires no minimum support to be specified in advance. They proposed an algorithm that automatically adjusts the minimum support so that the number of rules generated likes within a specified range.

In another study [8], association rules are applied to extract potentially useful knowledge from web usage data. They use Weka, a data mining software to discover association rules in web log data. A recent study in recommendation system based on association rules showed an improved result in performance and efficiency by combining association rule mining with user clustering [9].

The generation of association rules is done online in another study [10]. Due to the dimension of log files and high processing time, the first two phases are being achieved offline, in a batch process, while the recommendation phase usually performed online and real-time. Once the session has been processed, rules extracted and inserted, it will be deleted from the sessions table. As a result, the system has been a scalable model for recommendation since it can work with large datasets in real-time.

Recommendation aims to timely provide suitable and valuable information according to users' demand, and such information will be used as reference for supplementing decision-making [11]. there are several common and widely used technologies for personalization and recommendation systems. One of them is Content-based method (CB). In Content- based filtering, the user will be recommended items similar to the ones that the user preferred in the past. Meanwhile, in Collaborative filtering (CF), the user will be recommended items that similar people will prefer in the past [12].

Many proposed Hybrid method, that is the combination of Content-based method and Collaborative filtering. This method has been experimented in various studies with various source of data, and has proven to be more effective since it provides satisfying results, with improving the quality of recommendations, as shown in [13].

A scalable two-stage personalized news recommendation approach with a two-level representation [14] considers the exclusive characteristics of news items when performing recommendation. The first level contains various topics relevant to users' preference, and the second level includes specific news articles. They also presented a principled framework for news selection based on user interest with a good balance between the novelty and diversity of the recommended result.

Various methods to measure similarity between news articles for a news recommender system have been researched by Tintarev and Masthoff [15]. Commonly, news recommendation is performed using the combination of TF-IDF technique and cosine similarity measure [16]. However, they proposed two new approaches: SF-IDF (Synset Frequency - Inverse Document Frequency) that is similar to TF-IDF but uses WordNet synonym set, and Semantic Similarity by combining five semantic similarity measures. Cosine similarity approach for content-based news recommendation has also been researched by Kompan and Bielikova [17].

In this research, we combined the methods that have been studied before and developed a news recommendation system in Indonesian language. The recommended news generated are from the discovery of association rules from web log files, which is one of the collaborative filtering approach in recommendation system, and also from the pre-computed similarity between news articles, which is one of the content-based approach. Furthermore, we analyze the most popular news articles in the associated hashtag and append it to generate more solid news recommendation.

III. Proposed System

Web recommendation system in general is composed of 3 phases: data preparation and transformation, pattern discovery, and recommendation [6]. Figure 1 depicts the process of the proposed system.

Data is retrieved from the web server, in form of web server access logs that contain every transaction performed on the server. Then, data cleaning is performed to improve the accuracy of the recommendation result. Data filtering is then performed to make sure that the data processed is indeed necessary.

Session identification is executed to complete the data preparation process. After that, the pattern discovery phase

may begin, which has three levels: association rule discovery, news discovery from associated #hashtag, and discovery of similarity measure of news articles. Based on those three levels, the recommender systems generate the recommended articles.

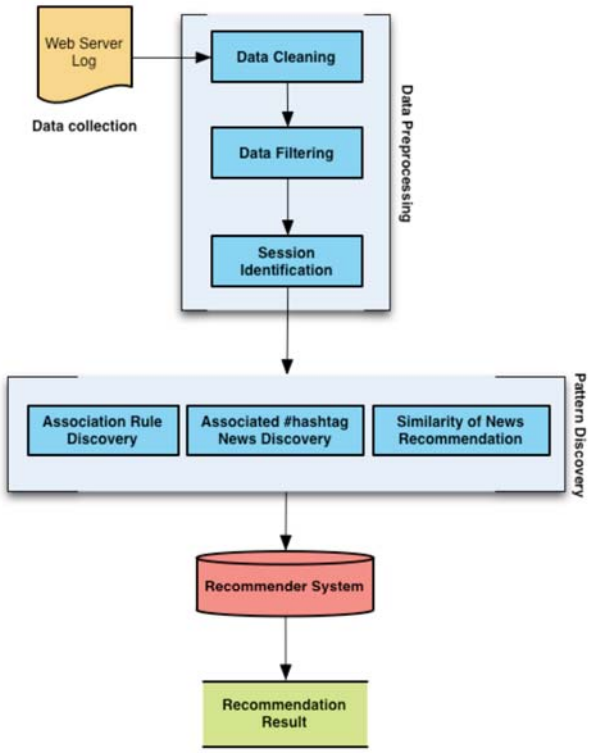


Fig. 1. Proposed System

A. Data Preprocessing

There are three steps in data preprocessing used in this study: data cleaning, data filtering, and session identification.

Every time a user browses a website, the web browser downloads an HTML document. The resources embedded in HTML files such as images, style files (javascript files, css files) are also automatically downloaded without the user explicitly requesting it. They might be advertisements that have no relationship with the content of the web page. These actions are recorded in the web log file, therefore making it extremely congested and filled with information unrelated to the content of the news articles.

Therefore, all data on the web log server that is not needed for processing is cleaned off, such as image files, javascript files, multimedia files, style files, spider requests, error requests, and web robot requests. The records with HTTP status codes over 299 or under 200 are also removed, since only records with HTTP status codes between 200 and 299 give successful response. The records with POST or HEAD method also should be removed since transactions from the user is the only relevant information that is needed to be processed.

Data filtering consists of filtering the data needed to use in this study. Since we need to generate only the title of news articles from the web log files, we cleaned off other unnecessary records, such as category URLs, homepage URL, and other pages such as the login and registration page.

The next step in data preparation is session identification. Session identification is done by using Rapidminer, by transforming the records of web log files into set of sessions. Rapidminer is an open-source data mining and data analytics tool which provides data mining and machine learning procedures.

B. Pattern Discovery

After the web logs are cleaned and sessions are identified, the process of pattern discovery can be performed. From each segmented and fragmented web log files, frequent itemsets and association rules are generated, and then the results are combined.

Association rule mining, which is one of the data mining techniques, discovered unordered correlations between items found in a database of transactions [18]. These interesting relationships can be represented in the form of association rules or frequent itemsets [19]. First introduced by [20], the association rule mining algorithms discovers all item associations or rules in the data that satisfy the user-specified constrains: minimum support, or called minsup, and minimum confidence, or called minconf [21].

Frequent itemsets are defined as "group of items occurring frequently together in many transactions" [22]. We use FP-growth to generate the frequent itemsets. The implementation is also done in Rapidminer. Because the FP-growth algorithm implementation in Rapidminer require the data to be in binomial form, further data preprocessing is necessary after session identification.

Given a set of items $I = \{i_1,i_2,...,i_k\}$, where each i is an item or news articles, and let $T = \{t_1,t_2,...,t_n\}$ be a set of all transactions where each transaction t is a set of items such that $t \subseteq I$. Association rules, on the other hand, capture relationships among pageviews based on the navigational patterns of users [6]. An association rule is a rule of the form $X \rightarrow Y$, where X and Y are subsets of I and $X \cap Y = \emptyset$. Association rules are mainly defined by two metrics: support and confidence. Support is the percentage of transactions that contain both X and Y among all transactions in data set, while confidence is the percentage of transactions that contain Y among transactions that contain X [7].

An example of association rule is as follows:

Example 1. [/p/bingkisan-ulang-tahun-dari-google-play]→ [/p/perubahan-desain-google-dahului-facebook] (support: 0.003, confidence: 0.750)

This rule says that 0.03% of transactions contain these two news articles, and for 75% transactions that contain the news article "Bingkisan Ulang Tahun dari Google Play" (Birthday present from Google Play) also contains the news article "Perubahan Desain Google Dahului Facebook" (Google+ preceded Facebook in design changes). This means that users who

read the former article also read the latter article 75% of the time.

C. News Recommendation System

In this phase, a recommendation system will generate recommended news articles as related news for every news article automatically. The recommended news is generated in three stages.

1) Recommended News from Association Rule Discovery: First, the recommended news is generated from frequent itemsets. Sets of recommended news will be generated based on the mined associative patterns.

However, there are problems regarding this approach. For the recently developed news provider websites, the lack of "related news" section in every news article has resulted in a high number of bounce rate. As a result, there are few news articles in each session. Few association rules can be generated, and not every news article has associative rules of other news articles.

To generate accurate association rules, each session ideally has around 5-7 news articles. Yet each session in this Indonesian news provider has only around 2-3 news articles based on our study. In order to increase the session length, we use the second level of the proposed news recommendation system.

2) Recommended News from Associated #hashtag: It is now common to have at least one #hashtag in every news article. A hashtag is considered as one category or subcategory. It may represent the broader level of topics, such as Politics, Music, or Technology; and it may also represent the more specific level topics such as Special News, Kpop, and Gadget.

One news article may have more than one hashtag, but at least must have one hashtag. These hashtags are useful to deliver related articles as recommendation since it cluster the news articles based on category.

To increase the user's involvement in the news provider website and to reduce bounce rate, recommended news articles in "related news" section will be generated by identifying the hashtags inside a news article, then suggest five news articles of the most popular articles in that category which is represented by the hashtag. Since a news article may have more than one hashtag, there might be more than five recommended articles. The number of recommended news articles will be selected based on the number of hashtags a news article has, and is designed to fit the related news section in each news article.

3) Recommended News from Similar News: The third level of recommendation system is done by generating recommended news articles based on the similarity between them. Similarities between the title of news articles are being precomputed by using semantic analysis, cosine similarity to be exact. Few algorithms to compute sentence similarity had been researched by [23] and semantic analysis has been chosen in this paper because of its robust performance and fast computation. The result of similarity between two article titles by using semantic analysis gives a higher rating compared to similarity using statistical approach.

IV. RESEARCH FINDINGS

We analyzed the web log files of an online Indonesian news provider from 1 October 2012 to 26 March 2013. By doing data cleaning, the large size of raw web log files which was 980.1Mb has been successfully reduced to 93.3Mb. However, the time taken to process the reduced file size is still unreasonably long, and therefore, the computer used to execute this process ran out of memory before the pattern extraction phase had been finished. Therefore, the cleaned web log files are then segmented based on months.

After being segmented, it turns out that segmented log files are varied in size, because it depends on the number of transactions that has been done in that one month. Each log file is then segmented again per day. Each file then will be around 400Kb-1.5Mb in size. The data will be easy and quick to be processed. The sample web log files after data cleaning is represented in Table I.

TABLE I. THE WEB LOG FILES AFTER DATA CLEANING

IP Address	Date Time	URL	Referrer
91.201.64.24	07/Mar/2013: 00:06:16 +0700	/p/spam-bisa-berbuah-trending-topic	-
65.52.0.95	07/Mar/2013: 00:08:57 +0700	/p/ceo-apple-tertarik-teknologi-beats	-
223.255.226.10	07/Mar/2013: 00:13:14 +0700	/p/jokowi-masyarakat-tak-suka-pejabat-eksklusif	http://beritagar.com
223.255.226.10	07/Mar/2013: 00:13:41 +0700	/p/diet-yang-tepat-untuk-pria	http://beritagar.com

Sample of the generated association rules are presented below. These association rules are generated from the frequent itemsets which its samples are presented in Table II.

TABLE II. SAMPLE OF THE GENERATED FREQUENT ITEMSETS

Support	Item 1	Item 2	Item 3
0.003	/p/konflik-sabah-paksa-tki-pulang	/p/bisnis-batu-bata-lesu-akibat-hujan	
0.003	/p/pesepeda-bawa-instagram-keliling-dunia	/p/bingkisan-ulang-tahun-dari-google-play	/p/7-maret-gaya-kotakkotak-mondriaan-dan-mondrian
0.003	/p/pesepeda-bawa-instagram-keliling-dunia	/p/bingkisan-ulang-tahun-dari-google-play	/p/perubahan-desain-google-dahului-facebook
0.003	/p/bingkisan-ulang-tahun-dari-google-play	/p/perubahan-desain-google-dahului-facebook	/p/7-maret-lagu-amal-we-are-the-world
0.003	/p/film-berwarna-pertama-buatan-indonesia	/p/7-maret-mochtar-lubis-dan-mac	/p/bisnis-batu-bata-lesu-akibat-hujan

• [/p/bingkisan-ulang-tahun-dari-google-play]→ [/p/perubahan-desain-google-dahului-facebook] (confidence: 0.750)

• [/p/bingkisan-ulang-tahun-dari-google-play]→ [/p/pesepeda-bawa-instagram-keliling-dunia, /p/perubahan-desain-google-dahului-facebook] (confidence: 0.778)

• [/p/bayi-sehat-tidak-butuh-suplemen-selain-asi]→ [/p/bahaya-air-putih-untuk-bayi, /p/pengaruh-makananminuman-ibu-pada-asi, /p/relasi-asi-dan- berat-badan-anak] (confidence: 0.857)

• [/p/facebook-juga-akan-pakai-hashtag]→ [/p/chris-messina-si-penemu-hashtag, /p/akun-dong-hae- diretas-elf-bereaksi] (confidence: 0.857)

The english translation of association rules above are:

• [birthday present from google play] → [google+ preceded facebook in design changes] (confidence: 0.750)

• [birthday present from google play] → [cyclist carries instagram around the world, google+ preceded face- book in design changes] (confidence: 0.778)

• [healthy baby doesn't need any suplement besides mother's breast milk] → [drinking water hazards for baby, impact of mother's meal to her baby, relation- ship between mother's breast milk and child's weight] (confidence: 0.857)

• [facebook will use hashtag too] → [chris messina, the founder of hashtag], [dong hae's account is hacked, elf reacted] (confidence: 0.857)

We perform an experiment of assigning support and confidence to be applied to the FP-growth algorithm in order to generate association rules. The experiment is done to select the suitable amount of support and confidence such that the association rules generated are optimal. We used a sample dataset that contains 2572 lines (records) of log files. The result of the experiment is presented in Table III.

TABLE III. NUMBER OF ASSOCIATION RULES GENERATED

Support	Confidence									
	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
0.01	*127*	*84*	*47*	*35*	*31*	*23*	*19*	*16*	*6*	*6*
0.02	*20*	*14*	*6*	*4*	*2*	*1*	-	-	-	-
0.03	*2*	*1*	-	-	-	-	-	-	-	-
0.04	-	-	-	-	-	-	-	-	-	-
0.05	-	-	-	-	-	-	-	-	-	-
0.06	-	-	-	-	-	-	-	-	-	-
0.07	-	-	-	-	-	-	-	-	-	-
0.08	-	-	-	-	-	-	-	-	-	-
0.09	-	-	-	-	-	-	-	-	-	-

To recommend news articles in the related news section of "Bingkisan ulang tahun dari Google Play" (Birthday present from Google Play) based on the popularity of news articles in associative #hashtag, we identify the #hashtag associated with that news article and the date it was created. Since the #hashtags are #tekno and #gadget, we find the most popular news articles in those two #hashtags or categories.

The similarity between news articles are done by comparing a particular news article to all news articles available for the time being. Table IV presents the similarity comparison of sample news articles to "Bingkisan ulang tahun dari Google Play" (Birthday present from Google Play), sorted by the rate of similarity between news articles.

TABLE IV. SIMILARITY OF SAMPLE NEWS ARTICLES

News Title	Similarity Result
perubahan-desain-google-dahului-facebook	0.8613972559064530
google-luncurkan-sistem-peringatan-tsunami-di-jepang	0.8376697156768023
picasa-dileburkan-ke-dalam-google	0.8347221919284370
bahaya-air-putih-untuk-bayi	0.8308834468966150
akun-dong-hae-diretas-elf-bereaksi	0.8276263019935370
jokowi-masyarakat-tak-suka-pejabat-eksklusif	0.8090434224276720
bayi-sehat-tidak-butuh-suplemen-selain-asi	0.7979124279015620
tips-memilih-celana-denim	0.7804312953296150
arti-penting-foto-jurnalistik	0.7794151494702410
facebook-juga-akan-pakai-hashtag	0.7726957209327540
keseharian-warga-permukiman-kumuh	0.7291831881157720

Recommended news articles in the "related news" section:

For a news article titled: "Bingkisan ulang tahun dari Google Play" (Birthday present from Google Play), the recommended news articles are:

• Perubahan desain Google+ dahului Facebook

• Pesepeda bawa Instagram keliling dunia

• Google luncurkan sistem peringatan tsunami di Jepang

• Picasa, dileburkan ke dalam Google+

The english translation of the recommended news articles above are:

• Google+ preceded Facebook in design changes

• Cyclist carries Instagram around the world

• Google published tsunami warning system in Japan

• Picasa, merged with Google+

V. CONCLUSION

This study proposed a news recommendation system which is implemented in one of Indonesia's online news website. We propose a three-level approach news recommendation system: based on association rule discovery, the most popular news articles in the same category, and similarity between news items.

This research shows that a three-level approach proposed showed promising results as recommended news. However, tweaking the system is necessary, such as adjusting the support and confidence for optimizing the association rule mining algorithm. By combining collaborative filtering approach and content-based filtering, experiment results show that the technique produces reliable news recommendation in Indonesian language.

ACKNOWLEDGMENT

The authors would like to thank Fakhrizal Lukman, a front-end web and mobile developer in the Indonesian news provider who has supported us in our research.

REFERENCES

[1] Cooley, R., Mobasher, B., Srivastava, J., "Web mining: Information and pattern discovery on the world wide web," in *International Conference on Tools with Artificial Intelligence*, pp. 558–567, IEEE, 1997.

[2] V. Chitraa and A. S. Davamani, "A survey on preprocessing methods for web usage data," *International Journal of Computer Science and Information Security*, vol. 7, No. 3, 2010.

[3] G. T. Raju and P. S. Satyanarayana, "Knowledge discovery from web usage data: Com- plete preprocessing methodology," *International Journal of Computer Science and Network Security*, vol. 8 No. 1, 2008.

[4] M. Helmy, A. Wahab, M. Norzali, H. Mohd, H. F. Hanafi, M. Farhan, and M. Mohsin, "Data pre-processing on web server logs for generalized association rules mining algorithm," December 2008.

[5] H. Peng, "Discovery of interesting association rules based on web usage mining," in *2010 International Conference on Multimedia Communications (Mediacom)*, pp. 272–275, IEEE, 2010.

[6] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Effective personalization based on association rule discovery from web usage data," in *3rd International Workshop on Web Information and Data Management*, WIDM '01, (Atlanta, Georgia, USA), pp. 9–15, ACM, November 9, 2001 2001.

[7] W. Lin, S.A. Alvarez, C. Ruiz, "Collaborative recommendation via adaptive association rule mining," *Data Mining and Knowledge Discovery*, 2000.

[8] M. Dimitrijevic and Z. Bosnjak, "Discovering interesting association rules in the web log usage data," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 5, pp. 191–207, 2010.

[9] H. Liu and X. Liu, "A personalized recommendation system combining user clustering and association rules with multiple minimum supports," in *2nd International Conference on Future Computers in Education*, 2012.

[10] D. Mican and N. Tomai, "Association-rules-based recommender system for personalization in adaptive web-based applications," in *10th International Conference on Web Engineering*, ICWE'10, (Vienna, Austria), pp. 85–90, Springer-Verlag, July 2010.

[11] A.M. Rashid, I. Albert, D. Cosley, S.K. Lam, S. M. McNee, J. A. Konstan, J. Riedl, "Getting to know you: Learning new user preferences in recommender systems," in *International Conference on Intelligent User Interfaces*, pp. pp. 127– 134, 2002.

[12] L. M. de Campos, J. M. Fernandez-Luna, J. F. Huete, M. A. Rueda-Morales,"Combining content-based and collaborative recommendations: A hybrid approach based on bayesian networks," *International Journal of Approximate Reasoning*, vol. 51, pp. 785–799, 2010.

[13] J. Liu, P. Dolan, and E. R. Pedersen, "Personalized news recommendation based on click behavior," in *15th International Conference on Intelligent User Interfaces*, IUI '10, (Hong Kong, China), pp. 31–40, ACM, February 7-10 2010.

[14] L. Li, D. Wang, T. Li, D. Knox, and B. Padmanabhan, "Scene: a scalable two-stage personalized news recommendation system," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, SIGIR '11, (New York, NY, USA), pp. 125–134, ACM, 2011.

[15] N. Tintarev and J. Masthoff, "Similarity for news recommender systems," in In Proceedings of the AHO 06 Workshop on Recommender Systems and Intelligent User Interfaces, 2006.

[16] M. Capelle, F. Frasincar, M. Moerland, and F. Hogenboom, "Semantics-based news recommendation," in *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*, WIMS '12, (New York, NY, USA), pp. 27:1–27:9, ACM, 2012.

[17] M. Kompan and M. Bielikova, "Content-based news recommendation," in *11th International Conference, EC-Web*, (Bilbao, Spain), Springer Berlin Heidelberg, September 1-3 2010.

[18] R. Cooley, B. Mobasher, J. Srivastava, "Data preparation for mining world wide web browsing patterns," *Knowledge and Information Systems*, pp. 5–32, 1999.

[19] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Pearson, 2006.

[20] R. Agrawal, T. Imielinski, A. Swami, "Mining association rules between sets of items in large databases," *SIGMOD*, pp. 402–411, 1993.

[21] B. Liu, W. Hsu, Y. Ma, "Mining association rules with multiple minimum supports," in *SIGKDD International Conference on Knowledge Discovery & Data Mining*, ACM, August 15-18 1999.

[22] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *20th International Conference on Very Large Data Bases*, VLDB '94, (San Francisco, CA, USA), pp. 487– 499, Morgan Kaufmann Publishers Inc., 1994.

[23] P. P. Tardan, A. Erwin, K. I. Eng, and W. Muliady, "Automatic text summarization based on semantic analysis approach for documents in bahasa indonesia," in *Under Review for International Conference of Information Technology and Electrical Engineering*, 2013.

Ontology and Semantic Matching for Diabetic Food Recommendations

Achmad Arwan
Department of Informatics
University of Brawijaya
Malang, East Java, Indonesia
arwan@ub.ac.id

Bayu Priyambadha
Department of Informatics
University of Brawijaya
Malang, East Java, Indonesia
bayu_priyambadha@ub.ac.id

Riyanarto Sarno
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, East Java, Indonesia
riyanarto@if.its.ac.id

Mohamad Sidiq
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, East Java, Indonesia
sidiq12@mhs.if.its.ac.id

Heri Kristianto
Department of Nurses & Medical Surgery
University of Brawijaya
heri.kristianto@ub.ac.id

***Abstract*—Foods recommendation for diabetes patients is indispensable for controlling blood sugar levels. Currently, the foods preparation is done by a nutrition expert. The patient's dependence on the nutrition experts is very high, thus the selection of foods could not be done independently. The Automation system to determine foods combination for diabetic patients is needed to solve these problems. In this study, the automation system has been designed and implemented. The technologies used in this research are the OWL and SWRL. There are few researches that explore an automation process of foods recommendation for diabetes patients using the technology of OWL and SWRL. Domain knowledge based on Ontology is needed to process foods composition automatically. However, using SWRL and OWL technology is not enough, because the accuracy of the words required. A semantic ontology understanding was added using weighted tree similarity method to specify the composition of foods for diabetic patients. 73% data were able to be correctly predicted by this method.**

Keywords*— *Diabetes Mellitus, Diet, Ontology, Foods Recomendation

I. INTRODUCTION

One of type of diabetes that has quick development process is type 2 diabetes [1]. Type 2 diabetes or Non-Insulin Dependent Diabetes Mellitus (NIDDM) can strike anyone regardless of age, generally the more attacking those who have overweight or obese. The early symptoms of type 2 diabetes are difficult to detect. This is indicated by one of three people who have type 2 diabetes do not know they have the disease. Type 2 diabetes is a chronic disease that occurs because the pancreas is unable to produce enough insulin (the hormone that regulates blood sugar) or when the body cannot effectively use the insulin that is produced. This condition can lead to high blood sugar levels and can cause serious damage to many body systems, especially the nerves and blood vessels. Diabetes can lead to serious complications such as heart disease (cardiovascular disease), neurological disease (diabetic neuropathy), kidney failure (diabetic nephropathy), and eye disease (diabetic retinopathy). For people whom suffer from type 2 diabetes controlling the food consumed is the important thing to do. Because, by controlling their food, patient of type 2 diabetes can control the condition of sugar level in their body.

Dietary control in diabetes patients is a very important. Nutritional foods is very diverse. Lack of information on nutritional diet on diabetic patients can make the condition of patients get worse. Diet in diabetes patients can be used as an attempt to control sugar levels in the blood.

Today, patient of diabetes consult about foods nutrition only with expert nutrition. Patients often experience of rising sugar levels suddenly only because of not knowing the levels of nutrients that ate. For non-medical patient, nutrition level is an information that they do not see in daily life. With this condition, there are needed some actions to make all patient have a knowledge about food and nutrition. To solve this problem, it is required a tool that make a patients can decide their menu by them self, and it will decrease the dependencies of patient of diabetes on the nutrition expert.

An automatic approach is one method that will propose in this research. This study proposed the construction of ontology-based foods recommendations for type 2 diabetes patients using ontology and semantic matching. Several researchers have conducted research on ontology to set food menu. Ontology is built as a domain knowledge about food menu which previous presented on [2]. The mechanism of that system is similar to knowledge-based system. There are knowledge and condition that will be queried into the knowledge. And then system will show the diet menu for patients. In this research, proposed some enhancement at the existed approach to increase the sense of semantic in the system. Some semantic search method will be adopted.

II. RELATED WORK

A. Previous Method

Several researchers have conducted research related to the determination of foods menu for diabetic patients. The purpose of some previous researches are to find a method that automatically determine the foods menu.Nutritional therapy is essential to avoid, control and slow down the process of complications in diabetic patients [3]. In such research, has been declared a method in the preparation of the ontology automatically with hierarchical clustering approach.

Automation determination system has been built using ontology and combination ontology with fuzzy algorithm known as fuzzy ontology [4]. The focus of those research was in diabetes cases [5][6][7] and the rest is in document searching [8]. There are uncertain a data value not precision data on a domain of knowledge. Fuzzy algorithm was used to solve the uncertainty.

Control system construction of foods for diabetic patients was conducted in [9]. In this study, there is a process of reasoning within ontologies. The ontologies which used in the study are foods ontology and menu ontology.

B. Analysis of Previous Method

Several researchers were conducted in several approach of constructing menu for diabetes patients. Some automation methods have been studied using ontology to represent the knowledge. There were some data that has been used for testing phase. Data were taken from several real diabetes patients. Then, the data was used as an object query to the ontology of domain knowledge. The result is shown as the arrangement of menu for diabetes patient.

The problems were found in previous approaches. System that uses the ontology as domain knowledge will be functioned if data is exact and contained in ontology. If the data is not contained by ontology, the system will not show the result. In the other side, there are possibilities that the data is not available in the ontology, but the system must be show out the result. The biggest question is how to solve this problem. There is a method called Weighted Tree Similarity that used to count degree of similarity two tree data [10]. In [11], the method is used to count degree of similarity of agents in e-business environment. Agents are represented as node-labeled tree, arc-labeled tree, or arc-weighted tree.

This research will propose a method to construct diabetes menu using ontology and semantic matching among ontologies using weighted tree similarity. The method and the direction of solving will be explained further.

III. RELATED THEORY

A. Daily Calorie Needs

To determine the nutritional recommendations on diabetes mellitus (DM) patient required daily calorie needs calculated in accordance with condition of the patients and the type of their activities [2]. Daily calorie needs is calculate by finding the desired weight using the Body Mass Index (BMI).

$$BMI = \frac{mass(kg)}{height(m)^2} \quad (1)$$

Then the calculation of BMI is used to classify whether patients fall into skinny, normal, or fat category.

The activities of the patient are also a supporting factor in determining daily calorie needs [2]. There are 3 types of daily activities of patients which are light, medium& heavy. The following types of activities the patient based on the work is shown on Table 1.

TABELI.
THE TYPE OF PATIENT ACTIVITIES

Light	Medium	Heavy
Clerk's Office	Student	Sailors
Store clerk	Light industrial employees	Labour
Teacher	Housewife	Dancers

Next is the metric table calories needs per kg according to the BMI calculation type of patient activities [2].

TABELII.
MATRIX NEEDS CALORIES PER KG IDEAL BODY WEIGHT

	Activity		
BMI	***Light***	***Medium***	***Heavy***
Obesity	25	30	35
Normal	30	35	40
Underweight	35	40	40-50

The calorie intake per day can be described in the following formula.

$$Calory = mass(kg) * matrix \quad (2)$$

To find out the ideal weight, the following formula is used.

$$W = (heigth - 100) - 10\% \quad (3)$$

For example patient with a weight of 100 kg and a height of 170 cm with the job as a teacher. The body mass index can be counted using the aforementioned formula as follow 100 / $(1.7)^2$ = 34. In the matrix of BMI, 34 are classified as obesity (over weight). To count ideal weight, we can count using formula 3, so the ideal weight can obtain as 63 kg. Using formula 3, there are some conditions, for patient that has gender as a man and has height greater than 160 has to decrease by 10% of height. If the patient has height less than or equal to 160 the calculation is be done without using decrease by 10% of height. On the other hand, female patients and has height greater than 150, it has to decrease by 10% of height, and otherwise if less than or equal to 150.

To count the ideal amount of calories consumed in a day, the number of calories per kg body weight multiplied by ideal body weight [2]. To find out the number of calories per kilogram of body weight, we can refer to the matrix needs calories per kg of body weight by combining BMI status and activity level of the patient. So we can obtain 25 calories,

because the example is a teacher and has the nutritional status of obesity. Occupation as teacher falls into the category of light activity. And to calculate calorie needs for a day, then the number of calories per kg ideal body weight multiplied by ideal weight, therefore the example patient is 1575 calories.

B. Ontology

The ontology is used to describe a concept that has the characteristic property & attribute [12]. Ontology consists of elements, attributes, class and object [13]. OWL is Web Ontology Language or a language for defining web ontology. Meanwhile, ontology is the science that is used to describe a variety of entities that exist in the world and how to represent them related [11][12]. OWL is built based on RDF schema. OWL and RDF has the same basic function: defines the classes, properties, and relationships. Nevertheless more than capabilities offer OWL RDF in defining complex relationships such as subclass, restriction, disjoint, constraint, and cardinality.

IV. METHODOLOGY

The methods used in this study are develops of foods ontology, developing of foods calorie ontology, developing rule and search foods menu. All the steps are figured in Fig. 1.

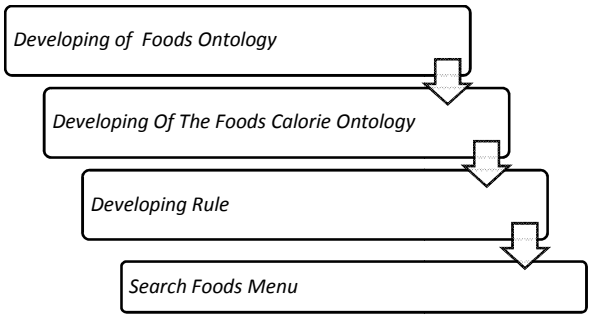


Fig.1.Research framework

A. Develops of Foods Ontology

Develops foods ontology is the most basic step in this research. Foods ontology is a an overview of knowledge about foods that includes of concepts, relations, attributes and individual foods. Knowledge of foods is reflected in the foods ontology which is a set of core knowledge in the search menu list for diabetes patient.

Knowledge of foods taken from some reference of foods and enriched with the information from nutrition expert [2]. All information about food reserved only for diabetic patients. The foods ontology is represented in Fig. 2.

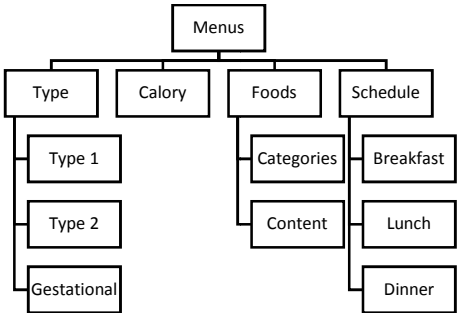


Fig.2.Foods Concept Hierarchy

B. Develops Of The Foods Calorie Ontology

Calories foods ontology is a knowledge representation that consists of information of diabetic patients and the number of calories of foods a day. Calorie foods ontology is spread from 1300 calories to 2300 calories. Hierarchical structure of the foods calorie ontology is depicted in Fig.3. This ontology will be used to process the matching of ontology of patients which is entered as the input system. The matching process occurs at the stage of search foods menu.

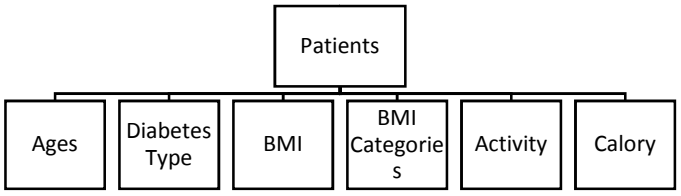


Fig. 3.Patients calorie needs ontology hierarchy

C. Develops the Rule

Rule in the ontology of serves as an inference engine. Rule served to do a classification or grouping the data categories in the ontology. In this study the language used as the definition of SWRL rule ontology.

SWRL is a combination of OWL-DL with RuleML [6] which was modelled in a Horn-claouse. Horn-clause is a representation of a conditional if-then that is more familiar. SWRL rule can be either C (x), P (x, y) SameAs (x, y) where C is an OWL description, P is an OWL property and x, y are the individuals or the data on the OWL. Examples of SWRL rules are as follows:

hasParent (?x1,?x2) ∧ hasBrother (?x2,?x3) ⇒ hasUncle (?x1,?x3)

Based on SWRL above can be interpreted as the following, x1 hasParent x2 and x2 hasBrother x3 then x1 hasUncle x3.

In this research the rule was made to classify the data automatically. Datas are grouped based on the food content. Some of the rules used are described as follows:

Food(?f), contain(?f, Vitamins_and_Minerals), hasKind(?f, "Fruit") -> Fruits(?f)

Food(?f), contain(?f, Vegetable_Protein) -> Vegetable_Protein_Foods(?f)

Food(?f), contain(?f, ?n) -> contained(?n, ?f)

Food(?f), hasKind(?f, "Milk") -> Milk(?f)

Food(?f), contain(?f, Animal_Protein) -> Animal_Protein_Foods(?f)

Food(?f), contain(?f, Fiber), contain(?f, Vitamins_and_Minerals) -> Vegetable_Foods(?f)

Food(?f), contain(?f, Carbohydrate) -> Carbohydrate_foods(?f)

D. *Search Foods Menu*

Search foods menu is done in several stages. The stages are receiving input ontology of patients, matching patient ontology with calorie foods ontology, and the final stage in the process of using SPARQL queries on foods ontology. The search flows of process illustrated in Fig.5.

To match the ontology we use Weighted Tree Similarity method [10]. Matching is performed on every leaf of the tree by considering the weight of each arc. Fig.4 shows a sample tree calculated similarity.

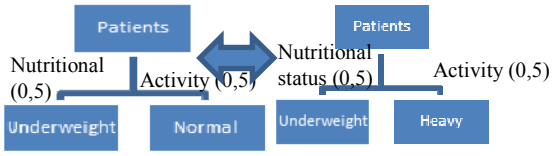


Fig.4. Example of Similarity Tree

Calculation of similarity of the tree above are:

Sim (nutrition) = 1 (0.5 + 0.5)/2 = 0.5
Sim (activity) = 0 (0.5 + 0.5)/2 = 0
Total = 0.5 + 0 = 0.5

Results calculation of the similarity of both the trees is 0.5. This method is used in the process of matching of ontology which is compare the patient's data and calorie. Both data is represented in tree. The highest degree of similarity resulted by this process will be chosen.

The next process is query the matching result to food ontology using SPARQL. The result will be showed the composition of food for diabetes patient. Fig. 5 will be described the flow of matching and searching of menu.

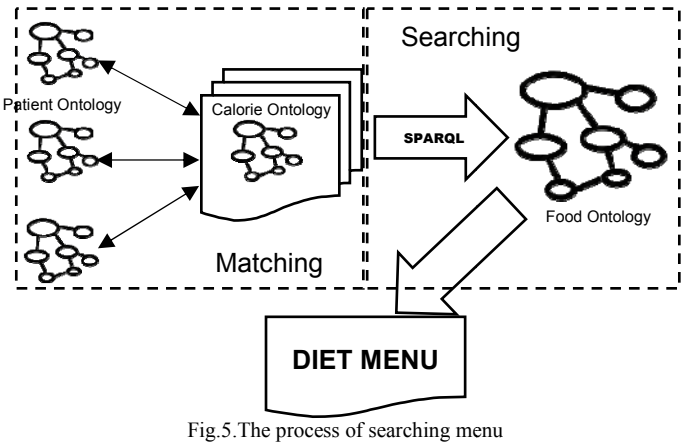


Fig.5.The process of searching menu

V. System Architecture

The proposed method will be implemented using several technology and tools. The input system or end user interface is implemented using Java Swing technology. In this layer, system is displayed as a desktop-based application. The second layer is the controller layer that act as a processor of the system. Several libraries are used in this layer. Jena and OWL API and JENA are used in this processor layer, beside java object oriented was conducted here. All of the tree or ontology in matching process is represented as an object. The last layer is the layer of data. This means that this layer act as storage of data. Data is described as an ontology that represents the domain knowledge. Ontology is built using tool named Protégé. Ontology contain of inference rule named SWRL. All of rules that define as SWRL was run or interpreted using inference engine. In this study, we employ the Pellet for reasoning engine. Pellet was equipped in Protégé, so we can easily use that feature. All of this description is represented in Fig. 6.

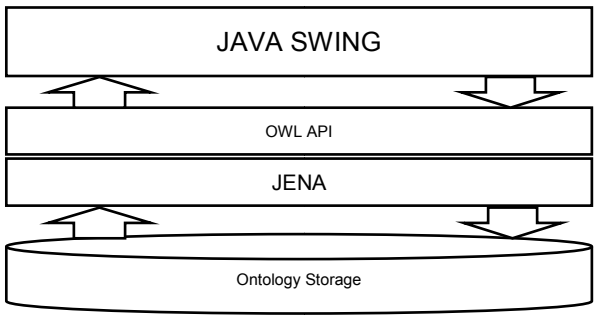


Fig. 6. System architecture

VI. Experiment

In the experiment of the search foods menu, the data used is the patient data including age, diabetes type, height, weight, nutritional status, activity and calories. Where in the process are the initial calculations prior to matching on ontology. Initial calculation does is to calculate the calorie and nutritional status, resulting in a system that was built, the values that should be included are age, diabetes type, height, and weight. The main display system is shown in Fig. 7.

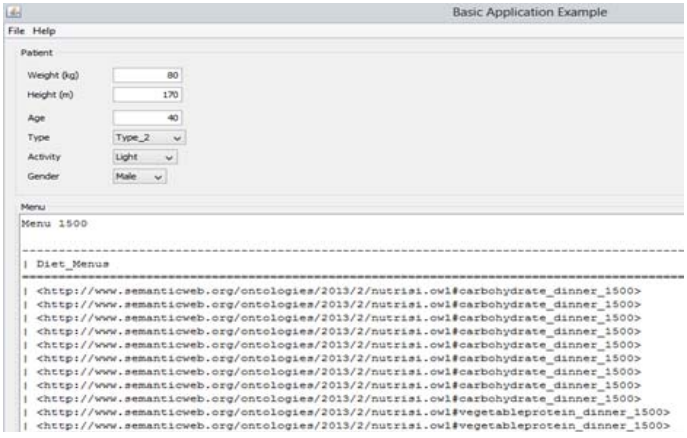


Fig.7. Main display system

The next process is the ontology matching of patients and calories foods. Ontology matching process is illustrated in Fig.8.

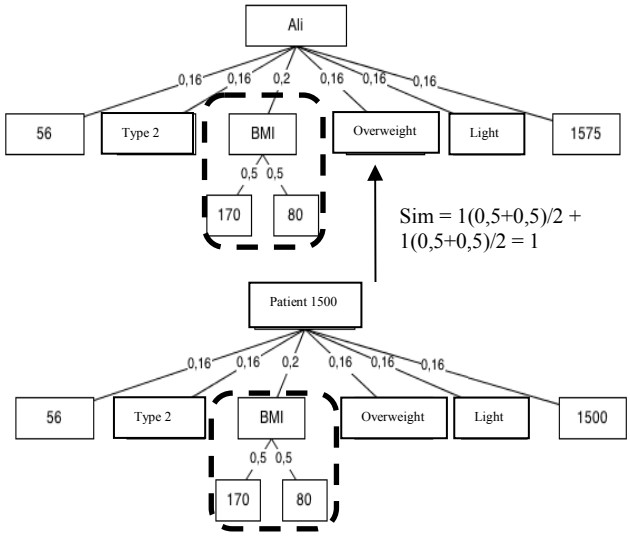


Fig.8. The first step ofsimilarity matching for menu

Fig.8 shows the description of the calculation process, where matching is done on the bottom branches. Explanation of the image above the calculation of similarity from the lowest branch is as follows:

$$\text{Sim} = 0.5\ (0.5 + 0.5) / 2 + 0.5\ (0.5 + 0.5) / 2 = 0.5$$

From the above calculation, obtained in common from the lowest node is 0.5. A value of 0.5 will put a value on it.

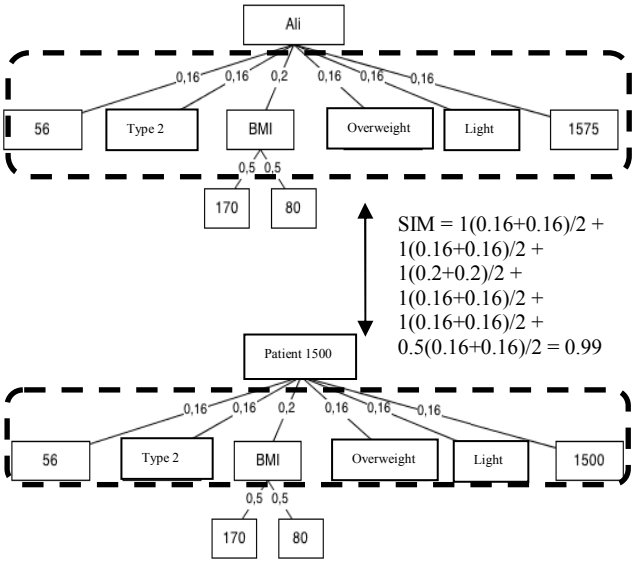


Fig.9. Second step of similarity matching for menu 1500

Fig.9 shows the continued process of calculation of matches from the points in the top level. From calculation the same way on the level below, then get that value matches from the patient's ontology and 1500 calories ontology was 0.99. There is likely value in common will be greater on the ontology of the other menus. Therefore the search process will be continued on menu ontology. And in the end found that, in the case of the example patient, a suitable menu for him is a menu with 1500 calories.

After the initial phase of ontology matching process is complete, then resume the search process the menu tree in the foods ontology. An overview of the foods ontology is reflected in Fig.10.

Fig.10. The foods ontology

The finding process of foods is done by running SPARQL in the ontology. The ontology has been equipped with reasoning that is defined using SWRL.

VII. RESULT

This section presents the results from the experiments. Ontology patient has 30 instances on diabetic patient ontology with vary conditions. Then through the user interface we have tested 30 patients with diabetes were randomized to look for individuals who have the highest level of similarity with it. Table III show the confusion matrix of this experiment.

TABELIII.
THE CONFUSION MATRIX OF EXPERIMENT

		Data Actual		Total
		True	***False***	
Prediction	***True***	(TP) 22	(TN) 0	(P') 22
	False	(FP) 8	(FN) 0	(N') 8
Total		(P) 30	(N) 0	30

Formula to calculate the level of accuracy of the data can be seen in the following formula.

$$ACC = \frac{TP+FN}{P+N} \quad (4)$$

Where, TP (True Positives) is the data that has true result and predicted as a true. TN (True negative) is data that is false but predicted as a true. FP (False positive) is the data is a true but predicted as false. And, FN (False negative) is data that is false and predicted as false. While P is the number of actual data of true, and N is the number of the actual false. So, it can be concluded that the accuracy of the defect detection system on this research is 0.733.

VIII. CONCLUSION

The determination of the list of menu foods for diabetic patients is crucial. List of foods menu based on the number of calories per day could help patients with diabetes to control their blood sugar levels. Setting the menu list can be done by building a foods ontology.

Search foods by using ontology can be done based on the context. Unlike the search in a relational database, if the data that is searched for is not found in the database, it will not generate an output. Semantic search method is used to calculate similaritas of data, so data doesn't have to be exactly same with the existing data. With the combination of some data then the context search will be formed on the method of semantic search.

Experiment results show that within 30 data sample of diabetic patients, system is able to recommend 22 data food recommendation (73%). It means that 73% of the data are correctly predicted. Further research will be focused on the methods of calculation of similaritas and accuracy improvement. Model calculations can be improved to make it more effective and efficient. Model ontology can also be improved, so that the process can be done quicker and more accurate.

REFERENCES

[1] G. Hu and J. Tuomilehto."Lifestyle and outcome among patients with type 2 diabetes" , Int. Congr. Series, vol. 1303. Aug. 2007. pp. 160–171.

[2] Sarwono .W,Gatut .S, Kartini .S,Rochamah .M, "*Cara mudah mengatur makanan sehari-hari seimbang & sesuai kebutuhan gizi*". FK University of Indonesia. 2004. ISBN 979-496-314-3.

[3] Li, H.C., Ko, W.M. "Automated Foods Ontology Construction Mechanism For Diabetes Diet Care". *IEEEMachine Learning and Cybernetics*. 2007.

[4] Lee, C.S., Wang, M.H., Hagras, H. "A Type-2 Fuzzy Ontology and Its Applicationto Personal Diabetic-Diet Recommendation". IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 18, NO. 2. 2010.pp. 374 – 395.

[5] Pramono, D., Setiawan, N.Y., Sarno, R. & Sidiq, M."Physical Activity Recommendation for Diabetic Patients Based On Ontology".*The 7th International Conference on Information & Communication Technology and Systems (ICTS) 2013*. May 15-16, 2013. Bali, Indonesia.

[6] Dewandono, R.D., Fauzan, R., Sarno, R. & Sidiq, M. "Ontology and Process Mining For Diabetic Medical Treatment Sequencing".*The 7th International Conference on Information & Communication Technology and Systems (ICTS) 2013*. May 15-16, 2013. Bali, Indonesia.

[7] Saadah, U., Sarno, R., and Yuhana, U.L. "Latent Semantic Analysis and Weighted Tree Similarity for Semantic Search in Digital Library". *The 7th International Conference on Information & Communication Technology and Systems (ICTS) 2013*. May 15-16, 2013. Bali, Indonesia.

[8] Sarno, R., Ghozali, K., Nugroho, B. A. & Hijriani, A. "Semantic Matchmaking using Weighted Directed Acyclic Graph". Surabaya, Institut Teknologi Sepuluh Nopember (ITS). 2011. pp. 1-6.

[9] Cantais, J., Dominguez, D., Gigante, V. "An Example of Foods Ontology for Diabetes Control". *Proceedings of the International Semantic Web Conference 2005 workshop on Ontology Patterns for the Semantic Web*. 2005.

[10] Sarno, R., Rahutomo, F. "Penerapan Algoritma Weighted Tree Similarity Untuk Pencarian Semantik". JUTI. Vol. 7, No. 1. 2008. Pp. 35-42.

[11] Sarno, R., Anistyasari, Y., and Fitri, R., "Semantic Search", Andi Publisher, 2012, ISBN 978-979-29-3110-5.

[12] Kuan M. "Using SWRL and OWL DL to Develop an Inference System for Course Scheduling". Master Thesis, Chung Yuan Christian University, Taiwan, R.O.C. 2004.

[13] Hebeler, J., Fisher, M., Blaze, R., Perez-Lopez, A. "Semantic Web Programming". Wiley Publishing Inc. 2009. Canada.

Predicting Latent Attributes of Twitter User by Employing Lexical Features

Elisafina Siswanto
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
elisafina.siswanto@gmail.com

Masayu Leylia Khodra
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
masayu@stei.itb.ac.id

***Abstract*—The rapid growth of social media, especially Twitter in Indonesia, has produced a large amount of user generated texts in the form of tweets. Since Twitter only provides the name and location of its users, we develop a classification system that predicts latent attributes of Twitter user based on his tweets. Latent attribute is an attribute that is not stated directly. Our system predicts age and job attributes of Twitter users that use Indonesian language. Classification model is developed by employing lexical features and three learning algorithms (Naïve Bayes, SVM, and Random Forest). Based on the experimental results, it can be concluded that the SVM method produces the best accuracy for balanced data.**

Keywords—age, job, classification, Twitter, lexical, machine learning.

I. INTRODUCTION

The rapid growth of social networking users, such as Twitter, in recent years has produced a very large amount of user-generated texts. In 2012, Twitter users were recorded at 500 million people, and the average number of tweets in one day was 55 million tweets [1]. This led to the growing of research within the field of information retrieval or text mining using these tweets to determine user personality (introvert, extrovert, etc.) [2] or user role in an event [3], to create demographic user profiles [4], etc. Information about the profiles of social networking users is very useful for the purpose of marketing, personalization, law enforcement, and others. The problem is Twitter only provides the name and location of its users and these information are not enough.

Our system predicts latent attributes of Twitter user based on his tweets. Latent attribute is an attribute that is not stated directly. In previous research, Rao [4] predicted several latent user attributes: gender, age, regional origin, and political orientation based on English-language tweets. In this paper, we predict age and job attributes of Twitter users based on the tweets that using Indonesian language. We specialize in researching tweets in Indonesian language because Indonesia is the fifth most Twitter users around the world. Another reason is the researches regarding the latent job attribute or Indonesian language tweets are still very rare.

Age and job attributes represent economic conditions and the level of user maturity and intelligence. It can be used for a variety of needs, for example, marketing. Both attributes could determine what goods that the user usually likes, or what the price of goods suitable for the user.

Unlike news, tweets contain non-standard language, many words are come from everyday language or regional language, a lot of abbreviations, and so forth. We employ preprocessing methods used in Indonesian Twitter sentiment analysis research [5]. Our classification model is built by extracting lexical features [4] and various methods of machine learning using WEKA tools [6].

The rest of this paper is organized as follows. In Section 2, related works about research in social media, especially analyzing Twitter is presented. Section 3 explains about our method, and Section 4-6 describes our data, preprocessing, and experiment. Concluding remarks are presented in Section 7.

II. RELATED WORKS

Most of the researches on Twitter can be divided into two major groups. The first group emphasizes the ways users interact. How to interact concepts are closely related to the number of followers and following, retweet, and mention. One of the studies conducted in this group is to determine the personality of the user based on the number of Twitter interactions [2]. Another study regarding this first group is how to calculate the influence of a Twitter user against other users [7] [8].

The second group more emphasis on the content of user posted tweets. This second group can also be divided into research that more emphasis on the content of a tweet or emphasize on all the tweets from a user. In the first type, a tweet would normally be considered as an instance. The content of the tweets are extremely diverse. Users can freely write down their feelings, weather conditions, traffic conditions, natural disasters, selling products, etc. The general research of this type was to extract important information from a collection of tweets that is written by a different user. As an example, the occurrence of earthquakes was detected in real time based on the user tweets [9]. Another research extracts information about the results of sports events such as live scoring [10]. Besides information extraction, classification can be employed such as for sentiment analysis of tweet [11] [12].

Another type of research is a study that considers all the tweets from a user as a single entity. Many researchers make

predictions to the latent attributes of a user. Commonly performed research is to predict the gender of the user [13] [4] and user's politics orientation (democrat or republican) [14] [4].

Research on Indonesian language tweets is rarely done. One of them is research that conducts sentiment analysis of Indonesian language tweets [5] [15]. Other researches extract traffic information [16] or online transaction [17]. Only one research predicted latent attribute gender of Twitter users [18].

III. METHOD

Researches to predict the latent attributes of a user are generally more emphasis on the words in the tweet, not the meaning of the tweet. It was found that each class would have a different style of word selection. For example, the words "love" and "cute" are words those are widely used by the female users [13]. We also found that each category of a class would have certain characteristics when viewed from the words used (see Table I).

TABLE I. TWEETS DATA EXAMPLE

Class	Tweets	
	Category	**Tweets sample**
Age	Under 20 years old	"Wkwkw, wes" adek kelas hoyal -_- SMA mana tujuan?? @dandann19"
		"Jektas tangi, padahal pengen melu ngibar bendera nang upacara mau"
		"Kamu tau tidak ?? Kapan saatnya aku mencintaimu dan aku membencimu"
	Over 20 years old	"Woo, akhirnya nongol lagi lagu indo yang beatnya enakeun. Walau liriknya rada2 absurd. --a"
		"@infoHendri ndak ada akses hen, satu2nya akses lewat network kampus, ada sih cara tricky, paper sm tapi format lain"
		"@jaunreza : Hampir 99% suami lebih sayang anak daripada istri. Karena anak adalah "Darah Daging""
Job	Student	"@Tarrajunita nek nilai'q sing neng ngisor kkn kimia, kkpi, ips, kwu, tpi kuwi ngenteni di rata" sek, dadi perlu remidi po orak..."
		"Buaah baca tweetnya bikin ngiri pengen jadi mahasiswa sana.."
		"25 Feb Seminar Proposal !!!! >:)"
	Employee	"overan dinas pagi dg @LikePrawirAkbar hari Libur tetep kerja semangat !!!!!!"
		"mau pulang kantor aja susah :("
		"Kangen gajiaaaan --"

Suppose a user who has a job as a student will use a lot of words that relate to school (such as "guru", name of the subject, "tugas", "pr", etc.) or college (like "mahasiswa", "proposal", "skripsi", "dosen", etc.). Likewise, employee category usually uses many words that relate to the job, office, work processes, and many more.

Age class also has the same related words. The words for under 20 years category are related to the school and love relationship, whereas over 20 years category frequently uses words with unfamiliar terms or words that relate to the family.

In this paper, we focused to perform the classification process at the lexical level. That means we build our classification model by employing common words of each category for both classes. Therefore, we split the tweets into a vector of word features. Then for each word, we will calculate the weight of its occurrences. These feature vectors will become the input for machine learning methods.

Before each tweet converted into a collection of words, we add some preprocessing process that has been adapted to the characteristics of the tweets in Indonesian language. Our preprocessing module is based on preprocessing used in Indonesian language Twitter sentiment analysis research [5]. These preprocessing processes will be further described in Section V.

IV. DATA

In constructing the tweets corpus, we have distributed a questionnaire where users will be asked to fill in the data his Twitter username, age (under or over 20 years) and job (student, employee, or other job). After data are collected, we use Twitter API [19] to retrieve all tweets from that users. Data tweets that can be retrieved are only tweets from accounts that are not locked. Tweets data would then be cleaned using a variety of preprocessing processes. Then we merge all of the tweets into one big tweet.

Our tweets corpus consists of all tweets from about 300 users. Maximal tweets data that can be extracted from a user, according to Twitter API is 3200 most-recent tweets. Tweets data that used for this research is only the tweet that is written in Indonesian language.

Categories for age are under 20 years old and over 20 years old. Categories for job data are a student or employee. Detailed distribution of data for both categories could be seen in Table II. In this paper, we consider an instance will represent a user and each instance will contain all of the user's tweets (up to 3200 tweets).

TABLE II. DETAIL USER DATA

Class	Number of users	
Age	Under 20 years old	165
	Over 20 years old	159
Job	Student	234
	Employee	72

V. PREPROCESSING

Before using WEKA tools, tweets that have been collected will be entered in preprocessing steps. We use preprocessing methods used in Indonesian language Twitter sentiment analysis research [5] and modifying several processes, such as adding "eliminate duplicate letter" and "ARFF file builder" processes and eliminate "convert number", "convert word", and "convert negation" processes. The processes performed for the tweets data are:

- Retweet removal: Retweet generally is a user doing a repost or quoting someone else's tweet. Retweet does not come from the users themselves and can cause data to be invalid, then retweet should be discarded. Retweet can be detected by the keyword "RT" or the words that are in quotation marks.
- Mention and link removal: On Twitter, mention used to tag other users within a tweet using the "@username" format. Mentions and links do not effect in the determination of the user attributes, because this paper emphasized the determination of the user attributes based on the language style that he used, not by how he interacted to another user.
- Duplicate letter removal: Twitter users generally use non-standard language, so there will be many typing errors. One of such errors is to extend the word by duplicating the letters, like "assssiiiiiikkkk bangeeeeet". Eliminating duplication of the letters is conducted by eliminating the letters within a word that is the same as the previous letters, so the example above will be shortened to "asik banget".
- Number removal: Numbers in predicting user attributes do not have a significant influence, for example, the nominal rate of the money "10 thousand", those numbers were not associated with a specific attribute. Therefore, every occurrence of numbers in tweets data will be removed.
- Stop word removal: What is meant by stop word are words that have no meaning, such as the words "ini", "itu", "dengan", etc. Stop word removal needs to be done because such words are generally words that have a high frequency of occurrence, so it is feared if these words are not removed would make the process of attribute classification, especially during learning process (the calculation of weights, etc.) will be disturbed. We use stop word list for standard Indonesian language and added by abbreviations of the words that commonly used in everyday life.
- Convert emoticon: Twitter users usually insert emoticons into their tweet. Emoticons (emotion icons) are one of the ways of expressing feelings textually. In this paper, we transform the appearance of emoticons into a more general form, given that there are so many ways to write emoticons. We use the modified emoticon's table converter as based on Sunni [5] (see table III).

TABLE III. EMOTICON CONVERSION

Conversion	Emotion
Emohappy	>:] :-) :) :o) :] :3 :c) :> =] 8) =) :} \(^.^)/ \(?)/ (~^.^)~ ^_^ ^^ ^^ :^)
Emolaugh	>:D :-D :D :)) =)) 8-D) 8D x-D xD =-D =D =-3 =3
Emosad	>:[:-(:(:-c) :c :-< :< :-[:[:{ >.> <.< >.<
Emohorror	D:< D: D8 D;) D= DX v.v D-':
Emotongue	>:P :-P :P X-P =P :-b :b
Emoshock	>:o :-o :o o_O o.O 8-o
Emoannoyed	>:\\ >:/ :-/ :-. :/ :\\ =/ =\\ :S
Emostraightface	:| ._. -_-" -_- :-|

- Punctuation Removal: after emoticons converted into text form as described in the previous point, then all punctuation marks that are still present in the tweets will be eliminated.
- Convert to lowercase: all tweets would then be converted to lowercase. This is done so the words that are written in the beginning of a sentence or in capital letters would have the same meaning as the words that written in small letters.
- ARFF file builder: After each tweet through the preprocessing processes, then all the tweets of a user will be combined into a single instance and written in a file with ARFF file extension. This file then will be used in the experiment stage using WEKA.

VI. EXPERIMENT

Our experiment would perform using WEKA tools. WEKA (Waikato Environment for Knowledge Analysis) is a collection of machine learning algorithm written in Java language. WEKA supports several data mining tasks, such as data preprocessing, clustering, classification, and many more [6].

The first thing to do is to change tweets that have a string data type to vector of words using the WEKA data-preprocessing feature. We use the word vector with size of 20 thousand different words. Comparison between training data and testing data that we use is 2:1.

Several methods of machine learning were employed to build the classification model. The methods are Naïve Bayes, Support Vector Machine and Random Forest. We compare the three methods to determine which method is the best performer in the classification process.

A. Age Classification

As described in the previous section, the data used for the age classification is 324 data (165 data are under 20 years old and 159 data are over 20 years old), where 110 data will be used as testing data. Table IV illustrates the results of experiments that we have.

TABLE IV. EXPERIMENT RESULT ON AGE ATTRIBUTE

Machine Learning	Accuracy
Naïve Bayes	66.3636 %
SVM (nu-SVC)	**77.2727 %**
Random Forest	70.9091 %

From the results in the table above it can be concluded that for the age classification, SVM method would produce the highest accuracy value. The results, especially by using SVM method, have produced slightly better accuracy when compared to the results of Rao [4]. Accuracy obtained by Rao [4] is 72.33% using the n-gram feature (unigram and bigram) and socio-linguistics.

B. Job Classification

Data for the job classification consists of 72 employees data and 234 students data with a total of 306 data. Of the total, 104 data are then retrieved from the data to be used as testing data. Table V illustrates the results of experiments that we have got for the job classification.

TABLE V. EXPERIMENT RESULT ON JOB ATTRIBUTE

Machine Learning	Accuracy
Naïve Bayes	70.1923 %
SVM (C-SVC)	71.1538 %
Random Forest	**73.0769 %**

From the table above, it can be seen that the Random Forest method would produce the highest accuracy. Unlike the previous age classification experiment, the classification of this work has imbalanced data. As a result, although the accuracy of the resulting looked quite high, but the accuracy was resulted from an imbalanced Confusion Matrix. Suppose in the SVM method, 104 of test data (30 employees and 74 students), all the data was predicted as a student (see table VIII).

There are several ways to handle the problem of classification of imbalanced data. One way to do it is using undersampling, which means to remove some of the majority data so that the new data that is generated would be balanced [20]. In this paper, we use the simple undersampling process by just taking 72 first data from the student data. Table VI illustrates the results of the data distribution after undersampling process.

TABLE VI. DETAIL USER DATA

Category	Detail Data	
Imbalanced	Student	234
	Employee	72
Balanced	Student	72
	Employee	72

By using data that have been balanced, we tried to re-classification. Table VII illustrates the results from the job classification using balanced data.

TABLE VII. EXPERIMENT RESULT ON JOB ATTRIBUTE (BALANCED)

Machine Learning	Accuracy
Naïve Bayes	67.3469 %
SVM (Nu-SVC)	**75.5102%**
Random Forest	67.3469%

From the data table above, it can be seen that the highest accuracy is obtained by the SVM method. If we considered of detail accuracy information, experiments using balanced data will generate the value of Confusion Matrix, Precision, Recall, and F-Measure better (See Table VIII).

TABLE VIII. DETAILED RESULT FOR BALANCED AND IMBALANCED DATA (USING SVM METHOD)

Machine Learning	Detailed Result
Imbalanced Data	=== Confusion Matrix === a b ← classified as 0 30 | a = Employee 0 74 | b = Student === Detailed Accuracy === Precision Recall F-Measure Class 0.000 0.000 0.000 Employee 0.712 1.000 0.831 Student 0.506 0.712 0.592 Weighted Avg.
Balanced Data	=== Confusion Matrix === a b ← classified as 15 10 | a = Employee 2 22 | b = Student === Detailed Accuracy === Precision Recall F-Measure Class 0.882 0.600 0.714 Employee 0.668 0.917 0.786 Student 0.787 0.755 0.749 Weighted Avg.

VII. CONCLUSION AND FUTURE WORKS

In this paper, we conduct a research to classified latent attributes age and job of a Twitter user based on the user's tweets data. The conclusions of this research are:

1. Data tweets especially from Indonesian users are written in non-standard Indonesian language. Using uptake words from the everyday language or regional language are common in tweets writing. In addition, many words are written using the abbreviation. This is one of the challenges in conducting research especially on the tweets data that using Indonesian language.
2. To overcome the point one problems, we perform preprocessing to reduce the occurrence of errors in the classification process.

3. The experimental results showed at the age and job classification, SVM method produces the highest accuracy values compared to other methods, especially for balanced data. We concluded that the SVM method is the most suitable method to be used for the classification with lexical features.
4. Imbalanced data also affects the accuracy of the classification results. Although the accuracy seems high, but when we perform the classification of minority's data, the resulting accuracy would be very small, even 0%.
5. To overcome our problem, undersampling process was employed on our imbalanced data. The result for the accuracy of the minority and the majority of the data is balanced. The accuracy of job classification that was balanced is smaller than the age classification, due to the reduced amount of training data and testing data is caused by undersampling process.

This research can still be further developed. Such as by adding some preprocessing processes to classify the words that have the same meaning but have different ways of writing. One way to do is to add the abbreviation dictionary, and the dictionary slang word commonly used by Twitter users who speak Indonesian.

How to handle imbalanced data can also be done better. Currently, we only use a simple undersampling process that took a number of first data. This can be improved by taking data that have more influence over the class, or by using other methods for handling imbalanced data.

Another improvement that can be done is to add predictions for a variety of other attributes, such as gender, marital status, preferred areas and others. The more complete predicted attributes that generated then the attributes can be more useful for a variety of fields; one of the fields is marketing.

VIII. REFERENCES

[1] (2013, March) Statistic Brain. [Online]. http://www.statisticbrain.com/twitter-statistics/

[2] Daniele Quercia, M. Kosinski, David Stillwell, and Jon Crowcroft, "Our Twitter Profiles, Our Selves: Predicting Personality with Twitter," in *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, Boston, MA, 2011, pp. 180 - 185.

[3] Munmun De Choudhury, Nikolas Diakopoulos, and Moor Naaman, "Unfolding the Event Landscape on Twitter: Classification and Exploration of User Categories," in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ACM, 2012, pp. 241-244.

[4] Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta, "Classifying Latent User Attributes in Twitter," in *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, ACM, 2010, pp. 37-44.

[5] Ismail Sunni and Dwi Hendratmo Widyantoro, "Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik," *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro dan Informatika*, vol. 1, no. 2, pp. 200-206, Juli 2012.

[6] University of Waikato. (2013, June) Weka 3: Data Mining Software in Java. [Online]. http://www.cs.waikato.ac.nz/ml/weka/

[7] Eytan Bakshy, Jake M. Hofman, and Winter A. : Watts, Duncan J. Mason, "Everyone's an Influencer: Quantifying Influence on Twitter," in *Proceedings of the fourth ACM International conference on Web Search and Data Mining*, 2011, pp. 65-74.

[8] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna P. Gummadi, "Measuring User Influence in Twitter: The Million Follower Fallacy.," *4th International AAAI Conference on Weblogs and Social Media*, vol. 14, no. 1, p. 8, 2010.

[9] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo, "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors.," in *Proceedings of the 19th International Conference on World Wide Web*, 2010, pp. 851-860.

[10] Liang-Chi Hsieh, Ching-Wei Lee, Tzu-Hsuan Chiu, and Winston Hsu, "Live Semantic Sport Highlight Detection Based on Analyzing Tweets of Twitter," in *Multimedia and Expo (ICME) IEEE International Conference*, 2012, pp. 949-954.

[11] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau, "Sentiment Analysis of Twitter Data," in *Proceedings of the Workshop on Languages in Social Media*, 2011, pp. 30-38.

[12] Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore, "Twitter Sentiment Analysis: The Good The Bad and The OMG," in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011, pp. 538-541.

[13] John D. Burger, John Henderson, George Kim, and Guido Zarrella, "Discriminating Gender on Twitter," in *Conference on Empirical Methods in Natural Language Processing*, 2011, pp. 1301-1309.

[14] Marco Pennacchiotti and Ana-Maria Popescu, "Democrats, Republicans and Starbucks Afficionados: User Classification in Twitter," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2011, pp. 430-438.

[15] Aqsath Rasyid Naradhipa and Ayu Purwarianti, "Sentiment classification for Indonesian message in social media," in *Cloud Computing and Social Networking (ICCCSN) International Conference*, 2012, pp. 1-5.

[16] Sri Krisna Endarnoto, Sonny Pradipta, Anto Satriyo Nugroho, and James Purnama, "Traffic Condition Information Extraction & Visualization from Social Media Twitter for Android Mobile Application," in *Electrical Engineering and Informatics (ICEEI) International Conference*, 2011, pp. 1-4.

[17] Masayu Leylia Khodra and Ayu Purwarianti, "Ekstraksi Informasi Transaksi Online pada Twitter," *Jurnal Cybermatika*, vol. 1, no. 1, 2013.

[18] Yudi Wibisono and Naufal Faruqi, "Penentuan Gender Otomatis Berdasarkan Isi Microblog Memanfaatkan Fitur Sosiolinguistik," *Jurnal Cybermatika*, vol. 1, no. 1, 2013.

[19] (2013, June) Twitter Developers. [Online]. https://dev.twitter.com/

[20] Zeping Yang and Daqi Gao, "An Active Under-Sampling Approach for Imbalanced Data Classification," *Computational Intelligence and Design (ISCID) Fifth International Symposium*, vol. 2, pp. 270-273, 2012.

Prediction of Reference Evapotranspiration with Missing Data in Thailand

Kitsuchart Pasupa
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, Thailand
Email: kitsuchart@it.kmitl.ac.th

Ek Thamwiwatthana
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, Thailand
Email: nupippo@gmail.com

***Abstract*—Artificial Neural Networks (ANNs) has been used in prediction of reference evapotranspiration for a recent decade. Its performance is competitive to a widely used method the so-called "Penman-Monteith" method. In this study, we aim to estimate the crop evapotranspiration by ANNs from climatic data in Thailand and compare the performance with the Penman–Monteith method. As missing data is inevitable, we also included the missing data situation into the study. This can be solved by expectation-maximization algorithm. The accuracy of the prediction decreases when the amount of missing values increases. Furthermore, we exploit the feature selection in the study. It shows that sunshine duration is the most important feature followed by temperature and wide speed, respectively.**

***Index Terms*—reference evapotranspiration; missing data; neural network; feature selection**

I. Introduction

Agriculture is an important mean of food production, and food is a basic necessity of human life. Nowadays, information and communication technology (ICT) begins to play a role in agriculture in order to reduce the cost and time of production. This is the so-called "precision farming". It can influence the agricultural market competitiveness. Moreover, ICT can be utilized to water supply management in order to cope with drought and rain risks.

Thailand is an agricultural country. According to the recent summary of the labor force survey in Thailand reported by National Statistical Office, Ministry of Information and Communication Technology in March 2013, 35.15% of Thailand's labor force is employed in agriculture [1]. Moreover, Thai agriculture is very competitive and highly required water supply management, therefore precision farming is needed.

One of the approaches to water supply management is to schedule irrigation for high water use efficiency which can be done by estimating the crop evapotranspiration (ET_c). ET_c can be calculated from a multiplication between crop coefficient and reference evapotranspiration (ET_0). The value of ET_0 can be measure from Lysimeter. This method is very expensive and can only be utilized by well-trained person, therefore there are many approaches introduced to indirectly estimate the value of ET_0 namely Penman-Monteith equation [2], Hargreaves equation [3], etc. Penman-Monteith equation is one of the most globally used method for ET_0 estimation [2]. It can achieve the highest accuracy among the other methods when there is enough data.

Artificial intelligence techniques were recently applied to estimate the value of ET_0 e.g. artificial neural network [4], [5], [6], [7]. Artificial neural networks (ANNs) are trained from data collected by sensors which are installed in the crop fields. The data consists of pairs of input objects (e.g. temperature, humidity, solar radiation, wind speed) and desired outputs (i.e, ET_0). Once the model has trained, it can predict the value of ET_0 from the input data collected from the sensors. Evapotranspiration is very complex and nonlinear model. However, only a single hidden layer ANN is enough to mimic the model together with six inputs which are minimum and maximum temperature, minimum and maximum relative humidity, wind speed, and solar radiation [4]. ANN was also successfully applied to estimate evaporation rate with air temperature, humidity, wind velocity and solar radiation [5].

Problems arise when a size of the crop field is very large. The cost of the sensor installations will be increased according to the size and some area may not be able to install the devices. Therefore, many researchers aim to analyze and report of which input features to be used in the prediction. Hence the cost and the number of devices could be reduced. Moreover, the computational time is reduced. There is an evidence that using inputs of air temperature, wind speed, humidity and solar radiation in Malaysia gives the highest accuracy of the prediction [5]. However, the selection of variables depends on the area where the experiment is conducted such as USA [4] Malaysia [5], Brazil [6], and Burkina Faso [7]. To the best of the authors' knowledge, there is no report about this matter in Thailand.

In real-world implementation, data loss is inevitable. Missing data can occur for many reasons: storage or sensor mechanisms are malfunction, system fails to respond when the data is transmitted. Therefore, it is not able to estimate the current reference evapotranspiration. We can use classic approaches to solve this problem. This can be done by mean substitution or list-wise deletion. However, these approaches can reduce the prediction accuracy. One of the popular statistical methods used for estimating missing values is Expectation-Maximization (EM) algorithm. It aims to compute the maximum likelihood estimation in the presence of missing data [8].

In this paper, we aim to predict the reference evapotranspiration by artificial intelligence techniques i.e. ANNs, linear regression (LR). We also study the relevance of the features using LR. Then we select sets of features and use them to train models. Moreover, we also apply EM algorithm to solve the missing data problem. The data used in the analysis is collected from two different provinces in Thailand.

The paper is organized as follows. Section II outlines the Penman-Monteith method, ANNs and EM algorithm. Section III explains experimental framework which includes the data collection, experimental results and discussion.

II. Methodologies

A. Penman-Monteith Method

It is a standard method which has been used for modeling evapotranspiration at Food and Agriculture Organization (FAO) of the United Nations [2]. The Penman-Monteith equation can be defined as,

$$ET_0 = \frac{0.408\Delta(r_n - G) + \gamma\frac{900}{T+273}U_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad (1)$$

where ET_0 is the reference evapotranspiration (mm/day), r_n is the net radiation at the crop surface (MJ·m^2/day), G is the soil heat flux density (MJ·m^2/day). T and U_2 are the mean daily air temperature ($^\circ$C) and the wind speed (m/s) at 2 meters height, respectively. e_s and e_a are the saturation vapor pressure (kPa) and actual vapor pressure (kPa), respectively. Δ is the slope vapor pressure curve (kPa/$^\circ$C), and γ is the psychrometric constant (kPa/ $^\circ$C).

B. Artificial Neural Network

Artificial Neural Network (ANN) is one of the nonlinear statistical data modeling technique. They aim to mimic the human brain functions and consist of weighted artificial neurons (nodes) in a layer. They can determine complex relationships between inputs and outputs of the data. A typical ANNs might have a hundred neurons and many layers. A basic structure of neural networks is shown in Fig. 1.

Consider the vector of m-dimensional inputs $\mathbf{x}=[x_1, x_2, ..., x_m]$ which have weight w_i associated with each input in a neuron. An output of the neuron, u, is a linear combination of inputs and weights:

$$u = \sum_{i=1}^{m} w_i x_i \quad (2)$$

The output of the neuron is fed into the activation function:

$$\hat{y} = \phi(u) \quad (3)$$

where $\phi(\cdot)$ is a sigmoid function:

$$\phi(u) = \frac{1}{1 + e^{-u}} \quad (4)$$

The feed forward ANNs can be adjusted in order to improve the performances i.e. number of nodes and activation function. Moreover, back propagation algorithm was applied in ANNs in order to reduce the error and increase robustness of the algorithm [9]. After the feed forward process, the error signals will be propagated backward through the network in order to adjust the weights in each node.

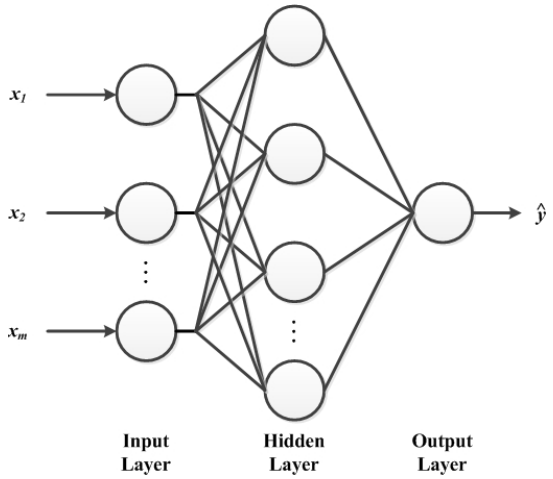


Fig. 1. Neural networks with single hidden layer.

C. Expectation-Maximization Algorithm

Expectation-Maximization (EM) algorithm is a widely used statistical technique to handle the incomplete data problem. Assume that $\theta^{(t)}$ is the t^{th} step in an iterative procedure. $\mathbf{X}$ is a set of observed data. $\mathbf{Z}$ is treated as missing values. The EM algorithm consists of two iterative steps which are as follows [8]:

- The expectation-step (E-step): In this step, it aims to calculate the expected value of the complete log-likelihood which can be calculated by,

$$Q(\theta|\theta^{(t-1)}) = E[\log p(\mathbf{X}, \mathbf{Z}|\theta)|\mathbf{X}, \theta^{(t-1)}]. \quad (5)$$

- The maximization-step (M-step): In this step, a new estimated is given by,

$$\theta^{(t)} = \arg\max Q(\theta|\theta^{(t-1)}). \quad (6)$$

III. Experiments

The data is collected from two automatic weather stations which are in Chiang Mai province and Ubon Ratchathani province. Both provinces are located to the northern and north eastern region of the country, respectively, as shown in Fig. 2.

The recorded climatic data is from 2006 to 2011. It was sampled and stored every hour. The features are listed in Table I. There is 2.2% of missing data. Hence, there is 11.18% of the samples which are not able to compute ET_0. The missing data was simply excised from the samples, therefore ET_0 can be calculated for every sample.

Next, we evaluate three different scenarios in this paper: prediction of reference evapotranspiration when (i) there is no missing data by different learning algorithms, (ii) there is different amount of missing data in real-world implementation, and (iii) the feature selection process is performed.

Fig. 2. Locations of the investigation areas in Thailand, namely Chiang Mai province (A) and Ubon Ratchathani province (B).

TABLE I
LIST OF THE FEATURES USED IN THE EXPERIMENT.

Index	Features
1	Day-of-year
2	Height above mean sea level (m)
3	Latitude (Radian)
4	Wind speed (m/s at 2 meters height)
5	Sunshine duration (hour/day)
6	Maximum temperature (°C)
7	Minimum temperature (°C)
8	Mean temperature (°C)
9	Relative humidity (%)

A. Scenario 1: No Missing Data

The experiment was run based on five-fold cross validation. The data is divided into five unique folds of roughly equal size. Then the algorithm is trained five times, leaving out one of the folds from training each time the machine is trained. The omitted fold will be used to compute the performance criterion, here, R^2 is used. R^2 is the coefficient of determination which measure how well output is predicted by the model. It can be calculated as follow:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (7)$$

where $\hat{y}$ is a prediction and $\bar{y}$ is the mean value of the target output.

We compared ANNs, and LR together. However, ANNs are required to tune their parameters to get the optimal performance, for example, the number of hidden layers and the number of nodes in the hidden layer. We again employed five-fold cross validation to tune the parameters in each training set. According to [4], the single hidden layer ANNs is sufficient for evapotranspiration model, therefore, we used the single hidden layer ANNs in this experiment. Hence, there is only one parameter to be tuned which is the number of node in the hidden layer. The optimal model of ANNs is with 38 nodes in the hidden layer. The experimental results are shown in Table II which reports R^2 of each algorithm. The results clearly show that ANNs yield the best performance. This is because the evapotranspiration is nonlinear model but LR is linear case.

TABLE II
PERFORMANCE COMPARISON BETWEEN ANN AND LR IN SCENARIO 1.

Algorithms	R^2
ANN (9-38-1)	**0.9999**
LR	0.9365

TABLE III
PERFORMANCE COMPARISON BETWEEN ANN AND LR WHEN MISSING DATA IS TAKEN INTO ACCOUNT.

Algorithms	Percentage of missing values	R^2 of re-estimated features	R^2 of the prediction
ANN	5%	**0.9771**	**0.9834**
LR			0.9204
ANN	10%	0.9472	0.9675
LR			0.9017
ANN	15%	0.9137	0.9443
LR			0.8759

In Thailand, there are three seasons which are as follows: (i) cool season (mid-November–mid-February), (ii) hot season (mid-February–mid-May), and (iii) rainy season (mid-May–mid-October) [10]. Fig. 3(a) and 3(b) shows the reference evapotranspiration computed by Penman-Monteith method and predicted by ANNs in Chiang Mai province and Ubon Ratchathani province, respectively. Clearly, they show that ANNs is competitive to the Penman-Monteith method. The reference evapotranspiration of both stations were at the highest point in April which is in the hot season. Then it reduced in rainy and cool seasons.

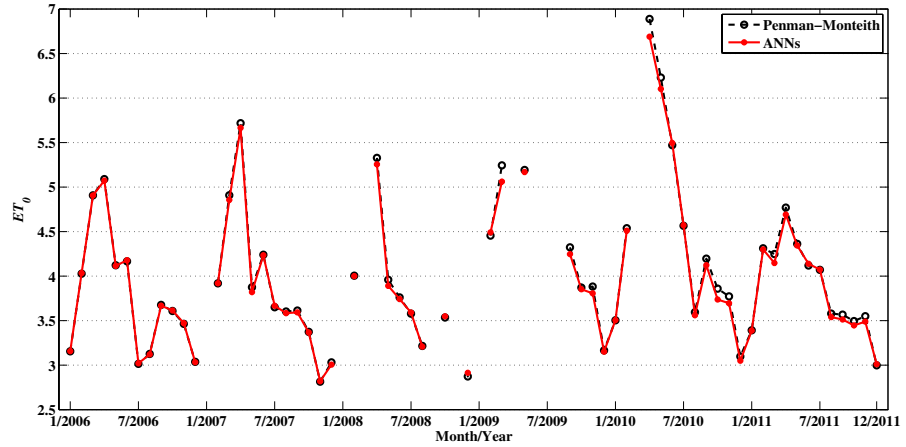
B. Scenario 2: Missing Data in Real-world Implementation

We again evaluated the results on five-fold cross validation and used the optimal models from the previous scenario. We randomly marked the test data as missing with 5%, 10%, and 15% of the data. The missing values were re-estimated by EM algorithm, then the data was trained by learning algorithms. Table III shows the experimental results. Using ANNs still gives the highest R^2 of the prediction in every cases. As expected, the performances of the re-estimated feature and the prediction of ANN and LR dropped when more missing data arises. It should be noted that only feature 4–9 can be missed as the others do not require sensors to collect the data.

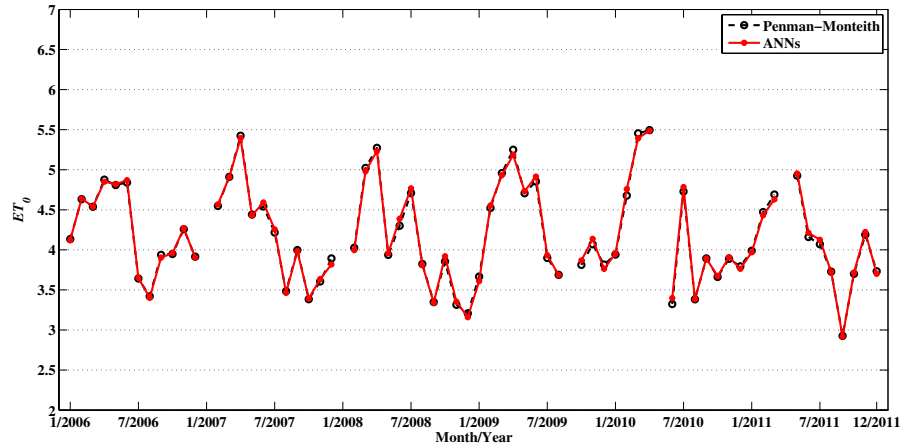
Fig. 4 shows comparisons of ET_0 computed by Penman-Monteith equation with ANNs (9-38-1) model (for the first random split) when there is no missing value, and 5%–15% of missing value. In the case which has no missing value, it has less scattered estimate than the other cases. When the amount of missing values increase, the more scattered estimate are.

C. Scenario 3: Selection of Features

The importance of the features used in the prediction is studied. LR seeks a linear combination of the variable to predict the outcome. The weight of the LR can roughly indicate the relevance of the features used in the prediction. According to the scenario 1, we have five linear regression



(a)



(b)

Fig. 3. Monthly reference evapotraspiration from 2006 to 2011 in (a) Ubon Ratchathani province and (b) Chiang Mai province by Penman-Monteith equation and ANNs.

models. Then we calculated the average of the absolute values of each element in w_i across five vectors. According to Fig. 5, it can be seen that sunshine duration is the most important feature. Moreover, we ranked the values of absolute weight in descending order which gives the following order:

$$5 > 7 > 8 > 4 > 9 > 1 > 6 > 3 \sim 2$$

In this scenario, we trained models by adding new features one-by-one according to how important the features are. Five-fold cross validation is used to evaluate the performances and search for the optimal parameters. Missing value cases were also considered too.

Table IV shows the performances of ANN and LR when different set of features are used. The classifiers with 8 features (i.e. feature 5, 7, 8, 4, 9, 1, 6, and 3) and 9 features gave the highest accuracy for both ANN and LR cases. Fig. 6 shows the relative improvement or worsening in the accuracy for ANN and LR when we considered new features in the process. In ANN, adding feature 7 (minimum temperature) to feature 5 (sunshine duration) could improve the performance by 40.99% of using only feature 5. In addition, when we considered mean temperature and wind speed in the process, the performances were improved 5.7% and 7.2%, respectively. Again, the performance were improved by 1.6% and 4.2% after adding relative humidity and day-of-year, respectively. Unfortunately, considering maximum temperature, latitude, and high above mean sea level in addition, did not improve much. Overall picture is much the same for LR.

Moreover, we also examined the feature selection process together with missing values case by ANN as shown in Table V. The overall picture is still much the same for the case with no missing values. The performances of the algorithms drop when the amount of missing values increases. Here, ANNs (8-49-1) with 5% missing values yield the best performance.

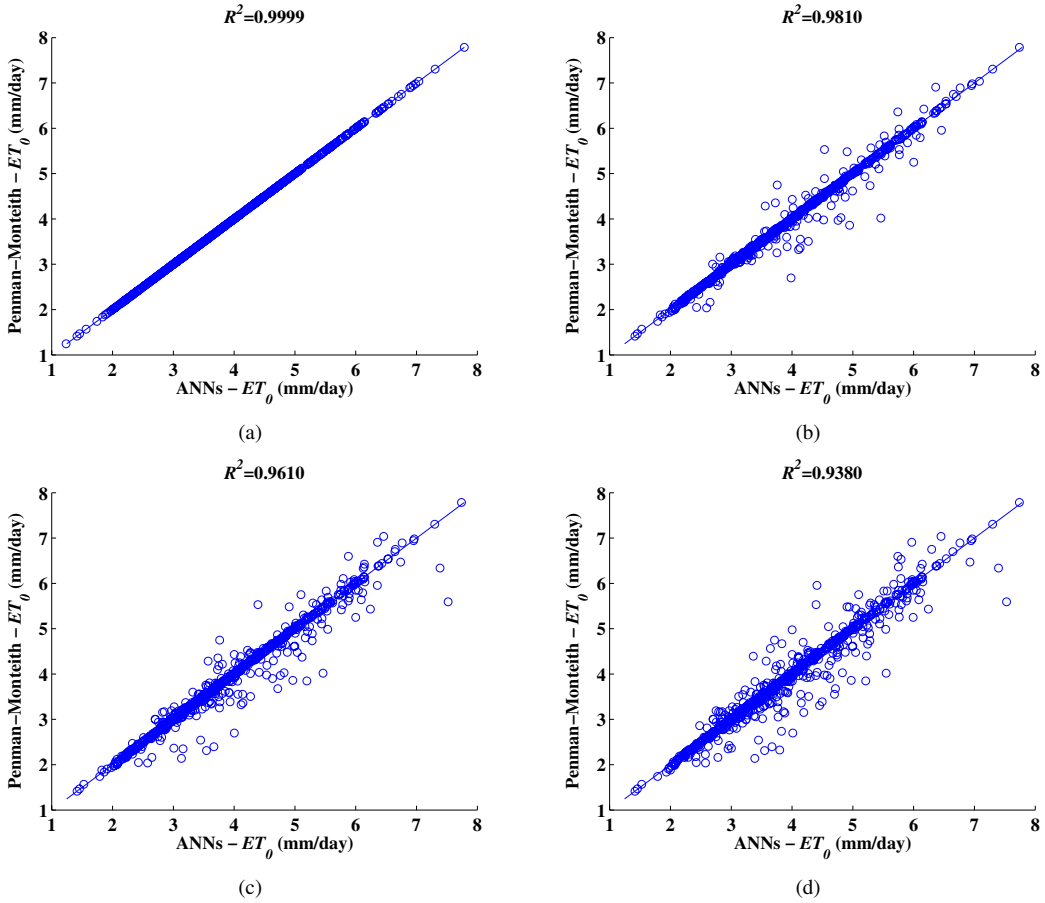


Fig. 4. Comparison of ET_0 computed by Penman-Monteith equation with ANNs (9-38-1) model when there is (a) no missing value, (b) 5% of missing value, (c) 10% of missing value and (d) 15% of missing value.

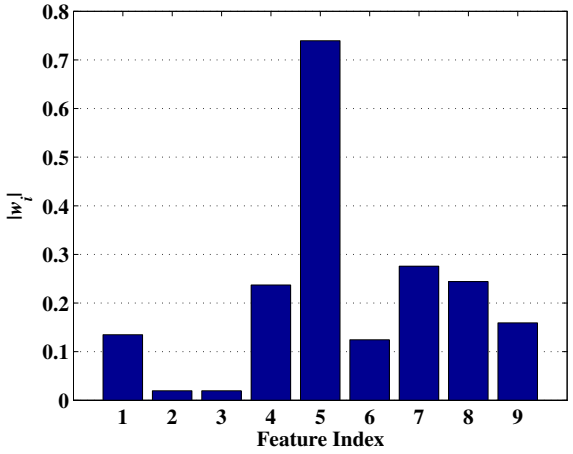


Fig. 5. An illustration of feature importance. It shows the absolute value of the linear regression weight vectors w_i average across all the models.

TABLE IV
PERFORMANCE COMPARISON BETWEEN ANNS AND LR WHEN FEATURE SELECTION IS CONSIDERED.

Features used	Algorithms	R^2
5	ANN (1-6-1)	0.5889
	LR	0.5709
5,7	ANN (2-10-1)	0.8303
	LR	0.8010
5,7,8	ANN (3-12-1)	0.8776
	LR	0.8511
5,7,8,4	ANN (4-32-1)	0.9410
	LR	0.9122
5,7,8,4,9	ANN(5-21-1)	0.9561
	LR	0.9260
5,7,8,4,9,1	ANN(6-29-1)	0.9963
	LR	0.9355
5,7,8,4,9,1,6	ANN(7-40-1)	0.9960
	LR	0.9355
5,7,8,4,9,1,6,3	ANN(8-49-1)	**0.9999**
	LR	0.9365
5,7,8,4,9,1,6,3,2	ANN(9-38-1)	**0.9999**
	LR	0.9365

IV. CONCLUSIONS

In this paper, we used ANN and LR to predict the reference evapotranspiration from the climatic data in Thailand. We have shown that ANN is competitive to the most widely used method the so-called "Penman-Monteith equation". The performance of LR is generally worse than ANN, however, this could be improved by using kernel regression as the evapotranspiration is nonlinear model. We also applied EM

TABLE V
PERFORMANCE OF ANNS WHEN FEATURE SELECTION IS CONSIDERED TOGETHER WITH MISSING VALUE SITUATIONS.

Features used	Algorithms	Percentage of missing values	R^2 of re-estimated features	R^2 of the prediction
5	ANN (1-6-1)	5%	0.9619	0.5752
		10%	0.9619	0.5752
		15%	0.9619	0.5752
5,7	ANN (2-10-1)	5%	0.9586	0.8046
		10%	0.9149	0.7610
		15%	0.8576	0.7140
5,7,8	ANN (3-12-1)	5%	0.9590	0.8544
		10%	0.9186	0.8306
		15%	0.8665	0.8070
5,7,8,4	ANN (4-32-1)	5%	0.9543	0.9086
		10%	0.9090	0.8803
		15%	0.8598	0.8520
5,7,8,4,9	ANN (5-21-1)	5%	0.9619	0.9329
		10%	0.9231	0.9098
		15%	0.8860	0.8838
5,7,8,4,9,1	ANN (6-29-1)	5%	0.9619	0.9772
		10%	0.9231	0.9527
		15%	0.8809	0.9272
5,7,8,4,9,1,6	ANN (7-40-1)	5%	**0.9740**	0.9772
		10%	0.9456	0.9674
		15%	0.9112	0.9335
5,7,8,4,9,1,6,3	ANN (8-49-1)	5%	**0.9740**	**0.9818**
		10%	0.9456	0.9605
		15%	0.9112	0.9390
5,7,8,4,9,1,6,3,2	ANN (9-38-1)	5%	**0.9740**	0.9835
		10%	0.9456	0.9617
		15%	0.9112	0.9391

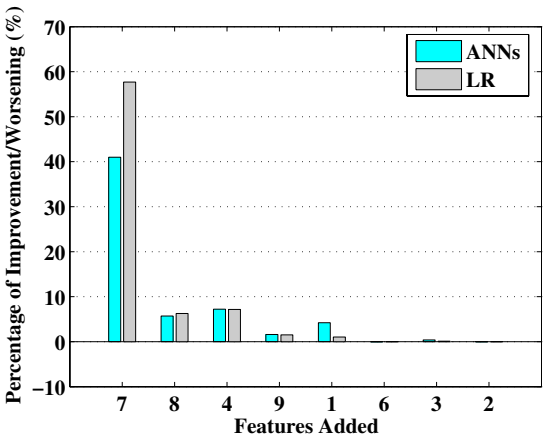


Fig. 6. Relative improvement/worsening in accuracy of ANN and LR when adding new features one-by-one according to how important the features are.

algorithm to solve the problem when missing data occur in real-world implementation. Therefore, we still can predict the reference evapotranspiration. When the amount of missing data increases, the performances of the re-estimated feature and the prediction of the learning algorithms will drop. However, this is still worthwhile. Moreover, the feature selection process was performed. We have shown the importance of the features used in the prediction. The most three importance features are sunshine duration, temperature and wind speed, respectively. In this study, we have only investigated on two automatic weather stations, therefore, we did not see much impact to the prediction with the height above mean sea level and latitude of the stations. However, if there is a number of automatic weather stations, these two features might have more effect to the prediction.

REFERENCES

[1] National Statistical Office, "Summary of the labor force survey in Thailand: March 2013," Ministry of Information and Communication Technology, Bangkok, Thailand, 2013.

[2] R. Allen, L. Pereira, D. Raes, and M. Smith, "Crop evapotranspiration-guidelines for computing crop water requirements," *FAO Irrigation and Drainage Paper*, vol. 56, pp. 15–78, 1998.

[3] G. Hargreaves and R. Allen, "History and evaluation of hargreaves evaportranspiration equation," *Journal of Irrigation and Drainage Engineering*, vol. 129, no. 1, pp. 53–63, 2003.

[4] M. Kumar, N. Raghuwanshi, R. Singh, W. Wallender, and W. Pruitt, "Estimating evapotranspiration using artificial neural network," *Journal of Irrigation and Drainage Engineering*, vol. 128, no. 4, pp. 224–233, 2002.

[5] M. Benzaghta, T. Mohammed, A. Ghazali, and M. Soom, "Prediction of evaporation in tropical climate using artificial neural network and climate based models," *Scientific Research and Essays*, vol. 7, no. 36, pp. 3133–3148, 2012.

[6] S. Zanetti, E. Sousa, V. Oliveira, F. Almeida, and S. Bernardo, "Estimating evapotranspiration using artificial neural network and minimum climatological data," *Journal of Irrigation and Drainage Engineering*, vol. 133, no. 2, pp. 88–89, 2012.

[7] T. Wang, S. Traore, and T. Kerh, "Neural network approach for estimation reference evapotranspiration from limited climatic data in Burkina Faso," *WSEAS Transactions on Computers*, vol. 7, no. 6, pp. 704–713, 2008.

[8] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[9] A. Dongare, R. Kharde, and A. Kachare, "Introduction to artificial neural network," *International Journal of Engineering and Innovative Technology*, vol. 2, no. 1, pp. 189–194, 2012.

[10] Royal Command of H.M. the King, "Climate," *Thai Junior Encyclopedia*, vol. 4, no. 6, 1978.

Preliminary Design of Static Indonesian Sign Language Recognition System

Rudy Hartanto, Adhi Susanto, and P. Insap Santosa
Department of Electrical Engineering and Information Technology, Gadjah Mada University
Jalan Grafika No. 2, Yogyakarta, 55281 Indonesia
Email: rudy@mti.gadjahmada.edu

Abstract— **Human computer interaction has a long history to become more intuitive. For human being, especially for the deaf, gesture of different kind is one of the most intuitive and common communication.**

In this paper we focus on creating a system to identified and translate hand gesture pose to Indonesian alphabets. Skin detections method is used to create a segmented hand image and to differentiate with the background. A contours is used to localize hand area. SIFT algorithm in advanced, were used to recognize the signed gesture.

The result shows that this system can operate well in translated hand gesture image of sign into Indonesian alphabets.

Keywords-hand gesture, human computer interaction, skin detection, contours, SIFT Algorithm.

I. INTRODUCTION

Sign language is a language that uses body gestures instead of sound to communicate. The common gesture used is a combination of the shape or pattern of hand, orientation and hand gestures, facial expressions as well as lip patterns. Sign language is the primary communication tool for people with hearing-impaired or deaf. Problem will arise when people with hearing-impaired communicates with normal people who do not understand sign language. The constraint in communication will affect the lives and interpersonal relationships within the community of hearing-impaired and deaf. Of course, this problem needs to be addressed to make the communication with the normal people can do smoothly, which in turn will increase the harmonious relationship between components of society.

The use of computers as a tool for sign language recognition has long been studied and developed by experts in the field of human computer interaction. Some research has begun to explore a variety of methods and algorithms used hand gesture recognition in sign language.

Ravikiran [1] proposed a finger detection method to recognize American Sign Language (ASL). Recognition process is done by tracking the finger boundary and the detection of the fingertip. Other studies of static hand gesture recognition to recognize the alphabet in ASL performed by Rahman [2]. The recognition is done by using a back propagation neural network. Yang [3] studied the hand gesture recognition using the Hidden Markov Model (HMM).

Three are a lot of sign language in the world, each has its own grammars and develops independently, so do in Indonesian sign language. This Paper is a preliminary research that will specifically explore static Indonesian sign language recognition. We propose Convexity defect algorithms to get the contours of the hand gesture pose and then using SIFT (Scale Invariant Feature Transform) algorithm to recognize hand gesture pose image and compare with the matching image in the database.

A single camera is used to capture hand gesture pose which then will be recognized by the system as an Indonesian alphabet. As author's known based on a variety of reference sources, there has been no research that utilizes a web camera to capture the static Indonesian sign language using a Convexity defect algorithm combined with SIFT algorithm to recognize hand gesture pose, and then translated into Indonesian alphabet.

II. RELATED WORKS

Deaf and hearing-impaired persons use a sign language for communication with each other, however they often find a difficulty to communicate with normal persons due to the the lack of public understanding of sign language which they use as a means of daily communication. One way to eliminate these obstacles is to create a system that can recognize and translate the sign language used by the hearing-impaired and deaf community, so it can be understood by the general public. Recently, a lot of research has been conducted in order to develop a variety of methods and algorithms that can be used to recognize and translate sign language into common spoken language.

Asriani [4] uses back propagation neural networks with 15 static class cue word to recognize static hand gesture images. In this study used. Segmentation process is performed using HSV filter and feature extraction is done by wavelet Haar decomposition up to level 2. Further classification process is done using back propagation neural network with architecture that consist of 4096 neurons in the input layer, 75 neurons in the hidden layer and 15 neurons in the output layer. The system has been tested using 225 validation data which has achieved up to 69% accuracy.

Yang [3] has developed a hand gesture recognition system using HMM (Hidden Markov Model) algorithm. With the assumption that hand gestures can only be detected as

fragmented regions along with clutter, so they have developed a new method of HMM method called frag-HMM, which allows a thought based on fragmented observations, using an intermediate grouping process. Based on the model developed, it can be shown how to perform three kinds of computations. The first is determining the best observation group for each frame, given a sequence of observation group for the past frame. This allows them to incrementally compute the best segmentation of the hand for each frame of a given model. The second one involves the computation of likelihood of a sequence, averaged over all the possibility of all possible states sequences and grouping. The third is the computation of the sequential series, maximized over all possible state and group sequences. The results showed improved recognition performance by 2% compared with hand segmentation process manually and 10% better than the previous segmentation that uses the skin hand color.

In 2009, Ravikaran et al. [1] introduce a fast and efficient algorithm to identify the number of fingers that opened in the representation of the hand gesture for the American Sign Language alphabet. Resolved detection based on the concept of finger tracking and detection limit of the fingertip. System does not require the hand to be equipped with special signs or input gloves.

Another study conducted by Tao Ni et al.[5]. They had developed a new interaction technique called the "roll-and-pinch menu" (rapMenu). They examine the design space for selecting menus remotely by using hands-free input gestures. Rap menu will be mapped intuitively hand postures and gestures to a radial menu layout with two menus, the user can rotate the hand to alter the wrist orientation to select a menu group, then touched her finger to a particular finger (forming a pinch gesture) to select specific item in the menu group.

III. THEORIES

A. Contour Detections

Contour is a point that represents a collection of curves in an image [6]. Contour is useful in many applications. It can be used for detection of contour shape and pattern recognition of an object. The detection of an object can be done by comparing the contours of the object being examined with predetermined contours.

Convex hull

Convex hull is a classic problem in geometry computation. Convex hull is described simply in a field as search a subset of the set of points on the field, so that if the points are then used to form a polygon will perform a convex polygon. A polygon is said to be convex if the line that connects two points in the polygon do not cut the line of the polygon. Convex hull of an object P is defined as the area surrounding the smallest convex polygon P. Therefore, for a given set of N points $\{p_0, p_1, p_2, \ldots, P_N\} \varepsilon P$, it can be stated that the hull H can be constructed by M point of the set N to make a minimum area convex polygon. From Fig. 1 it can be stated that the convex hull is made by taking the interior angle θ, of three adjacent points {p1, p0, p9}. If $\theta > \pi$ then regarded as reflex points p_0 and p_0 are not members of M. Final set of H is {p1, p9, p7, p5, p} [7].

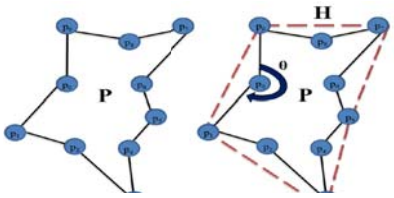


Figure 1. Generalization logic convex hull [7].

The searching of convex hull from a set of points Q (CH (Q)) is to find a smallest convex set containing all points in Q. Convex hull of a set of points Q (CH (Q)) in n dimensions is a whole slice of all convex sets containing Q. So that for an N points {p1, p2, ..., pN} P, the convex hull is the set of convex combination is expressed as:

$$CH(Q) \equiv \left\{ \sum_{j=1}^{N} \lambda_j p_j ; \lambda_j \geq 0; \sum_{j=1}^{N} \lambda_j = 1 \right\} \quad (1)$$

Convexity Defect

Convexity defects is a feature that used to find a defect between the convex hulls formed by the contours of polygon. Defect is useful to find a feature in a polygon, one of which is to detect human fingers [8]. For more details on the method convexity defect can be seen in Fig. 2.

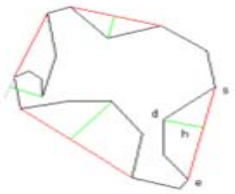


Fig. 2. Convexity defects of a polygon [8].

Fig. 2 above shows the convex hull depicted with red lines covering polygons with black contour lines. The symbol "s" and "e" indicates the "start point" and "end point" of the convexity defect. While the symbol "d" represents the "depth point", i.e. the contour point which lies between the "s" and "e" which is the furthest point of the convex hull contour lines, denoted with "se". The symbol "h" is the depth of the defect which is the distance from "d" to the line "se" [9]

Start point, end point, depth points, and depth, are the four elements that will be used to find the feature of a polygon so that it can be further processed.

Segmentation of Skin Color (*Skin Detection*)

Skin color segmentation is widely used for face recognition applications, detection of body, or limbs. The objective is to detect the skin color of each image captured by the camera. Image obtained generally have RGB format. To detect the skin color, RGB format is converted to YCrCb format to separate the intensity Y with chromacity expressed in two variables Cr and Cb. In this model we only use the color information Cr and Cb, so that the effect of changes in intensity can be eliminated. In the saturation region of light caught on camera, the value of

Cr and Cb is very stable, so that Cr and Cb value is reliable information for the color classification process. Conversion from RGB to YCrCb can be obtained using (2):

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.29900 & 0.587000 & 0.114000 \\ -0.168736 & -0.331264 & 0.500000 \\ 0.500000 & -0.418688 & -0.081312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2)$$

Where Y is the luminance (intensity of color), is the blue color component, and Cr is the red color component [10].

After the conversion to YCrCb format, then the image converted to the binary form using a thresholding process to simplify the next process. Skin color segmentation is done by determining the value range of the Y, Cr and Cb. For every pixel in the image, if it is in the range, it will be considered an object, while outside the range it would be considered as a background. [10].

B. Searching the Center Point of Hand Image.

The process of finding the center point of the hand image after segmentation can be obtained using (3).

$$\bar{x} = \frac{\sum_{i=0}^{k} x_i}{k} \ , \ \frac{\sum_{i=0}^{k} y_i}{k} \quad (3)$$

Where x_i and y_i are the x, y coordinates of ***i*** pixels in the area of hand, and ***k*** indicates the number of pixels in the area.

To get a hand size, a circle can be made by increasing the radius of the circle from the circle center coordinates until the circle reach first black pixels. When the algorithm has reached the first black pixel, it will going back to the previous value of the radius.

This algorithm uses assumption that when the circle reaches the first black pixel, after drawing the growing circle, then the length of the center is the radius of the hand. In this approach, image segmentation will play a very important part, because it marks the hand boundary by black pixel of segmentation border.

C. Feature Extraction and Hand Gesture Pattern Recognition

To recognize the pattern, the input image is matched with the image of the database. The first step is extracted image features that have many properties that suitable for use in image matching on the object or scene images. This feature is invariant to image scaling and rotation, as well as in brightness. They are much localized in both spatial and frequency domain, reducing the possibility of disruption by occlusion, clutter, or noise. A large number of features can be extracted from typical images with efficient algorithms [11].

Here are four major computational steps used to generate a set of image features.

1. Extreme scaled-space detection: The first computation step is finding all scalable locations and image. This process can be implemented efficiently by using the Gaussian difference function to identify potential interest points that invariant to scaling and rotation.

2. Key point localization: At each candidate location, details of the model was fitted to determine the location and scale. Key points are selected based on the measurement stability.

3. Orientation marking: For each key-point location, one or more orientations are marked based on local image slope direction. Throughout subsequent operations performed on image data that have transformed relative to the marked orientation, scale, and location of each feature, which thus becomes the invariant against the transformation.

4. Key-point descriptor: Local image slope is measured on a scale that is selected in the region around each key point.

The above approach is called the Scale Invariant Feature Transform (SIFT), which will transform image into the scale invariant coordinates relative to local features. For matching and recognition, first the SIFT features extracted from the set reference images and then stored in a database. The new image will be matched with comparing every feature of the new image with the previous database and finding candidate matching features based on Euclidean distance of its vector.

Key-point descriptor is very specific, which allows a single feature to find a match with a good chance in a large database features. Correct match can be filtered from the complete set of matches by identifying a subset of the key points that meet the object and the location, scale, and orientation in the new image [11].

IV. IMPLEMENTATION

The general architectural design for Indonesian static sign language recognition system is shown in Fig. 3. The input of the system is hand gesture pose images of Indonesian sign language alphabets captured from the camera (webcam). The output of the system is the text resulted from the alphabetic Indonesian sign language recognition system.

A. Segmentation

Segmentation process is used to identify the object which in this case is the image of the hand. One of the simple algorithm is detecting the difference between the user's hands skin color with the background, which is known as skin detection algorithm. This method will detect the skin color of the captured image. Image obtained from camera webcam has RGB format.

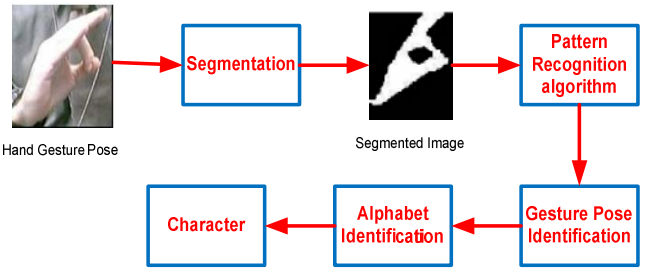


Fig. 3. System Architecture

To make the skin color model, RGB format is converted to YCrCb format to separate between intensity and chromacity

(colors), each of which is expressed by Cr and Cb respectively. This conversion can be obtain by (4).

$$Y = 0.299*R + 0.587*G + 0.114*B \quad (4)$$
$$Cr = (R - Y)*0.713 + 128$$
$$Cb = (B - Y)*0.564 + 128$$

After the conversion to YCrCb format then the image will be converted into binary format with the thresholding process. Skin color detection is done by determining the range of values of the Y, Cr and Cb, which is the color value of user skin used in determining whether the pixel value will be white (an arm object) or black (a background).

$$w(x,y) = \begin{cases} 255, & if\ Cr1 < Cr < Cr2\ \ dan\ Cb1 < Cb < Cb2 \\ 0, & otherwise \end{cases} \quad (5)$$

Where Cr1 and Cr2 as well as CB1 and CB2 are an acceptable hand skin color value range. Value range of the Y, Cr and Cb used in this research are 0 to 255, 131 to 185, and 80 to 135. This value can be adjusted according to the user's skin color.

For each pixel in the image, if it is in the range, it will be considered as skin color and its value will be changed to 255 (white), otherwise, it will change the value to 0 (black). With this method, will results a binary image with white pixel representing the user's hand and the black pixel represents the background.

B. Contour detection

Contour is actually a pixel series of same value. Contour found if there is a difference of point's values with its neighbors. This process based on the results of edge detection in a binary image from skin detection results between white pixels (objects that have a color resembling hand) and black pixels (objects other than hands or background). The error possibility of this method is quite high if the objects in an image have almost the same color. Fig. 4 show a flowchart of contour detection process.

The contour resulted from the process will be stored in the pointer. Each contour can be accessed individually. To ensure that each captured image is only have one contour, it is necessary to determine the largest contour areas. So that only the contour with the largest area who will set up its contours, the others contours will be ignored. The contour search results are shown in Fig. 5.

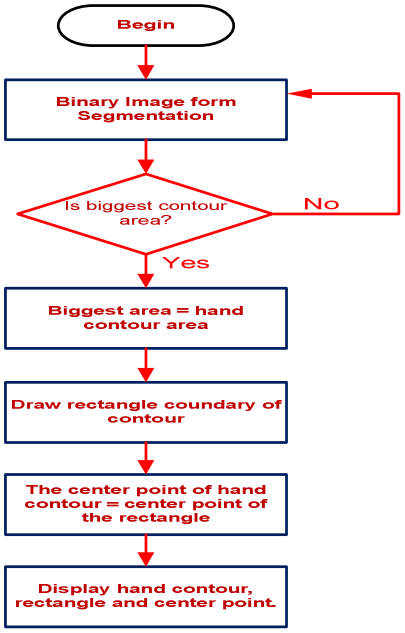


Fig.4 Contour detection flowchart

Fig. 5 Figure hand covered contours.

C. Identification the Pattern of Hand Gesture

Identification process to recognize the alphabet pattern of hand gesture is shown in Fig. 6. The input is a segmented image resulted from the previously process with its background has been removed (replaced with black). The input image is processed by using SIFT algorithm to extract a number of key points (features) which would then be compared with the image of the hand from a database consisting of 24 poses that form the sign language alphabet as shown in Fig. 7. The system output is an image database that match the input image, which means point to one particular alphabet.

The comparison process is done by first extracting features in the form of key points characteristic and the orientation of the input image and the image database. Based on the value and position of these key points will be carried out the comparisons to find images in the database that match with the characteristics of the input image.

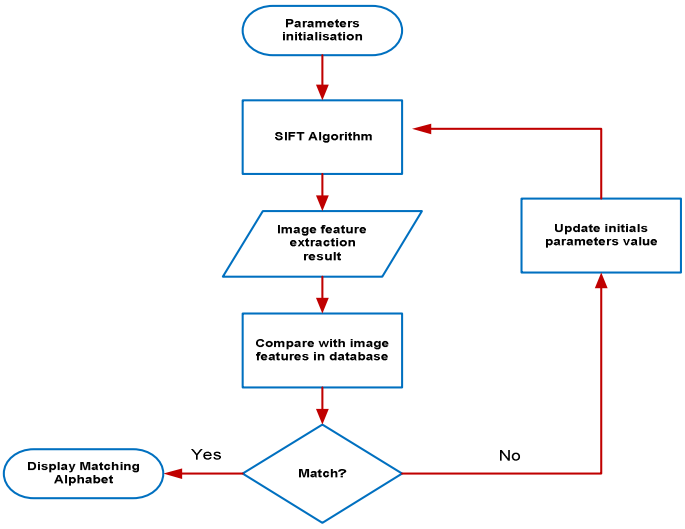


Fig. 6 Flowchart the process of image matching recognition and hand gesture.

The process of matching features between the key points of the input image to the image database based on the number and location will be done one by one, until there is a match the number and location of the key points.

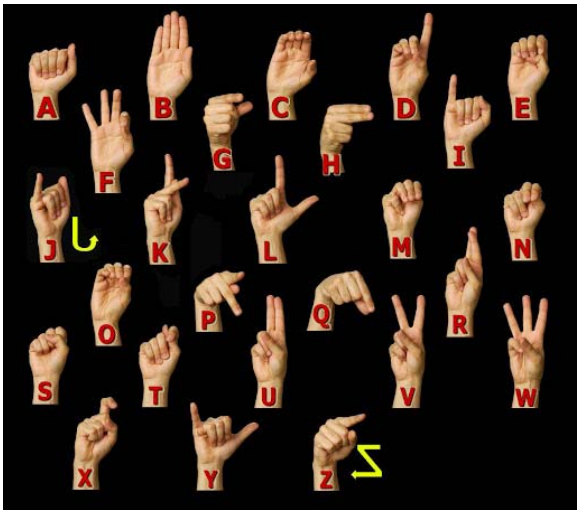


Fig. 7 Image poster hand gesture representing 26 alphabet is stored in a database.

In this research the alphabet 'J' and 'Z' is not tested, because contains dynamic elements, so that the system cannot recognize correctly. The experimental results for the 24 characters who the image signs are already stored in the database are shown in Table 1. The sample image have differences with the image Stores in database in terms of image size and orientation because it's very difficult to make exactly the same.

Table 1 shows that the performance of system to recognize hand gesture pose image of Indonesian sign language is still relatively small, with an average of 62.6%. Approximately 37.4% of the hand gesture image pose is not recognized, and approximately 12.8% pointed to wrong alphabet.

TABLE I. THE IMAGE MATCHING RESULTS OF INDONESIAN SIGN FOR 25 ALPHABETICAL (A TO Y) IN %.

Alphabets	Being Recognized	False recognized
A	60	20
B	65	0
C	70	0
D	60	15
E	55	20
F	60	10
G	65	15
H	75	15
I	70	10
K	60	15
L	70	0
M	55	20
N	55	25
O	70	0
P	60	10
Q	60	15
R	55	25
S	60	20
T	60	20
U	60	10
V	65	10
W	65	10
X	60	15
Y	70	5
Average	62,6	12,8

D. Discussion

From the observed data can be concluded that the system is able to recognize the image of the hand gesture pose of Indonesian sign language representing the Latin alphabet, although their identification success rate is still relatively low at around 62.6%. Table 1 shows that the percentage of successful hand gesture recognition poses of each Indonesian sign language alphabet is not the same, because some alphabet has the similar shape of gesture poses, making it difficult for the system to identified.

Fig. 7 shown that pose of hand gesture image representing A, E, M, N, S, and T alphabet has the form or a pose that nearly equal, so that the key features of these points are also relatively similar. Furthermore the image size that relatively small at around 320x240 pixels also contribute to that low performance. The consideration of image size selection due to the duration of computing process that is more than 0.6 seconds. The length of computing time is caused by the algorithm that must extract key points feature of each database image to be compared with the input image key point features. So in each of the matching process will extract the key point's features of the 24 images from database. Thus, the system cannot yet be used for online identification, because the speed of the hand gesture image sampling is still low, less than 2 frames per second (2 FPS).

In practice it is very difficult or almost impossible to get the image of the hand gesture pose exactly the same as the one stored in the data base, even if done by the same person. Similarly orientation or inclination of hand in sign language is very difficult to demonstrate constantly. This factors led to lower the recognition percentage.

E. Advantages and Limitations of System

Based from the observation and discussion, it can be summarize the advantages and limitations of had been build.

Pros:

1. The system has been able to implement some of the algorithms works well in identifying Indonesian static sign image.
2. The image feature matching algorithm SIFT has been functioning well and can recognize the image of Indonesian sign.
3. System is made relatively simple but quite effective.

Cons:

1. The computational process is still quite long, so it is simply not possible to be used real-time yet. It's required to optimize the algorithms as well as the size of the image that will improve the performance of recognition systems.
2. Recognition percentage is still relatively low, which is about 62.6%, this is due to the difficulty in demonstrating the sign constantly.
3. The size of the image is used as the image database is still relatively small, considering the relatively long computation process.

V. CONCLUSION

A. Conclusion

Based on observations and discussions can be deduced as follows.

1. Indonesian static gesture recognition systems can be implemented and work as expected.
2. The recognition process of Indonesian alphabet gesture performed using SIFT algorithm approach.
3. Performance of the system in recognizing Indonesian sign is still relatively low, which is about 62.6%.
4. The computing time required is still large, so that the system is still not able to be used real-time.

B. Suggestions for further studies

Based on the limitations of the system, for further research is recommended to:

1. To optimize all of the algorithm used to shorten the computational time.
2. Increase the percentage of the recognized image from the input signal by increasing the size or resolution of the image, so that the number of features that serve as the key point of the unique features of each image will be more numerous and unique.

REFERENCES

[1] Ravikiran, K. J Mahesh, S. Mahishi, D. R, S. S and N. V. Pujari, "Finger Detection for Sign Language Recognition," in *Internal Multi Conference of Engineers and Computer /scientists*, Hong Kong, 2009.

[2] M. A. Rahman, Ahsan-Ui-Ambia and M. Aktaruzzaman, "Recognition Static Hand Gesture of Alphabetic ASL," *IJCIT,* vol. 2, no. 01, 2011.

[3] R. Yang and S. Sarkar, "Gesture Recognition using Hidden Markov Models from Fragmented Observations," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, 2006.

[4] F. Asriani and H. Susilowati, "Pengenalan Isyarat Tangan Statis pada Sistem Isyarat Bahasa Indonesia Berbasis Jaringan Syarat Tiruan Perambatan Balik," *Makara, Teknologi,* vol. 14, no. 2, 2010.

[5] T. Ni, M. R.P. and B. D.A., "Tech-note: rapMenu: Remote Menu Selection Using Freehand Gestural Input," in *IEEE Symposium on 3D User Interfaces, 2008 3DUI,* Reno, NE, 2008.

[6] H. Kim and F. D.W., "Interaction With Hand Gesture for a Back Projection All," *Computer Graphics International, IEEE Computer Society,* 2004.

[7] M. M. Youssef, "Hull Convexity Defect Features for Human Action Recognition," The School of Engineering of The University of Dayton, 2011.

[8] G. Bradski and A. Kaehler, Learning OpenCV, Gravenstein Highway North, eSbastopol, CA 95472: O'Reilly Media Inc., 2008.

[9] INTEL, Open Source Computer Vision Library, Intel Corporation, 2001.

[10] H. Park, "A Method for Controlling Mouse Movement using Real-time Camera," Brown Computer Science, 2010.

[11] D. Lowe, "Distinctive Images Features from Scale Invariant Key-points," *International Journal of Computer Vision,* vol. 6, no. 2, pp. 91-110, 2004.

Rapid Nitrogen Determination of Soybean Leaves Using Mobile Application

Marcelinus A.S. Adhiwibawa
Ma Chung Research Center for Photosynthetic Pigments
Universitas Ma Chung
Malang, Indonesia

Christian Tantono
Dept. of Information Technology
Universitas Ma Chung
Malang, Indonesia

Kestrilia R. Prilianti
Dept. of Information Technology
Universitas Ma Chung
Malang, Indonesia

Monika N.P. Prihastyanti
Ma Chung Research Center for Photosynthetic Pigments
Universitas Ma Chung
Malang, Indonesia

Leenawaty Limantara
Ma Chung Research Center for Photosynthetic Pigments
Universitas Ma Chung
Malang, Indonesia

Tatas.H.P. Brotosudarmo*
Ma Chung Research Center for Photosynthetic Pigments
Universitas Ma Chung
Malang, Indonesia
tatas.brotosudarmo@machung.ac.id

***Abstract*—Nitrogen is one of the important nutrients elements for the growth of soybean plants. In this paper we propose mobile application that can be used nondestructively to estimate the nitrogen content of soybean leaves. We named this software "Mata Daun". The primary concept of this software is to relate the RGB (Red, Green, Blue) value of the captured soybean image with its nitrogen content. Furthermore, the captured image is processed into Enhanced Color Visibility (ECV) index using digital image processing method for the ease of software algorithm process. Calibration process and field trial were conducted to found the relation between ECV index and soybean leaves nitrogen content. The calibration result showed that the nitrogen readings by this application had a fairly strong relationship (R^2 =0.70) with the soybean leaves nitrogen content (Agriexpert CCN-6000 readings). The field test result also gave the same strong positive relationship between predicted and real soybean leaves nitrogen content (R^2 =0.93).**

Keywords—nitrogen; soybean; RGB; image processing; mobile application (key words)

I. INTRODUCTION

Nutrients deficiency, particularly nitrogen, can lead to a physiological imbalance in soybean plants. One of the impact is a reduction in yield during harvest[1,2]. Addition of fertilizer is needed to fulfill the shortage of necessary nutrients on plants. Prior to fertilizer application, information about the levels of nitrogen in plant is required to avoid excessive nitrogen fertilizer application. High amount of nitrogen in plant tissue can increase the attack from pests and diseases and also can harm the plant if the dose is high [3,4,5]. The most accurate method to determine levels of nitrogen in plant tissue is laboratory test. The laboratory test result requires longer time due to the complexity of sample preparation. Usually, it needs 3-7 days from sample preparation to the result. The complexity of laboratory testing process is one of the barriers for farmer to check nitrogen level of plant leaves at the beginning of the growing season. The lack of farmer's knowledge about nitrogen status of their plant can drive to excessive use of fertilizer that can be harmful to their plant and surrounding environment. In this paper we propose a mobile application called "Mata Daun". This mobile application is intended to help soybean farmers in determining real time plant nutrient status at low cost. With the ability to estimate the nitrogen content of soybean leaves nondestructively and rapidly, we hope this mobile application can improve the farmer's knowledge about their soybean plant nutrient condition.

This research is sponsored by Ma Chung Research Grant under the grant number 002/MACHUNG /LPPM-MRGV/X/2 012.

II. METHODS

A. Hardware

Samsung smart mobile phone type Wonder I8150 was used to conduct all the mobile application experiment. This mobile phone embedded with 1.2 Gigahertz processor, 2 Gigabyte Read Only Memory, 5 megapixels camera and 2.3.5 Ginger Bread Android operating system. The white paper background reference was a sheet of 80 gram A4 size white paper.

B. Software

Mobile application was developed based on JAVA language. It is compatible with Android version 2.3 up to 4.0.

Digital image processing algorithm was embedded to obtain estimation of the nitrogen content of soybean leave being observed. Mobile device's camera captured reflected light from soybean leaf surface and white paper background reference. The captured light was then converted into RGB (Red, Green, Blue) format. Afterwards, global thresholding method was applied into RGB image to automatically identify the RGB value of soybean leaves and white paper background references. Both RGB value was then converted into Enhanced Visibility (ECV) index. ECV index used in the "Mata Daun" application is a development from Color Visibility (CV) index which is used in the "Bai Khao" mobile application [6]. The improvement that was made is the color correction mechanism to deal with illumination variation problem. Color correction was made by the addition of the corrected white paper background RGB value to the index equation. Below is the description of ECV index equation:

$$ECV = \frac{(Rs+(255-Rr)}{255} + \frac{(Gs+(255-Gr)}{255} + \frac{(Bs+(255-Br)}{255} \quad (1)$$

Rs= Average red value of soybean leave image

Rr= Average red value of white reference image

Gs= Average Green value of soybean leave image

Gr= Average Green value of white reference image

Bs= Average Blue value of soybean leave image

Br= Average Blue value of white reference image

Figure 1 shows the illustration of the "Mata Daun" mobile application usage to estimate soybean leaves nitrogen content. The color correction algorithm implementation in this mobile application could significantly minimize the variation of ECV index caused by the sun light illumination changes. Color correction method was also used by Sumriddetchkajorn [7] to minimize bias of predicted chlorine concentration resulted from their mobile application reading.

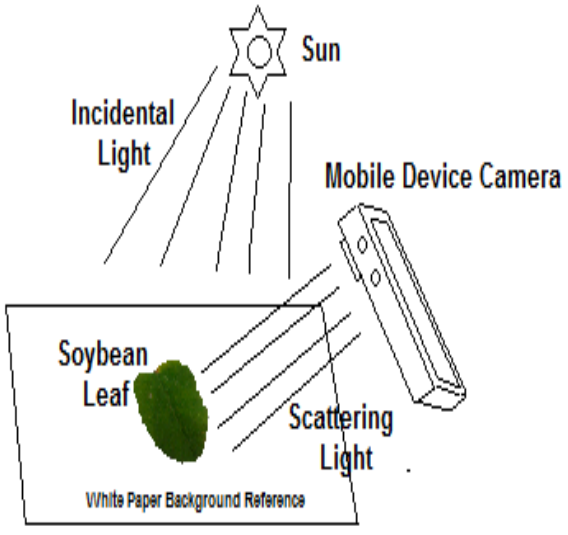


Figure 1. Illustration of the mobile application usage for soybean leaves nitrogen estimation.

III. PERFORMANCE EVALUATION

A. *Enhanced Color Visibility Index Stability Under Different Ilumination Conditions*

An experiment was conducted in shaded and open area which is differ in illumination value (Shaded area=1,266±57 lux, open area illumination =9,833±152 lux) to verify the performance of the color correction mechanism. Test was performed at daylight (09.00 A.M- 11.00 A.M) under trees (shaded area) and under the sun (open area). In this experiment 40 samples were used, thus the total sample was 20 for each conditions. The result in figure 2 shows that predicted nitrogen meter value by ECV index in both area gave relatively the same response (coefficient of variation = 0.05 or 5%). There is no significant difference between ECV in shaded and open area. Through this validation test, the reliability of ECV index in different illumination condition was confirmed.

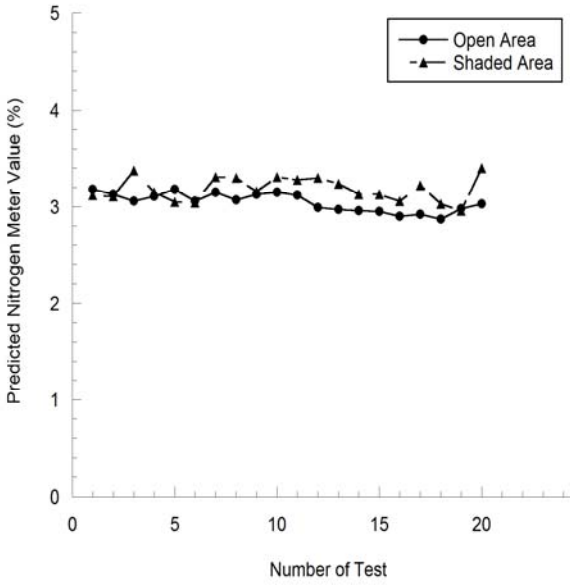


Figure 2.Ilumination changes response of Predicted Nitrogen Meter Value by Enhanced Color Visibility (ECV)

B. *ECV Index and Nitrogen Content Relationship*

In this section, an experiment was conducted to understand the relationship between ECV index and nitrogen content of soybean leaves. 30 soybean leaves were used for this purpose. An Agriexpert CCN-6000 nitrogen meter was used to determine nitrogen content of soybean leaves. Power trend curve was fitted to draw the relationship between nitrogen content obtained from nitrogen meter reading of soybean leaves and ECV index. The power equation is Y=235.04*X^(-1.17) (R^2=0.70). This equation was used as a mathematical model in our mobile application to obtain nitrogen estimation from ECV index.

Although the coefficient of determination of the model is not as high as in "Bai Khao" mobile application [6], it still

show good nitrogen reading value. In the future we plan to add the size samples of our model to improve the accuracy. We also plan to use other approximation methods, instead of mathematical model (i.e. neural network or fuzzy), to enhance our prediction about the relationship between soybean leaves nitrogen content and ECV index. In this research, nitrogen laboratory test for "Mata Daun" result comparison has not been conducted, this comparison research is still in progress in our on going research.

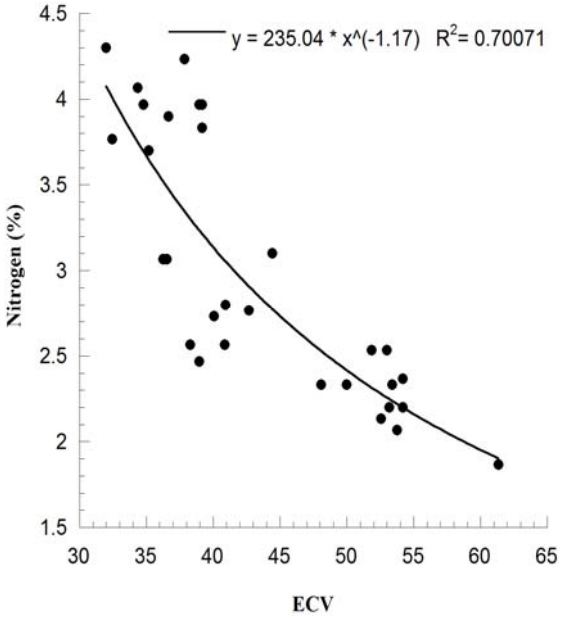


Figure 3. Relationship model of nitrogen meter CCN-6000 Agriexpert reading and Enhanced Color Visibility (ECV) Index

IV. FIELD TEST

To know the performance of our mobile application on the field, a field test experiment was conducted. 20 soybean leaves samples were used to obtain the correlation value between predicted value and real value of soybean leaves nitrogen content reading. The image acquisition process on the field is shown in figure 4, while the result page is shown in the figure 5. The mobile application result page shows the value of predicted soybean leaves nitrogen. We also designed this mobile application with an embedded web service function. In this way users could save the reading result on the database and access those results anytime as long as it is connected with internet. For future development, we plan to expand this mobile application capability to give a suggestion to farmer about their soybean plant condition based on the mobile application reading. This development is intended to give the farmer a solution about their plants continuously. Based on the experience of "Mata Daun" users on the field, users need bigger and simple result screen which is easier to read on the field. As the result, figure 6 shows fairly strong relationship between predicted and real soybean leaves nitrogen content. It shows that the coefficient of determination is 0.93 which is proved the reliability of mobile application reading on the field. For the next development we will combine this test with fertilizer treated plot to enhance the accuracy of the reading result.

Figure 4. Image acquisition process using "Mata Daun" mobile application : acquisition process on field (left) and screenshot of the result (right)

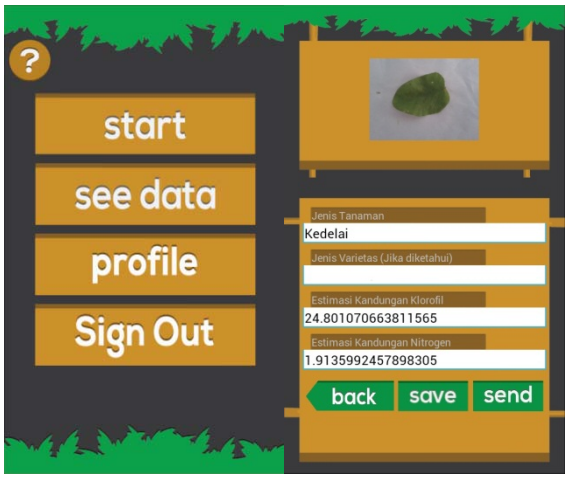


Figure 5. Screenshot of "Mata Daun" mobile application: main page (left) and result page (right)

Based on our field test, to get the best nitrogen meter value prediction with "Mata Daun", it is recommended to use this mata daun version only with Panderman, Anjasmoro and Kaba Soybean variety. The others variety maybe will deliver different result regarding to different physical characteristics such as leaf thickness which is directly connected with soybean leaf nitrogen concentration. Beside the soybean variety, the position of the user while take a leaf picture also give slightly different result. It is also recommended to take a picture 20 cm above the soybean sample leaf. A stable measurement position will give a stable prediction.

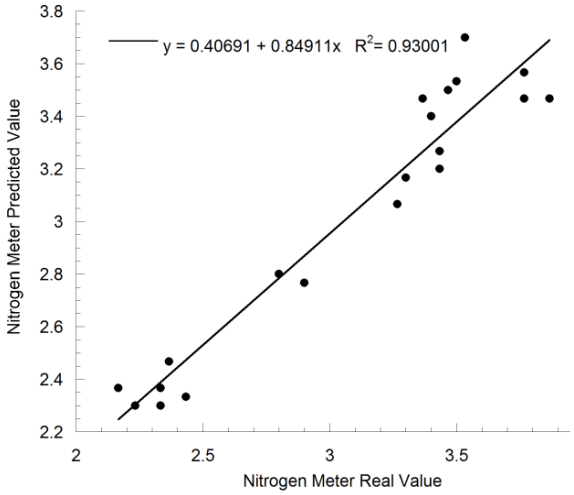


Figure 6. Correlation of nitrogen meter CCN-6000 Agriexpert predicted and real value

V. CONCLUSIONS

This paper proposes mobile application for the determination of soybean leaves nitrogen content. Our aim is to help farmers to conduct their own plant nitrogen determination test just in time and at low cost with reliable result. Our experiments shows that "Mata Daun" mobile application predicted value was reliable to be used at the field. With the strong relationship between predicted value and real value these mobile application can be as an alternative for conventional plant leaves nitrogen test. We do hope with the application of this software by the farmer, their productivity could be increased.

ACKNOWLEDGMENT

Authors would like to thank to Lembaga Penelitian dan Pengabdian Masyarakat (LPPM), Universitas Ma Chung for sponsored this research with grant number 002/MACHUNG /LPPM MRGV/X/2 012.

REFERENCES

[1] Shaahan, M.M.; A.A. El-SAyed, E.A.A; Nour, A.E.1999. Predicting Nitrogeb, Magnesium and Iron Nutritional Status in Some Perennial Crops Using a Portable Chlorophyll Meter. *Scientia Horticulturae*, **82**:339-348

[2] Fageria, N.K.; Baligar, V.C.; Jones, C.A. 2011. Growth and Mineral Nutrition of Field Crops. CRC Press. New York.

[3] Foth, H.D. 1994. Dasar-Dasar Ilmu Tanah. Penerbit Erlangga. Jakarta

[4] Xian, L.Z.; Xiao-Ping, Y.; Heong, K.L.; Cui. 2007. Effect of Nitrogen Fertilizer on Herbivores and Its Stimulation to Major Insect Pest in Rice. *Rice Science*, **14**(1): 56-66.

[5] Hoffland, E.; Jeger, M.J.; Beusichemn, M.L. 2000. Effect of Nitrogen Supply Rarte on Disease Resistance in Tomato Depends on The Pathogen. *Plant and Soil*, **218**:239-247.

[6] Intaravanne, Y.;Sarun S. 2012. BaiKhao (Rice Leaf) App: A Mobile Device-based Application in Analyzing the Color Level of the Rice Leaf for Nitrogen Estimation in Optoelectronic Imaging and Multimedia Technology. Ed. Tsutomu Shimura; Guangyu Xu; Linmi Tao; Jesse Zheng. Proc. Of SPIE vol. 8558.

[7] Sumriddetchkajorn, S.; Kosom C.; Yuttana I. 2013. Mobile Device-based Self-referencing Colorimeter for Monitoring Chlorine Concentration in Water. *Sensor and Actuators B*, **182**:592-597.

Refactoring Rules Effect of Class Cohesion on High–Level Design

Arwin Halim
Department of Computer Science
STMIK Mikroskil
Medan, Indonesia
arwin@mikroskil.ac.id

Petrus Mursanto
Enterprise Computing Laboratory
Universitas Indonesia
Depok, Indonesia
santo@cs.ui.ac.id

***Abstract*—Various refactoring rules and their impact on class cohesion at high-level designs have been investigated. Early measurement and improvement of internal quality attributes such as cohesion, allows developers to avoid extensive review, frequent modification and rewriting of specifications, designs, and source codes. The impact of refactoring is obtained by comparing class cohesion of a design before and after refactoring applied. A set of class cohesion metrics that suits four properties of a good class cohesion metric was chosen based on theoretical analysis and supported by empirical evidence. We found that refactoring rules does not always improve class cohesion values.**

Keywords—class cohesion; high-level design; object-oriented; refactoring; software quality

I. INTRODUCTION

Refactoring is a technique to improve quality of object-oriented software without altering external behavior of the system. It is focused on improving internal structure of the system instead [1]. The key of refactoring is redistributing classes, variables, and methods of the system to facilitate future adaptations and extensions [2]. Refactoring is useful at higher levels of abstraction than source code and low-level design [2]. The process of refactoring is usually accomplished subjectively and has no clear explaination on the effect of quality factor [3,4]. Therefore, we need to know proper refactoring rules to obtain the desired quality.

Cohesion is an important internal object-oriented software quality attribute. Cohesion shows the relationship and interaction between elements in class. The metric is called Class Cohesion. Class Cohesion can be used as an early indicator of defects in class design [5]. Higher value of class cohesion results in more understandable, modifiable, and maintainable design. Many experimental results show that high cohesive system must have low coupling in consequence [6].

Metric is an indicator that pair certain characteristics on observed entity into a numeric value. The benefit of metric depend on what will be accomplished form the results of measurements performed, for example selecting the most optimal software design that fit predetermined criteria [7]. Class cohesion metric has been proposed in several studies. Most applications are available to calculate metrics based on source code, it is available only at the end of development though [8]. Measurement of quality in the early design phase can reduce the development of an extensive review, modification, and rewriting the specification, design, and source code [8]. Cohesiveness can be measured based on the high-level design and low-level design [9]. To analyze the effect of refactoring on cohesion of the high–level design, we apply the metric proposed by Dallal and Briand [9]. With several refactoring techniques described in [1,10], we assessed the quality improvement obtained.

This paper is organized in the following structure. Section 2 describes related research on this topic. Section 3 describes the metrics used to calculate cohesion value on high-level design and refactoring techniques applied. Section 4 presents a sample case of investigation and discussion of the results obtained. Section 5 describes the conclusion and future work.

II. RELATED WORK

Some of the catalog that contains refactoring techniques have been introduced by Fowler [1,10] and Kerievsky [4]. These catalogs contain a guide to implement the refactoring. The detailed picture of restructuring and refactoring formally from the practical standpoint has been proposed by Bart and Mens [2]. Research related to refactoring and quality factor can be seen in [3,4,11-14]. The studies investigate the refactoring effect on software quality such as internal and external quality attributes.

One of the internal quality attributes is cohesion. Several metrics have been developed to measure the value of cohesion. Measurement cohesion in class design can be performed on high-level design (HLD) and low-level design (LLD). Measuring class cohesion at the high-level design conducted at the design phase has been done several researchers.

Briand et al [8] proposed Ratio of Cohesive Interaction (RCI) metric that based on data declarations and subroutines. The metric consider only the data to data (DD) and data to subroutine (DS) interactions. The RCI metric is defined as the ratio of the number of cohesive interactions of a module to the total number of possible cohesive interactions. The RCI metric does not take the indirect interactions into account. This metric do not consider a method invocation interactions and inheritance relations.

Dallal [15] proposed metric based on distance between pairs of attributes to compute their degree of similarity. This metric considers three types of interactions, i.e. method-method interaction, attribute-attribute interactions, and attribute-method interactions. Interactions between class members constructed from UML class diagram that available at high-level design phase. This metric do not consider an indirect interaction and method invocation interaction. Dallal and Briand [9] proposed the Similarity-based Class Cohesion (SCC) metric. The SCC metric is high-level design-based class cohesion metric based on realistic assumptions, mathematical rules, and can be automated. This metric utilize information from the UML communication diagram and the UML class diagram. This metric supports all types of interactions between class members, both directly and transitively.

Many metrics have been proposed to measure the value of cohesion LLD, such as the Chidamber and Kemerer [16], Hitz and Montazeri [17], Bieman and Kang [18], etc. Bela et al [19] proposed a metric with a conceptual approach to measure the value of coupling and cohesion to predict fault prone on the design. M. Alshayeb [20] examine the effect of refactoring on the five metrics cohesion LLD. This study focused on three refactoring techniques, i.e. extract class, extract method, and extract a subclass. The effect of refactoring is measured by comparing the value of cohesion before and after refactoring on design class. This investigation shows that overall cohesion has been improved by refactoring.

III. Related Work

In this section, we will introduce the high-level of cohesion metrics that will be used in Section III-A, and determine refactoring techniques which may affect the level of cohesion in Section III-B. In this paper, refactoring techniques will be applied to every class and will be measured cohesion values before and after the refactoring, then compare the results to determine the impact of refactoring.

A. Identifying High Level Cohesion Metrics

Briand et al [21] introduced the four properties that serve as a guide to develop a good cohesion measure that consists of non-negative and standardization; minimum and maximum; monotonicity, and cohesion will not be increased when two modules are combined.

Dallal research [22] revealed that only 42% of cohesion metrics are measured in the course of a study that meet the criteria of a good cohesion metric properties. One of the high level of cohesion metric that meets all the criteria above property is shown in the paper Dallal & Briand [9].

Dallal & Briand [9] introduced a Similarity-based Class Cohesion (SCC) used during the High Level Design (HLD) of the object-oriented software, based on the theory, and strongly supported by empirical evidence. This metric apply two types of matrices that can be formed with class diagram and communication diagram. That diagrams were found in high-level design with Unified Modeling Language (UML). The UML information will be used to produce four types of interaction between class member:

1) Direct Method Invocation (DMI) matrix indicate the interaction between method and explicitly defined in the UML communication diagram,
2) Method Invocation (MI) matrix indicate the direct and transitive method-invocation interactions and is derived from the DMI matrix,
3) Direct Attribute Type (DAT) matrix indicate the use of the attribute types instead of the method parameters and extracted from the UML class diagram,
4) Attribute Type (AT) matrix indicate the interaction of methods and attribute types directly and transitively, and is derived from the DAT and MI matrices.

Let k is the number of methods and l is the number of distinct attribute types. The SCC metric utilize the MI and AT matrices to describe the interaction as follows.

1) Method-Method through Attributes Cohesion (MMAC). The MMAC metric can be calculated by Equation 1.

$$MMAC(C) = \begin{cases} 0 & \text{if } k = 0 \text{ or } l = 0, \\ 1 & \text{if } k = 1, \\ \frac{\sum_{i=1}^{l} x_i(x_i-1)}{lk(k-1)} & \text{otherwise} \end{cases} \quad (1)$$

where x_i is the number of 1s in the *i*th column of the AT matrix.

2) Atribute-attribute Cohesion (AAC). The AAC value can be calculated by Equation 2.

$$AAC(C) = \begin{cases} 0 & \text{if } k = 0 \text{ or } l = 0, \\ 1 & \text{if } l = 1, \\ \frac{\sum_{i=1}^{k} y_i(y_i-1)}{kl(l-1)} & \text{otherwise.} \end{cases} \quad (2)$$

where y_i is the number of 1s in the *i*th row of the AT matrix.

3) Atribute-Method Cohesion (AMC). The AMC value can be calculated by Equation 3.

$$AMC(C) = \begin{cases} 0 & \text{if } k = 0 \text{ or } l = 0, \\ \frac{\sum_{i=1}^{k} \sum_{j=1}^{l} AT_{ij}}{kl} & \text{otherwise.} \end{cases} \quad (3)$$

where AT_{ij} is the value at row i and column j of the AT matrix.

4) Method-method Invocation Cohesion (MMIC). The MMIC value can be calculated by Equation 4.

$$MMIC(C) = \begin{cases} 0 & \text{if } k = 0, \\ 1 & \text{if } k = 1, \\ \frac{\sum_{i=1}^{k} \sum_{j=1, i \neq j}^{k} MI_{ij}}{k(k-1)} & \text{otherwise.} \end{cases} \quad (4)$$

where MI_{ij} is the number of 1s in the ith column of the AT matrix.

5) Similarity-based Class Cohesion (SCC). Based on MMIC, MMAC, AAC and AMC matrix, the SCC matrix can be calculated by Equation 5.

$$SCC(C) = \begin{cases} 0 & \text{if } k = 0 \text{ and } l \geq 1 \\ 1 & \text{if } k = 1 \text{ and } l = 0 \\ MMIC(C) & \text{if } k > 1 \text{ and } l = 0 \\ \frac{k(k-1)[MMAC(C)+2MMIC(C)+l(l-1)AAC(C)+2lkAMC(C)}{3k(k-1)+l(l-1)+2lk} & \text{otherwise} \end{cases} \quad (5)$$

These metrics are not used to measure the class cohesion with maximum precision, but as expected when an initial assessment and decision. It can be used as an early indicator of quality to detect any defect.

B. *Identifying Refactoring Rules*

Fowler et al [1,10] has explained 72 refactoring techniques and has grouped them into 6 categories. We will choose the refactoring rules that represent each category and have indicated an influence on class cohesion. The selection process is based on the interaction refactoring rule attributes, methods, and attributes the direct method and transitive.

The following refactoring rules are grouped in six categories and may have impact on class cohesion in the design stage.

1) Composing method represented by:
   a) Extract method: turn the fragment into a method whose name explains the purpose of the method.
   b) Inline method: put the method's body into the body of its callers and remove the method.
   c) Replace temp with query: extract the expression into a method, replace all references to the temp with the expression. The new method can then be used in other methods.
   d) Substitute algorithm: replace the body of the method with the new algorithm.
2) Dealing with generalisation represented by:
   a) Extract superclass: create a superclass and move the common features to the superclass.
   b) Form template method: get the steps into methods with the same signature, so that the original methods become the same. Then you can pull them up.
   c) Pull up constructor body: create a superclass constructor and call this from the subclass methods.
   d) Pull up field: move the field to the superclass.
   e) Pull up method: move the method to the superclass.
3) Making method calls simplers represented by:
   a) Paramaterize method: create one method that uses a parameter for the different values
   b) Replace parameter with method: remove the parameter and let the receiver invoke the method.
4) Moving features between objects represented by:
   a) Introduce foreign method: create a method in the client class with an instance of the server class as its first argument.
   b) Move method: Create a new method with a similar body in the class it uses most. Either turn the old method into a simple delegation, or remove it altogether.
5) Simplifying conditional expressions represented by decompose conditional. This technique will extract methods from the condition, then part and else parts.
6) Organizing data represented by replace magic number with symbolic constant. This technique create a constant, name it after the meaning, and replace the number with it.

IV. CASE STUDY

The following case study illustrates how applying the refactoring rules are shown in section III-B and its impact on class cohesion in section III-A. To measure the cohesion class, a complete annotated UML class diagram is required.

A. *Original Design*

The data used in this research is part of the Long Example Refactoring [23]. This case study is an electricity utility charging. The utility has several kinds of customers (residential and disability) and charges them in different ways. The original UML class diagram shown in Fig 1.

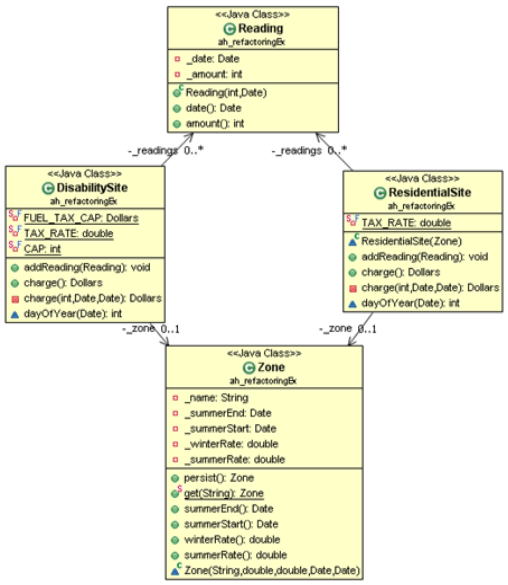


Fig. 1 A Class Diagram from the Preliminary Design

Looking at Fig 2, it is an example of the original UML communication diagram illustrating the message flow between systems objects. When predicting the cohesion of specific class, a complete communication diagram for all classes is not necessary, because the measurements were made for each class individually [9].

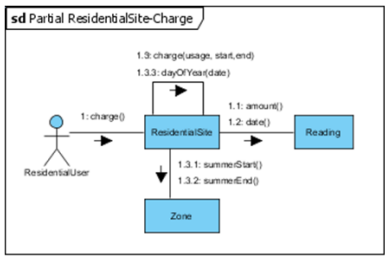


Fig. 2 A Communication Diagram from the Preliminary Design

B. *Refactored Design*

The refactoring techniques applied was adopted from the refactoring catalog which is defined by Fowler et al [1,10]. Fig 3 shows the refactoring version of the preliminary design using refactoring rules in section III-B.

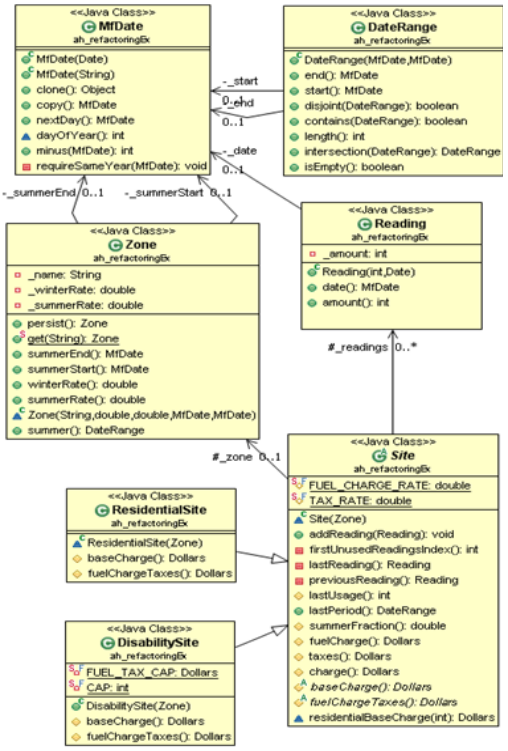


Fig. 3 A Class Diagram from the Preliminary Design

Example of the applied refactorings on the preliminary design:

1) Extract superclass. We create a new class (class Site) and move method addReading from ResidentalSite class and DisabilitySite class to Site class.
2) Form template method. We use this technique because it looks identical method in the ResidentalSite class and DisabilitySite class, but the detail operation is different.
3) Pull up constructor body. We create a superclass (Site class) constructor and invoked from a subclass constructor.
4) Pull up field. We move the zone and readings field from subclass to parent.

For instance, Fig 3 has a difference from the preliminary design of the number of classes, the hierarchy of classes, the number of fields and methods. These changes will alter the message flow between systems objects from the preliminary design, as shown in Fig 4.

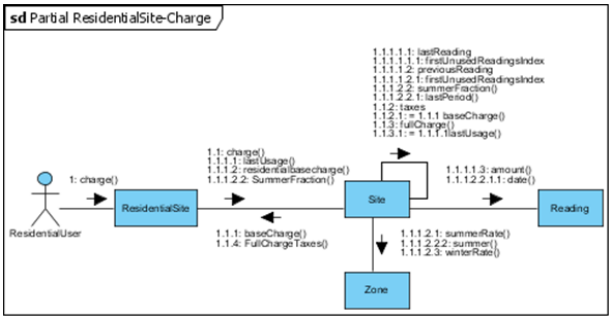


Fig. 4 A Communication Diagram from the Preliminary Design

In this section, the impact of our selected refactoring rules on preliminary design is described. Every implementation of refactoring rules on the design may change the structure, method, or attribute of classes. These changes will be observed by measuring the value of class cohesion for each class that has been changed. The measurement performed at each stage of refactoring. The changing values will indicate the impact of each refactoring rules towards class cohesion on high-level design.

We observed fifty-three results of measurement based on the effect of applying fifteen refactoring rules that have been described in section III-B. Their impact on the value of matrices and metrics is shown in six different ways: '↑' means increase the value, '↗' means may increase the value, '–' means has no effect, '↘' means may decrease the value, '↓' means decrease the value, and '↕' means the effect is varied.

In this part, we will observe the interaction changes that occur due to the implementation of the refactoring rules as shown in Table I and summarize its impact on the categories of refactoring in Table II. The DMI and AT columns indicate the impact of refactoring rules on the number of direct interactions after refactoring. That impact was obtained from the UML diagram after refactoring. The MI and AT columns indicate the number of direct and transitive interaction after refactoring. The both columns will be utilized to measure the cohesion metrics. Looking at Table 1 and Table 2, we notice that for most of the categories of refactoring rules relative increased the interaction between class members.

Table 1 Refactoring Rules Effect on Interaction between Class Member

No	Refactoring rules	DMI	MI	DAT	AT
1	Decompose Conditional	↑	↑	–	↗
2	Etract Method	↑	↑	↑	↑
3	Extract Super Class	–	–	–	–
4	Form Template Method	↑	↑	↗	↗
5	Inline Method	↓	↓	–	↘
6	Introduce Foreign Method	↑	↑	–	–
7	Move Method	↓	↓	↘	↘
8	Parameterize Method	↑	↑	↑	↑
9	Pull Up Constructor Body	–	–	–	–
10	Pull Up Field	–	–	↗	↗
11	Pull Up Method	↑	↑	–	–
12	Replace Magic Number with Symbolic Constant	–	–	–	–
13	Replace Parameter with Method	↗	↗	↘	↗
14	Replace Tempt with Query	↑	↑	↑	↑
15	Substitute Algorithm	–	–	–	–

Table 2 The Category of Refactoring Rules Effect Interaction between Class Members

No	Refactoring category	DMI	MI	DAT	AT
1	Composing methods	↗	↗	↑	↑
2	Dealing with generalization	↑	↑	↗	↗
3	Moving features between object	↗	↗	↗	↗
4	Making method calls simpler	↕	↕	↘	↘
5	Organizing data	–	–	–	–
6	Simplifying conditional expressions	↑	↑	–	↗

Information from MI and AT matrices will be used to measure the SCC metric. The SCC metric is depending on four metrics as shown in section III-A. Table 3 shows the experimental results to illustration the refactoring rules effect on class cohesion. The MMAC, AAC, AMC, and MMIC columns indicate the impact of refactoring rules on the value of metrics that will affect the class cohesion. The last column indicates the cohesion value. This value will be used to determine the result of measurement. Looking at Table 3, the impact of refactoring rules can be grouped into five criteria, i.e. groups which do not affect, increase, decrease, may increase, and may decrease the metric.

Table 3 Refactoring Rules Effect on Class Cohesion

No	Refactoring rules	MMAC	AAC	AMC	MMIC	SCC
1	Decompose Conditional	↘	↓	↘	↑	↘
2	Etract Method	↗	–	↗	↑	↗
3	Extract Super Class	–	–	–	–	–
4	Form Template Method	↓	↘	↓	↓	↓
5	Inline Method	↗	↑	↗	↓	↗
6	Introduce Foreign Method	↓	–	↓	↓	↓
7	Move Method	↑	↗	↑	↑	↑
8	Parameterize Method	↑	↑	↑	↑	↑
9	Pull Up Constructor Body	–	–	–	–	–
10	Pull Up Field	↘	–	↘	–	↘
11	Pull Up Method	↓	–	↓	↘	↘
12	Replace Magic Number with Symbolic Constant	–	–	–	–	–
13	Replace Parameter with Method	↘	↘	↘	↘	↘
14	Replace Tempt with Query	↑	–	↑	↑	↑
15	Substitute Algorithm	–	–	–	–	–

There are four refactoring rules that do not affect the class cohesion, i.e. extract super-class, pull up constructor body, replace magic number with symbolic constant, and substitute algorithm. When we investigated the classes, the extract super-class and pull up constructor body techniques change the class hierarchy. The method on the class hierarchy has been included in the measurement of class cohesion, so it does not give effect. The substitute algorithm technique also does not give any effect. This technique changes the internal flow in the method, but does not change the number of interactions. In this case study, the replace magic number with symbolic constant technique also does not give effect, but this technique has the potential to affect the class cohesion because it may introduce new data type which may often or never accessed by another method.

There are three refactoring rules that increase the class cohesion, i.e. parameterize method, move method, and replace temp with query. When we investigated the classes, the parametrize method and replace temp with query techniques create a new method that has the intensive and complex functionality and accessed at least once. The more intensive and complex functionality of the new method will increase the class cohesion. The intensive and complex functionality of the new methods will increase the number of interactions. A lot of interactions that occur in class will increase the class cohesion. The move method technique also increase the class cohesion. This technique move the method that is using or used by more features of another class, so that method should be moved to the correct class. It can make the classes simpler and highly relatedness of class members.

There are the refactoring rules that decrease the class cohesion, namely introduce foreign method. This technique will create a new method with a specific task and less interaction with other methods. The specific methods rarely accessed, so that the class cohesion is reduce. There are two refactoring rules that may increase the class cohesion, i.e. extract method and inline method. The extract method technique classifies a fragment of interaction into a new method. This method increases the chance that other methods can use a method when the method is finely grained, so that the class cohesion may increase. The inline method technique removed the obvious method that does simple delegation to another method in class. This elimination is not guarantee the class cohesion increasing. If the obvious method and its callers invoke the same method, then the techniques will not increase the class cohesion.

There are five refactoring rules that may decrease the class cohesion, i.e. decompose conditional, form template method, pull up field, pull up method, and replace parameter with method. The form template method, pull up field, and pull up method techniques can only performed in the class hierarchy. When we investigated, the techniques will not affect the class cohesion in the subclass, but has the potential to reduce the class cohesion of the super-class. There is no effect in the subclass, because all inherited methods have been included and the distinct types of all directly and transitively accessible attributes have been consider [9]. The decompose conditional and replace parameter with method techniques make a special method from the condition or parameter. Special method are rarely accessed, so that the class cohesion may decrease. The summary of measurement results of refactoring rules based on categories which has been described in [1,10], towards class cohesion on high-level design can be seen in Table 4.

Table 4 The Category of Refactoring Rules on Class Cohesion

No	Refactoring category	MMAC	AAC	AMC	MMIC	SCC
1	Composing methods	↗	↗	↗	↗	↗
2	Dealing with generalization	↘	↘	↘	↘	↘
3	Moving features between object	↕	↗	↕	↕	↕
4	Making method calls simpler	↗	↗	↗	↗	↗
5	Organizing data	–	–	–	–	–
6	Simplifying conditional expressions	↘	↓	↘	↑	↘

The impact of refactoring rules can be grouped into five criteria, i.e. groups which do not affect, may increase, may decrease, and is varied the class cohesion. We notice that refactoring rules that include composing methods and making method calls simpler category may increase the class cohesion value, the dealing with generalization and simplifying conditional expressions may decrease the class cohesion, the organizing data do not affect the class cohesion, and the moving features between object category are not able to predict

the class cohesion. Looking at Table IV, we notice that for most of the classes in the design, the refactoring rules cannot guarantee an increase or decrease in class cohesion. To increase the class cohesion, the implementation of refactoring rules must improve the interaction between attributes and methods, improve frequency of an accessible method, and associate the data type of attributes and methods in the classes.

V. CONCULUSIONS AND FUTURE WORK

We noticed that the refactoring rules effect of class cohesion on high-level design at hand varied immensely and showed no consistent trends. There are fifteen refactoring rules that are applied in fiftythree measurements. We found that refactoring rules relatively increase interaction between class members. The interaction is not a guarantee of improvement in class cohesion. In addition of no effect, our investigation shows that refactoring rules may improve class cohesion in certain categories and may reduce in others. Having considered the results, we are unable to validate that refactoring rules as whole practice improves class cohesion.

In particular, our assessment has shown that there are four techniques have no affect, three techniques increase, one technique reduces, two techniques may increase and five techniques may decrease the class cohesion. Based on group of refactoring rules, our assessment has shown that there are two categories may increase, two categories may decrease, one category has no affect, and one category is not able to predict the class cohesion.

Further investigation is required to include development of tool for automatic metrics calculation and summary in an understand form. Future work is gaining practical experience with applying the refactoring rules in order to identify the most useful or frequent rules, and measure more systems in the different context to validate this results. In term of productivity, this results are contradictive to the conclusion of previous work by Alshayeb [20]. However, the previous papers addressed to the case with limited implementation of refactoring rules. Altogether, we believe that our findings are particularly relevant, as this work is a case study in free context and free domain by applying more refactoring rules.

REFERENCES

[1] Fowler, M. and Beck, K. and Brant, J. and Opdyke, W., ”Refactoring: Improving the Design of Existing Code,” Harlow, England: Addison-Wesley, 2002.

[2] Bois, B.D. and Gorp, P.V. and Amsel, A. and Eetvelde, N.V. and Stenten, H. and Demeyer, S. and Mens, T., ”A Discussion of Refactoring in Research and Practice,” IEEE Transactions on Software Engineering, vol. 27, no. 6, pp. 513–530, 2002.

[3] Bois, B.D. and Mens, T., ”Describing the impact of refactoring on internal program quality,” International Workshop on Evolution of Large-scale Industrial Software Applications (ELISA), pp. 37–48, 2003.

[4] Bryton, S. and Abreu, F.B. e, ”Strengthening Refactoring : Towards Software Evolution with Quantitative and Experimental Grounds,” Four International Conference on Software Engineering Advances, pp. 570–575, 2009.

[5] Joshi, Padmaja and Joshi, Rushikesh K., ”Cohesion Analysis, Concept Analysis,Refactoring,” 13th European IEEE Conference on Software Maintenance and Reengineering 2009, pp. 237–240, 2009.

[6] Chen, Zhenqiang and Zhou, Yuming and Xu, Baowen and Zhao, Jianjun and Yang, Hongji, ”A Novel Approach to Measuring Class Cohesion Based on Dependence Analysis,” Proceedings of the International Conference on Software Maintenance (ICSM02), pp. 377–383, 2002.

[7] Halim, A. and Sudrajat, A. and Sunandar, A. and Arthana, I. K. R. and Megawan, S. and Mursanto, P. , ”Analytical Hierarchy Process and PROMETHEE Application in Measuring Object Oriented Software Quality,” IEEE International Conference on Advanced Computer Science and Information Systems, pp. 165–170, 2011.

[8] Briand, Lionel C and Morasca, Sandro and Basili, Victor R, ”Defining and Validating Measures for Object-Based High-Level Design,” IEEE Transactions on Software Engineering, vol. 25, no. 5, pp. 722–743, 1999.

[9] Dallal, Jehad Al and Briand, Lionel C., ”An Object-Oriented High-Level Design-Based Class Cohesion Metric,” Information and Software Technology, Elsevier , vol. 52, no. 12, pp. 1346–1361, 2010.

[10] Fowler, M. ”Refactoring home page”, [Online] Available: http://www.refactoring.com. 2012

[11] Alshayeb, M., ”Empirical investigation of refactoring effect on software quality,” Information and Software Technology, vol. 51, pp. 1319–1326, 2009.

[12] Alshammari, B. and Fidge, C. and Corney, D., ”Security Metrics for Object-Oriented Class Designs,” Ninth International Conference on Quality Software, pp. 11–20, 2009.

[13] Mealy, E. and Carrington, D. and Strooper, P. and Wyeth, P., ”Improving Usability of Software Refactoring Tools,” Object oriented Programming Systems, languagaes, and applications, ACM, 2006.

[14] Moser, R. and Abrahamsson, P. and Pedrycz, W. and Sillitti, A. and Succi, G., ”A case study on the impact of refactoring on quality and productivity in an agile team,” Balancing Agility and Formalism in Software Engineering, pp. 252 – 266, 2008.

[15] Dallal, J.A., ”A Design-Based Cohesion Metric for Object-Oriented Classes,” Journal of Computer Science, vol. 1, pp. 195–200, 2007.

[16] Chidamber, Shyam R. and Kemerer, Chris F., ”A Metrics Suite for Object Oriented Design,” IEEE Transactions on Software Engineering, vol. 20, no. 6, pp. 476–493, 1994.

[17] Hitz, M. and Montazeri, B., ”Measuring Coupling and Cohesion In Object-Oriented Systems,” Proceedings of the International Symposium on Applied Corporate Computing, pp. 25–34, 1997.

[18] Bieman, J.M. and Kyoo, Kang,B., ”Cohesion and Reuse in an Object-Oriented System,” Proceedings of the 1995 Symposium on Software reusability, vol. 11, pp. 259–262, 1995.

[19] Ujhazi, B. and Ferenc, R. and Poshyvanyk, D. and Gyimothy, T., ”New Conceptual Coupling and Cohesion Metrics for Object-Oriented Systems,” Working Conference on Source Code Analysis and Manipulation, 2010.

[20] Alshayeb, M., ”Refactoring Effect on Cohesion Metrics,” International Conference on Computing, Engineering and Information, pp. 3–7, 2009.

[21] Briand, Lionel C and Daly, John W and Wust, Jurgen, ”A Unified Framework for Cohesion Measurement in Object-Oriented Systems,” Empirical Software Engineering, vol. 3, pp. 65–117, 1998.

[22] Dallal, Jehad Al, ”Mathematical Validation of Object-Oriented Class Cohesion Metrics,” International Journal of Computers, vol. 4, no. 2, pp. 45–52, 2010.

[23] Fowler, M. ”Chapter 15: A Longer Example,,” [Online] Available: http://www.refactoring.com/rejectedExample.pdf. 2011.

SARIMA (Seasonal ARIMA) Implementation on Time Series to Forecast The Number of Malaria Incidence

Adhistya Erna Permanasari, Indriana Hidayah, Isna Alfi Bustoni
Department of Electrical Engineering and Information Technology
Gadjah Mada University, Jl. Grafika no. 2, Yogyakarta 55281, Indonesia
adhistya@ugm.ac.id, indriana.h@ugm.ac.id, alfiebie@gmail.com

Abstract- **The usefulness of forecasting method in predicting the number of disease incidence is important. It motivates development of a system that can predict the future number of disease occurrences. Fluctuation analysis of forecasting result can be used to support the making of policy from the stake holder. This paper analyses and presents the use of Seasonal Autoregressive Integrated Moving Average (SARIMA) method for developing a forecasting model that able to support and provide prediction number of diasease incidence in human. The dataset for model development was collected from time series data of Malaria occurrences in United States obtained from a study published by Centers for Disease Control and Prevention (CDC). It resulted SARIMA $(0,1,1)(1,1,1)_{12}$ as the selected model. The model achieved 21,6% for Mean Absolute Percentage Error (MAPE). It indicated the capability of final model to closely represent and made prediction based on the Malaria historical dataset.**

Keywords – disease forecasting; time series; SARIMA.

I. INTRODUCTION

Disease forecasting is one of an important area of medical area. It also supports policies making by stakeholders such as health services and healthcare needs [1]. There is no single approach to disease forecasting, and so various methods have often been adopted to forecast aggregate or specific health conditions. Meanwhile, there are no specific methods to match the choices of disease forecasting approaches that are often applied. The selection of an appropriate method is important to achieve a better prediction.

Time series analysis regarding forecasting model is widely used in various fields such as energy demand prediction, economic field, traffic prediction, and in the health support. Indeed, predicting the number of disease incidence need to be focused because the obtained result is needed for further decisions.

Many researchers have developed different forecasting methods to predict human. The number of dengue cases in a population was projected using SARIMA model [2]. The model was developed based on the reported monthly incidence of dengue from 1998 to 2008, and then it was validated using data in 2009. The selected model showed that the monthly cases could be predicted using one, two and twelve months prior. This model indicated that the number of prediction was close to the historical data.

Monthly number of onchocerciasis data in Mexico was analysed using software R [3]. Data was collected from Mexican Secretariat of Health between 1988 and 2011. ARIMA model was constructed to estimate onchocerciasis cases for two years ahead. The results reported a decreasing trend of the disease over time.

The number of human incidence of Schistosoma haematobium at Niono, Mali was projected online by using exponential smoothing method [4]. The method was used as a core of a proposed state-space framework. Data was collected from 17 community health center in the range of 1996 to 2004. The final framework could assist to manage and to assess the transmission and intervention impact of S. haematobium.

Three different methods were used to forecast the SARS epidemic in China [5]. The existing time series was computed by AR(1), ARIMA(0,1,0), and ARMA(1,1). The result of this study was used to monitor the dynamic of SARS in China based on the daily data. Hence, the result could be used to support the disease reports.

A Bayesian dynamic model also could be used to monitor influenza surveillance as one factor of SARS epidemic [6]. This model was developed to link pediatric and adult syndromic data to the traditional measures of influenza morbidity and mortality. The findings showed the importance of modeling influenza surveillance data, and recommend dynamic Bayesian Network.

The monthly data of Cutaneous leishmaniasis (CL) incidence in Costa Rica from 1991 to 2001 was computed by using seasonal autoregressive models. This work was studying the relationship between the interannual cycles of the diseases with the climate variables using frequency and time-frequency techniques of time series analysis [7]. This model supported the dynamic link between the disease and climate.

The application of additive decomposition method was used to predict Salmonellosis incidence in US [8]. This method was selected because of the relatively constant trend of the historical data. Fourteen years historical data from 1993 to 2006 was collected to compute the forecast values until 12 months-ahead.

This paper analyses the empirical results for evaluating and predicting the number of zoonosis incidence by using Autoregressive Integrated Moving Average (ARIMA). This model is selected because of the capability to correct the local trend in data, where the pattern in the previous period can be used to forecast the future. Thus, this model also supports the modeling of one perspective as a function of

time [9]. Due to the seasonal trend of time series used, the Seasonal ARIMA (SARIMA) is selected for the model development.

The remainder of the paper is structured as follows. Section II introduces basic theory of SARIMA. Section III describes time series data collection. Section IV reports model development. Finally, Section V presents conclusion.

II. SEASONAL ARIMA (SARIMA)

This section describes the basic theory of Autoregressive Integrated Moving Average (ARIMA). The general class of ARIMA (p,d,q) comes from three parts: d is the level of differencing, p is the autoregressive order, and q is the moving average order [10]. The ARIMA model is shown in (1) as

$$z_t = \delta + \emptyset_1 z_{t-1} + \emptyset_2 z_{t-2} + \cdots + \emptyset_p z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}^{+} \quad (1)$$

where z_t is level of differencing, the constant is notated by δ, while ϕ is an autoregressive operator, a is a random shock corresponding to time period t, and θ is a moving average operator.

Seasonal ARIMA (SARIMA) is used when the time series exhibits a seasonal variation. A seasonal autoregressive notation (P) and a seasonal moving average notation (Q) will form the multiplicative process of SARIMA as $(p,d,q)(P,D,Q)_s$. The subscripted letter 's' shows the length of seasonal period. For example, in a hourly data time series $s = 7$, in a quarterly data $s = 4$, and in a monthly data $s = 12$.

In order to formalize the model, the *backshift* operator (B) is used. The time series observation backward in time by k period is symbolized by B^k, such that $B^k y_t = y_{t-k}$

Formerly, the backshift operator is used to present a general stationarity transformation, where the time series is stationer if the statistical properties (mean and variance) are constant through time. The general stationarity transformation is presented below:

$$z_t = \nabla_s^D \nabla^d y_t = (1 - B^s)^D (1 - B)^d y_t \quad (2)$$

where z is the time series differencing, d is the degree of nonseasonal differencing used and D is the degree of seasonal differencing used.

Then, the general form of SARIMA model SARIMA (p,P,q,Q) is

$$\emptyset_p(B)\emptyset_p(B^s) z_t = \delta + \theta_q(B)\theta_Q(B^s) a_t \quad (3)$$

George Box and Gwilym Jenkins studied the simplified step to obtain the comprehensive information of understanding ARIMA model and using the univariate ARIMA model [10],[11]. The Box-Jenkins (BJ) methodology consists of four iterative steps:

1) Step 1: Identification.

This step focus on selection of the order of regular differencing (d), seasonal differencing (D), the non-seasonal order of Autoregressive (p), the seasonal order of Autoregressive (P), the non-seasonal order of Moving Average (q) and the non-seasonal order of Autoregressive (Q). The number of order can be identified by observing the sample autocorrelations (SAC) and sample partial autocorrelations (SPAC).

2) Step 2: Estimation

The historical data is used to estimate the parameters of the tentatively model in Step 1.

3) Step 3: Diagnostic checking

Diagnostic test is used to check the adequacy of the tentatively model.

4) Step 4: Forecasting

The final model in step 3 is used to forecast the forecast values.

This approach is widely used to examine the SARIMA model because of the capability to capture the appropriate trend by examining historical pattern. The BJ methodology has several advantages, involving extract a great deal of information from the time series using a minimum number of parameters and the capability in handling stationery and non-stationary time series in non-seasonal and seasonal elements [12],[13].

III. DATA COLLECTION

Malaria disease dataset was selected for model development because these incidences can be found in any country. Malaria is a mosquito-borne infectious disease came from a parasite. It causes symptoms including fever, chills, and flu-like illness. Without further treatment, the patients can progress to coma or death. WHO reported that it was etimated 219 million cases of malaria occurred worldwide with 660,000 death cases in 2010 [14].

This study collected time series data of Malaria occurrences in United State for the 204 month period from January 1993 to December 2009. The data was obtained from the summary of notifiable diseases in United States from the Morbidity and Mortality Weekly Report (MMWR) that published by Centers for Disease Control and Prevention (CDC). The original data is plotted as presented in Fig. 1.

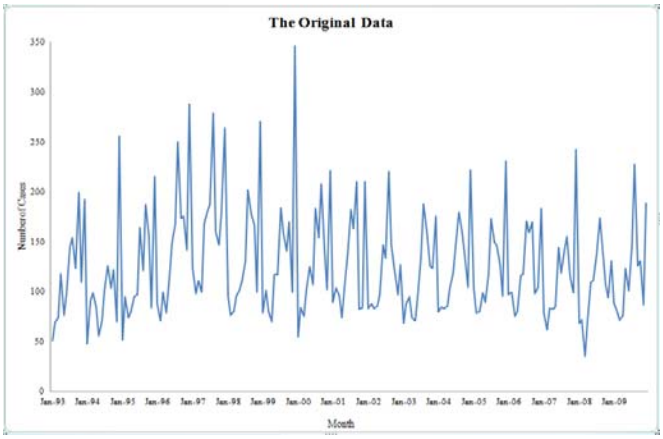


Figure 1. Malaria Original Dataset.

Fig. 1 shows the line chart of the historical data from 1993 until 2009. The x axis represents the specific year while the y axis represents the associated number of incidence. Since time series plot of the historical data exhibited the seasonal

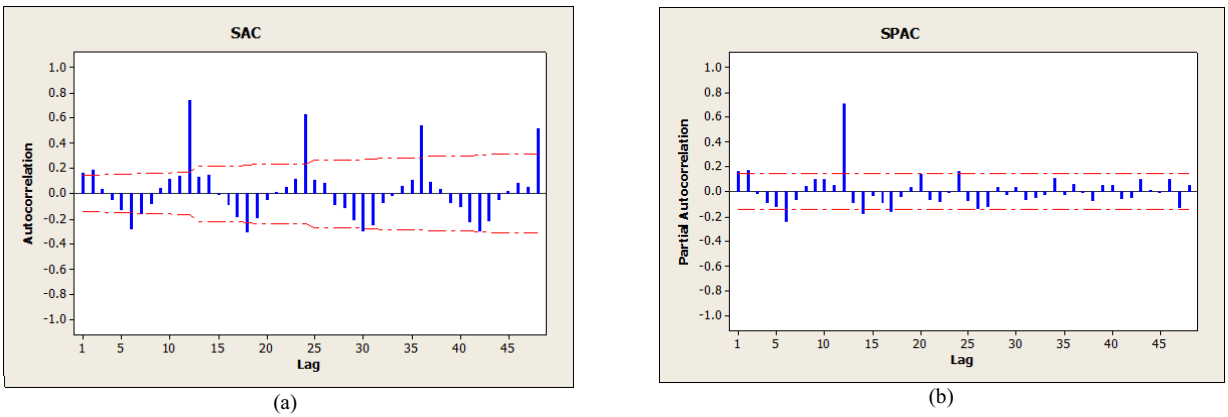
variations which present similar trend every year, then SARIMA was chosen as the appropriate approach to develop a model prediction.

IV. Model Development

This following section discusses the result of BJ iterative steps to forecast the available dataset.

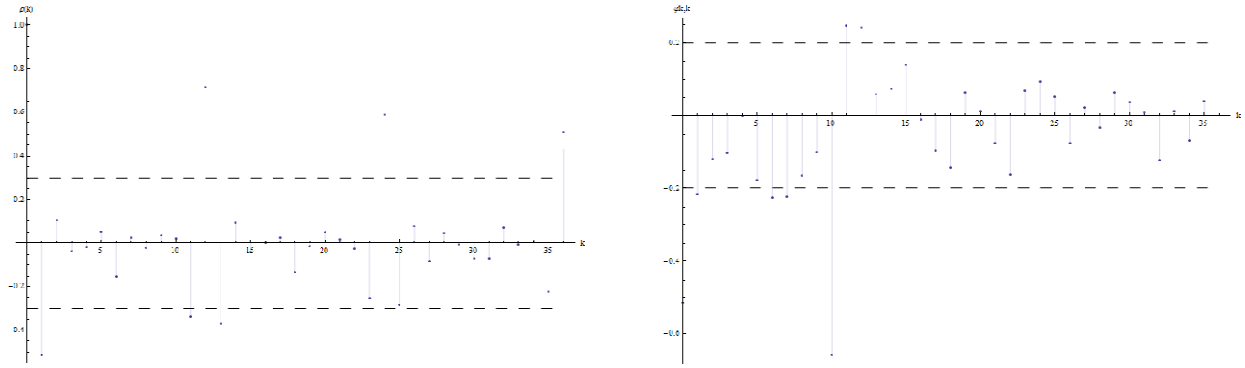
A. Step 1: Identification

The first step in BJ method is time series identification. It is done by plotting sample autocorrelations (SAC) dan sample partial autocorrelations (SPAC) based on the original data. In this model development, four periodical data was selected to illustrate the plot. Fig. 2 shows SAC and SPAC correlogram of the original data. Based on these figures, it can be observed that the correlogram of time series is likely to have seasonal cycles especially in SAC where some lags have values > |0.2|. Then, it implies level non-stationary. Furthermore, three variations of differencing are applied to the original time series. They are the regular differencing, the seasonal differencing, the reguler and seasonal differencing (Fig. 3, 4, and 5) respectively. In these figures, x-axis shows number of lags, where the value of each lag is showed by y-axis.



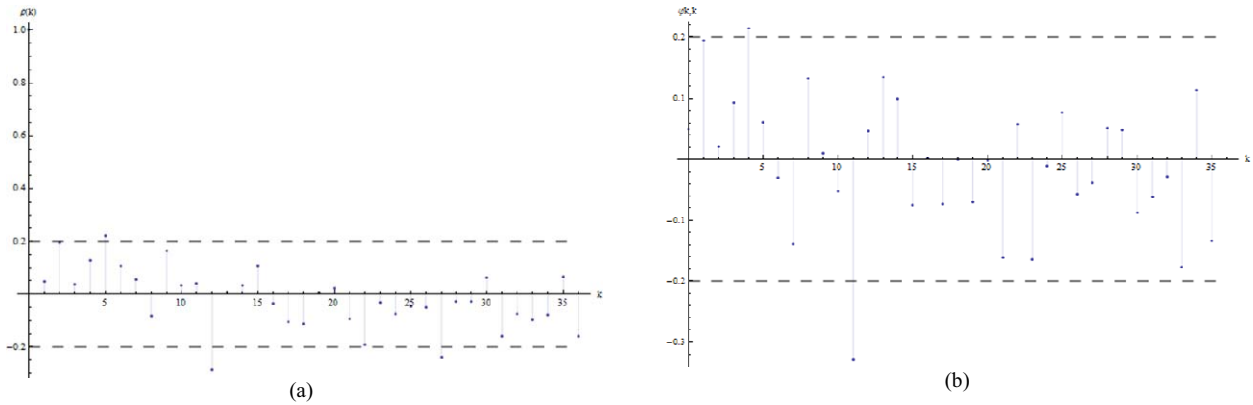
(a) (b)

Figure 2. (a) SAC Correlogram of the Original Data; (b) SPAC Correlogram of the Original Data.



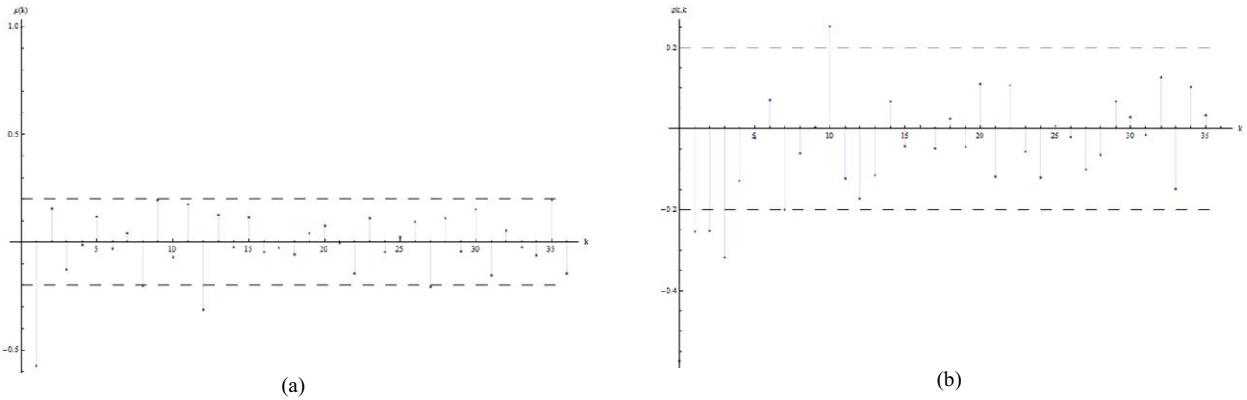
(a) (b)

Figure 3. (a) SAC Correlogram of Regular Differencing; (b) SPAC Correlogram of Regular Differencing.



(a) (b)

Figure 4. (a) SAC Correlogram of Seasonal Differencing; (b) SPAC Correlogram of Seasonal Differencing.



(a) (b)

Figure 5. (a) SAC Correlogram of Regular and Seasonal Differencing; (b) SPAC Correlogram of Regular and Seasonal Differencing.

Selection of whether to use regular or seasonal differencing was based on the correlogram. In order to develop ARIMA, time series should be stationary. The observation of correlogram Fig. 3, 4, and 4 indicates that more spikes are found in the regular differencing and seasonal differencing. Then, the regular and seasonal differencing is chosen for the model development.

B. Step 2: Estimation

This step aims to determine the suitable model based on the observation on step 1. To produce the model, the separated non-seasonal and seasonal model was computed first. It was followed by combining these models to describe the final model.

- Nonseasonal level: SAC cuts off at lag 1 and SPAC dies down. Then, the tentatively moving average nonseasonal q model:
  $$z_t = \delta + a_t - \theta_1 a_{t-1} \quad (4)$$
- Seasonal level: SAC *dies down* and SPAC dies down. It results the combination model of autoregressive-moving average seasonal orde (P,Q):
  $$z_t = \delta + \emptyset_{1,12} z_{t-12} + a_t - \theta_{1,12} a_{t-12} \quad (5)$$
- Since the final model consists of nonseasonal and seasonal level, then equation (4) and (5) are combined to get the final model as shown in equation (6):
  $$z_t = \delta + \emptyset_{1,12} z_{t-12} + a_t - \theta_1 a_{t-1} - \theta_{1,12} a_{t-12} \quad (6)$$
  Rewrite (6) yields (7)
  $$z_t = \delta + a_t + \emptyset_{1,12} z_{t-12} - \theta_1 a_{t-1} - \theta_{1,12} a_{t-12} \quad (7)$$

Values $\theta_1 a_{t-1}$and $\theta_{1,12} a_{t-12}$ from (7) are unified to form the multiplicative term (8):

$$(-\theta_1)(-\theta_{1,12}) a_{t-13} = \theta_1 \theta_{1,12} a_{t-13} \quad (8)$$

The form $-\theta_1$ and $-\theta_{1,12}$ are multiplied and added the negative numbers (-1 and -12) after the *t* in the random shock subscripts *a*. Then, the overall tentatively model become (9):

$$z_t = \delta + a_t + \emptyset_{1,12} z_{t-12} - \theta_1 a_{t-1} - \theta_{1,12} a_{t-12} + \theta_1 \theta_{1,12} a_{t-13} \quad (9)$$

The final SARIMA model has arrived to: $(0,1,1)(1,1,1)_{12}$

Next, the model was used to compute the estimation output of model coeffisient using Minitab software. It produced the following results: MA(1) = 0.9561; MA(2) = -0.1010; SMA(12) = 0.9174 and Constant = -0.04831.

C. Step 3: Diagnostic checking

SAC and SPAC were calculated using residuals values of the original data. It yielded the autocorrelation values that closed to |0.2| and within the 95% confidence interval. It indicated that the selected model was fit.

D. Step 4: Forecasting

The final SARIMA model $(0,1,1)(1,1,1)_{12}$ was used to forecast the values of the 36 months-ahead from t_{205} through t_{240} as presented in Table I. While, the whole forecasting plot is shown in Fig. 6.

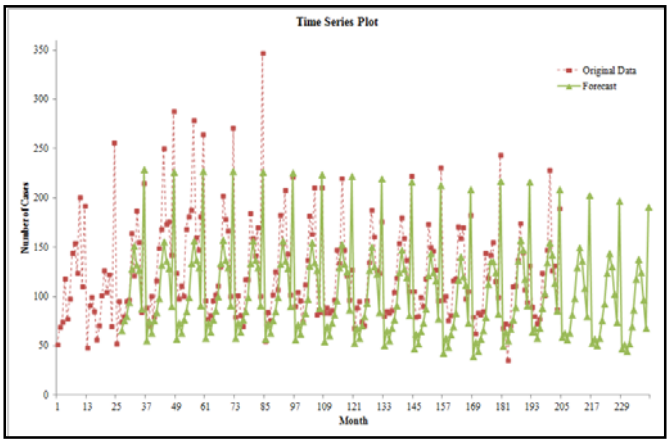


Figure 6. Time Series Plot of The Original Data and Forecast Values.

TABLE I. THE FORECASTING RESULT.

Month	Prediction		
	2010	2011	2012
January	70.482	62.913	56.920
February	71.205	65.744	59.703
March	64.899	59.390	53.301
April	71.001	65.443	59.306
May	91.320	85.714	79.529
June	104.908	99.254	93.020
July	137.838	132.135	125.853
August	164.604	158.853	152.522
September	141.918	136.119	129.740
October	115.832	109.985	103.558
November	87.110	81.214	74.738
December	200.933	194.989	188.466

E. Error Measures

The accuracy of the forecasting can be evaluated using error measures. It is achieved by comparing the original data and the forecast values. In this paper, Mean Absolute

Percentage Error was used as the error measure. The result showed MAPE value for the selected model was 21.6%. Thus, the empirical result indicated that the model was able to accurately represent the Malaria historical dataset.

V. CONCLUSION

The prediction of the future incidence of disease is important to make a better policy. In this paper, the use of forecasting method was applied to predict the number of Malaria incidence in US based on monthly data. The adjusted model prediction was developed by using SARIMA model based on the historical data. SARIMA model can be obtained by using four iteratively Box-Jenkins steps and provide the prediction of the number of human incidence in other zoonosis to help the stakeholder make further decisions. The result indicate that SARIMA $(0,1,1)(1,1,1)_{12}$ was the fit model. The model was also be able to represent the historical data with MAPE value 21.6%. A further work is still needed to evaluate and apply other forecasting methods into the time series in order to obtain better accuracy of forecast value.

REFERENCES

[1] I. Soyiri and D. Reidpath, "An overview of health forecasting," *Environmental Health and Preventive Medicine,* vol. 18, pp. 1-9, 2013/01/01 2013.

[2] E. Z. Martinez*, et al.*, "A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil," *Revista da Sociedade Brasileira de Medicina Tropical,* vol. 44, pp. 436-440, 2011.

[3] E. E. Lara-Ramírez*, et al.*, "Time Series Analysis of Onchocerciasis Data from Mexico: A Trend towards Elimination," *PLoS Negl Trop Dis,* vol. 7, p. e2033, 2013.

[4] D. C. Medina*, et al.*, "State–Space Forecasting of Schistosoma haematobium Time-Series in Niono, Mali," *PLoS Neglected Tropical Diseases,* vol. 2, pp. 1-12, 2008.

[5] D. Lai, "Monitoring the SARS Epidemic in China: A Time Series Analysis," *Journal of Data Science,* vol. 3, pp. 279-293, 2005.

[6] P. Sebastiani*, et al.*, "A Bayesian Dynamic Model for Influenza Surveillance," *Statistics in Medicine,* vol. 25, pp. 1803-1825, 2006.

[7] L. F. Chaves and M. Pascual, "Climate Cycles and Forecasts of Cutaneous Leishmaniasis, a Nonstationary Vector-Borne Disease," *PLoS Medicine,* vol. 3, pp. 1320-1328, 2006.

[8] A. E. Permanasari*, et al.*, "Forecasting of Zoonosis Incidence in Human Using Decomposition Method of Seasonal Time Series," in *NPC 2009*, Tronoh, Malaysia, 2009, pp. 1-7.

[9] E. S. Shtatland*, et al.*, "Biosurveillance and outbreak detection using the ARIMA and logistic procedures," in *SUGI 31*, Cary, NC: SAS Institute, Inc, 2006, pp. 197-31.

[10] B. L. Bowerman and R. T. O'Connell, *Forecasting and Time Series An Applied Approach*, 3rd ed.: Duxbury Thomson Learning, 1993.

[11] S. Makridakis and S. C. Wheelwright, *Forecasting Methods and Applications*: John Wiley & Sons. Inc, 1978.

[12] J. G. Caldwell. (2006, *The Box-Jenkins Forecasting Technique*. Available: http://www.foundationwebsite.org/

[13] C. Chia-Lin*, et al.*, "Modelling and forecasting tourism from East Asia to Thailand under temporal and spatial aggregation," *Math. Comput. Simul.,* vol. 79, pp. 1730-1744, 2009.

[14] WHO. (2013, June 16). *Malaria Fact Sheet*.

Selecting Suitable Solution Strategies for Classes of Graph Coloring Instances Using Data Mining

Nur Insani
Mathematics Education Department
Yogyakarta State University
Indonesia
nurinsani2001@yahoo.com

Kate Smith-Miles
School of Mathematical Sciences
Monash University
Australia
kate.smith-miles@monash.edu

Davaatseren Baatar
School of Mathematical Sciences
Monash University
Australia
davaatseren.baatar@monash.edu

***Abstract*—The Maximal Independent Set (MIS) formulation tackles the graph coloring problem (GCP) as the partitioning of vertices of a graph into a minimum number of maximal independent sets as each MIS can be assigned a unique color. Mehrotra and Trick [5] solved the MIS formulation with an exact IP approach, but they were restricted to solving smaller or easier instances. For harder instances, it might be impossible to get the optimal solution within a reasonable computation time. We develop a heuristic algorithm, hoping that we can solve these problems in more reasonable time. However, though heuristics can find a near-optimal solution extremely fast compared to the exact approaches, there is still significant variations in performance that can only be explained by the fact that certain structures or properties in graphs may be better suited to some heuristics more than others. Selecting the best algorithm on average across all instances does not help us pick the best one for a particular instance. The need to understand how the best heuristic for a particular class of instance depends on these graph properties is an important issue. In this research, we use data mining to select the best solution strategies for classes of graph coloring instances.**

Keywords—graph coloring; integer linear programming; algorithm selection; data mining; algorithm footprint

I. INTRODUCTION

Many practical problems can be formulated as combinatorial optimization problems, which can then frequently be expressed naturally in terms of graphs and as integer linear programs. One of the most famous problems in graph theory is graph coloring. Graph coloring, or proper coloring, has been a popular research topic since its introduction to solve the map coloring problem more than 150 years ago. Consider an undirected graph $G = (V, E)$ where V is a set of vertices (nodes) and E is the set of edges, with $|V| = n$ and $|E| = m$. The Graph Coloring Problem (GCP) is defined as coloring the vertices $v \in V$ such that no two adjacent vertices $(u, v) \in E$ are assigned the same color. The most common type of graph coloring seeks to minimize the number of colors for G.The minimum number of colors with which the vertices of G can be colored is called the chromatic number of G, denoted by $\chi(G)$. The coloring problem is then to determine $\chi(G)$ and to find a coloring of G that uses $\chi(G)$ colors. GCP can be formulated in many different ways.

A formulation of GCP based on maximal independent sets was introduced by Mehrotra and Trick [5], and it is known as the Maximal Independent Set (MIS) formulation. They formulated the GCP as a problem of partitioning the vertices of a graph into a minimum number of maximal independent sets as each maximal independent set can be assigned a unique color. The MIS formulation does not have a symmetry problem like the standard model because it does not involve assigning a specific color to a vertex but it simply defines which vertices receive the same color. One of the advantages of the MIS formulation is the lower bound provided by the linear relaxation of the formulation is at least as good as the one provided by the standard formulation, and probably better. The computational results show that Mehrotra and Trick's approach can consistently solve instances with up to 70 vertices for random graphs and 250 vertices for random geometric graph. Thus, they have successfully solved two NP-hard problems, the Maximum Weight Independent Set (MWIS) problem and an Integer Programming (IP) pricing

sub-problem, and so far it gives the most efficient exact methods to solve GCP of small to medium sizes.

However, the fact that the method involves solving two NP-Hard problems means that it will not scale up to larger instances. Previous experiments showed that most of the time of the column generation technique is spent in solving the pricing sub-problem, while solving the master problem of LP relaxation takes a small fraction of the whole time. The MIS formulation introduced by Mehrotra and Trick has some interesting properties: it does not have a symmetry problem: and a vertex $v \in V$ is not dominated if and only if the corresponding constraint is facet defining such that the linear relaxation of the formulation provides good lower bound. These properties make it attractive and raise some interesting questions such as: Can a feasible (optimal) solution to LP relaxation lead us to a good feasible (or even optimal) solution to the GCP? What is the gap between the best integral solution obtained from an optimal solution of the LP relaxation and optimal solution of GCP? How does one construct a good integral solution efficiently from the optimal solution of the LP relaxation? Can the final pricing sub-problem used to get a good approximate solutionto the GCP? Hence to answer the above questions, we propose four different strategies to solve the GCP by exploiting the fractional solution of the LP relaxation. They are LP-IP which solves the LP relaxation exactly and constructs an integral solution using the exact IP method, LPapprox-IP which solves the LP relaxation approximately and constructs an integral solution using the exact IP method, LP-intHeuristic which solves the LP relaxation exactly and constructs an integral solution using a rounding heuristic, and LPapprox-intHeuristic which solves the LP relaxation approximately and construct an integral solution using a rounding heuristics.

Although heuristics might be able to find good feasible solutions extremely fast compared to exact approaches, there is still significant variation in performance that can only be explained by the fact that certain structure properties in graphs may be better suited to some heuristics more than others. The need to understand how the best heuristic for particular classes of instances depends on these graph properties is an important issue. Time and complexities can be minimized by knowing which algorithm to use for a particular instance rather than testing all algorithms by trial and error. Selecting the best algorithm on average across all instances does not help us pick the best one for a particular instance. Wolpert and Macready's [15] No-Free-Lunch (NFL) theorems tell us that if an algorithm does particularlywell on one class of problems then there are likely to be other classes of instances where it will perform poorly. Instances arising from some applications are actually not as hard as the theory predicts; while others can be just as bad as the worst case; so it is valuable to understand how an algorithm performs on various kinds of instances. The features of instances can help determine this, and some recent work [13] has begun to show how the features of graphs determine the performance of graph coloring heuristics like DSATUR [1] and tabu search [4]. In this research, we extend such analysis to understand how our four proposed approaches above performs on average across all test instances, as well as to perform data mining analysis to determine how the features of graphs affect algorithm performance. Specifically, we determine if there are particular classes of instances where one approach dominates the others, and aim to characterize the features of such instances. Thus, by using the powerful facet-defining MIS formulation for GCP, the goals of this research are firstly to find out if computational efficiencies can be obtained by using the LP relaxation to yield integer feasible solutions, and whether the LP needs to be solved exactly or if an approximation is sufficient. Likewise, we determine if the integer feasible solution needs to be solved exactly from the LP relaxation, or does a heuristic rounding procedure yield the same results? We will compare the proposed heuristic algorithms to general purpose heuristics commonly used for graph coloring, and downloaded from Joe Culberson's graph coloring web resources page [3]. We will also study the features of a collection of the graphs and determine if these features are predictive of algorithm performance for classes of instances. All of the computational experiments will use the DIMACS benchmark instances and some additional randomly generated instances (geometric and random graphs) from generators available from Culberson's web site [17].

II. PRELIMINARY CONCEPTS

A. Formulation of GCP

In this section, we introduce the classical Integer Linear Programming models for GCP. Let C be an upper bound on the number of colors. Then a straight forward ILP model for GCP can be defined using $(C + 1)|V|$ binary variables [2]. The standard formulation is

$$min \sum_{c=1}^{C} y_c$$

s.t.

$$\sum_{c=1}^{C} x_{vc} = 1 \qquad \forall v \in V \qquad (1)$$

$$x_{vc} + x_{wc} \leq y_c \qquad \forall (v,w) \in E, \forall c \in \{1,2,\dots C\} \qquad (2)$$

$$x_{vc}, y_c \in \{0,1\} \qquad \forall v \in V, \forall c \in \{1,2,\dots C\} \qquad (3)$$

where $x_{vc} = 1$ if color c is assigned to vertex v and $x_{vc} = 0$ otherwise. The binary variables y_c indicate whether color c is used in some vertices, i.e. $y_c = 1$ if $x_{vc} = 1$ for some vertex v. Constraint (1) ensures that each vertex receives exactly one color, while constraint (2) ensures that adjacent vertices have different colors. Finally constraint (3) imposes the variables to be binary.

In [6], Mehrotra and Trick proposed an alternative formulation for the GCP. The formulation relies on the fact that any set of vertices that have the same color is an independent set. In other words, independent sets can be used to represent the set of vertices that have the same color. Based on that the GCP can be formulated as (The Maximal Independent Set Formulation):

$$min \sum_{j \in J} x_j$$

s.t.

$$\sum_{j \in J; v \in S_j} x_j \geq 1 \quad \forall v \in V \tag{4}$$

$$x_j \in \{0,1\} \quad \forall j \in J \tag{5}$$

where J is the index set of all maximal independent sets of G. The binary variables x_j indicate whether the vertices in the maximal independent set S_j could be assigned to the same color or not. In other words, $x_j = 1$ implies that vertices that are in the maximal independent set S_j could have the same color, while $x_j = 0$ implies that the vertices are not required to have the same color. Thus the objective function is to find the minimum number of maximal independent sets that cover all vertices of the graph G. Constraint (4) ensures that every vertex v in the graph must belong to at least one maximal independent set (i.e. must receive at least one color). Constraint (5) states that variable x_j must be binary.

B. Heuristics Approach to the GCP

Good heuristic algorithms are essential to tackle hard optimization problems. Using heuristics one might be able to find good feasible solution quickly. Moreover, heuristics are useful to tighten the bounds and consequently to reduce the search space, especially in enumerative approaches such as branch and bound/cut/price. This motivates the large amount of literature concerning the heuristic and metaheuristic approaches for GCP. Some of the most popular heuristics and metaheuristics found in the literature are DSATUR [1] and tabu search [4].

DSATUR is one of the most well-known exact algorithms and is exact for bipartite graphs [1]. It is a sequential vertex coloring that successively colors the vertices sorted in predetermined order based on the saturation degree of a vertex. Saturation degree of a vertex is the number of different colors connected to the vertex. DSATUR works by subdividing the problem into sub-problems. TABU Search (TS) is a higher level heuristic algorithm for solving combinatorial optimization problems. It is an iterative improvement procedure that starts from any initial feasible solution and attempts to determine a better solution. The version of this algorithm proposed by Hertz [4] for graph coloring is referred to as TABUCOL.

C. Algorithm Selection

Research into algorithm performance has led to the question of which algorithm is the most suitable for particular domains and certain instances. In her research, Smith-Miles [12] has formulated the questions: For a particular problem domain, what are the features or characteristics of particular instances of the problem that are likely to correlate with good or bad algorithm performance? Can we model the relationship between these characteristics and algorithm performance? To answer those questions, we can use the algorithm selection framework proposed by John Rice in 1976 [7] as follows: For a given problem instance $x \in P$, with features $f(x) \in F$, find the selection mapping $S(f(x))$ into algorithm space A, such that the selected algorithm $\alpha \in A$ minimizes the performance mapping $y(\alpha(x)) \in Y$ [8]. By using this model, there are four essential components that can be derived for graph coloring problems:

1. The problem space P: the input graphs defined by a set of vertices and edges.
2. The feature space F: the features or characteristics of the graph.
3. The algorithm space A: all algorithms that we consider to solve the graph coloring problems.
4. The performance space Y: the minimum number of colors found by an algorithm.

The collection of data describing $\{P, F, A, Y\}$ is known as meta-data [12]. The role of data mining is to learn the relationship between the feature space and the performance space so that performance can be predicted based on the features of an instance alone. In addition, data visualization techniques can be employed to gain insights into the relationships in the high dimensional feature space. Data mining methods will be utilized in this research to provide insights into the strengths and weaknesses of the solution strategies used in the research.

III. SOLUTION STRATEGIES BASED ON LP RELAXATION

We present our approaches and strategies developed to give answers to the research questions. It includes the preprocessing process before we begin the main computation, the pricing problem, and our four different approaches to solve the final solution of LP relaxation to obtain a feasible integral solution. From now on, without loss of generality, we may assume that all vertices in a graph are not dominated. If the graph has dominated vertices then using some preprocessing we can always reduce the CG problem into a smaller problem where no vertices are dominated.

A. LP Relaxation and Column Generation

We developed four greedy heuristics that we use to solve the pricing problem. We use the best solution from these heuristics to generate a new column for the restricted master problem (RMP) if the reduced cost is negative. If the heuristics fail, we solve the MWIS exactly. However, we terminate the solver whenever it finds a column with negative reduced cost. In this way, MWIS problem is solved exactly only to prove the optimality. The four greedy heuristics are Heuristic Based on the Maximum Weight, Heuristic Based on the Net Marginal Weight, Heuristic Based on the Weight Ratio, and Heuristic Based on the Weight of Neighborhoods.

B. Generating Feasible Integer Solution for GCP

We introduce the column generation approach that we use to solve the LPrelaxation of the MIS formulation of the GCP in the previous section. From any feasible solution of the LP relaxation we can construct a feasible solution to GCP. In this section, we describe the exact approach that we use to find the best solution for GCP that could be obtained guided by the optimal (most likely fractional) solution of the LP relaxation. Then we introduce some heuristic algorithms that could be used to find a good integral solution using the optimal solution of the relaxed problem.

Exploiting the fractional solution, one might construct different integral feasible solutions to the GCP. The most straight forward approach is by using integer programming. Therefore, after solving the LP relaxation, we remove all columns (independent sets) that are not in the optimal solution from the final restricted master problem. Then we solve it as an integer program. From a computational point of view, this approach might not be a very efficient way to construct an integral solution to the GCP using the fractional solution. However, it will provide us the best solution to the GCP that could be constructed using the fractional solution.

However, for large problems it might be impossible to get the optimal solution within a reasonable computation time. Therefore, we have developed a heuristic algorithm; we refer to it as greedy rounding. The heuristic is computationally efficient and could be used for large problems to get a reasonably good integral solution to the GCP. This greedy rounding constructs integral solutions, guided by the fractional solution from the LP relaxation, using three different heuristic algorithms and provides the best solution as the final feasible integral solution of the GCP. Our three different heuristic algorithms are Heuristic based on Weight, Heuristic based on Number of Cover, Heuristic based on Least Cover.

C. Solution Strategies for GCP

Based on our proposed algorithms above, we use a number of different ways to find an integral feasible solution to the GCP:

1. LP-IP: Solve the LP relaxation exactly and construct an integral solution using the exact method.
2. LPapprox-IP: Solve the LP relaxation approximately and construct an integral solution using the exact method.
3. LP-IntHeuristic: Solve the LP relaxation exactly and construct an integral solution using the rounding heuristic.
4. LPapprox-IntHeuristic: Solve the LP relaxation approximately and construct an integral solution using the rounding heuristic.

We do expect that the quality of the integral solution will be better for the approaches listed higher in the list. Those alternative approaches are naturally derived from the fact that we have two challenging problems: MWIS and final IP problems. Approximation of the LP relaxation can be obtained by terminating the column generation approach early.

IV. NUMERICAL RESULTS

We test our approaches on the benchmark data sets which are widely used by many researchers and on some randomly generated data set. All algorithms are implemented and tested using CPLEX12.01 embedded in C++ on a machine with i5 processor, 3.33GHz, and 4 GBRAM. Throughout this paper, the CPU times are reported in seconds. The instances used for the experiments are DIMACS benchmarks instances taken from ftp://dimacs.rutgers.edu/pub/challenge/graph/ and randomly generated instances. The DIMACS benchmark set includes: artificial graphs, register allocation graphs, random graphs, Leighton graphs and flat graph. All of geometric and random graphs are randomly generated using built-in generator from an open source Python package called Networkx. These generators are available at http://networkx.lanl.gov/reference/. The data set can be divided into two groups: easy and hard. We call it as easy instances as these instances are solved exactly by some approaches proposed in literature in reasonable amount of CPU time. There are 171 instances in this group and most of them can be solved within few seconds. We also have 28 hard instances. We call them hard instances as they are hard to solve exactly in a reasonable amount of time and most of the proposed heuristic approaches in the literature use these instances as a benchmark data set.

Applying each solution strategy to each graph, we can draw some conclusion about the average performance of each strategy. Generally, LP-IP strategy produces the best feasible integral solutions among other heuristics, while the LPapprox-intHeuristic solves the problems fastest compared to the other heuristics. The LP-IP and the LPapprox-IPsolve more problems with the same solution quality compared to DSATUR, while the LP-intHeuristic and the LPapprox-intHeuristic strategies solve more problems with the same solution quality, compared to tabu search.

In terms of trade-off between the solution quality and the computational time, though the LP-IP provides feasible solutions with the smallest gaps to the optimal solution, compared to the other heuristics, with a maximum gap of 20%, this heuristic requires much more time to solve the problems on easy instances. The longest time to solve the easy instances is 1396 seconds and the longest computational time to solve the harder instances is almost 7000 seconds. The LPapprox-intHeuristic approach provides feasible solutions with the maximum gap the same as the solution obtainedby the LP-IP for the easy instances, but it produces larger maximum gap for the harder instances, which is 160%, where the LP-IP only has 150% as the maximum gap. However, the computational time spent to solve the GCP on both types of instances are faster compare to the other heuristics, up to 1407 seconds and 4034 seconds, respectively. Although the third approach, the LP-intHeuristic strategy, solves the easy and harder instances much faster compared to the two previous strategies, where the maximum time spent for easy instances is 1019 seconds and for the harder instances is 3847 seconds (half time from LP-IP time), their feasible integral solutions obtained are bit worse, up to 33% for easy instances and 177% for the harder ones. However, the solutions produced by this strategy are much better compared to the solutions obtained by DSATUR which has the maximum gap of 43% and 128%for easy and harder instances, respectively. The LPapprox-intHeuristic solves all instances with the fastest time, where the maximum time spent for easy instances is 73 seconds and for harder instances is 1005 seconds; they are 95% and 86% faster compared to the first strategy, LP-IP strategy. Unfortunately, its solution qualities are much worse among

other proposed approaches and DSATUR, but they are better than the solution obtained by tabu search and Greedy approaches. Our fourth proposed approach, LPapprox-intHeuristic strategy, gives feasible integral solutions with the maximum gap to the optimal solution up to 45% for easy instances and 190% for harder ones. Therefore, DSATUR is better to solve the harder instances compare to the LPapprox-intHeuristic. The two general heuristics, tabu search and Greedy, give similar solution quality but they have the largest maximum gaps to the optimal solution compared to the others, which are about 71% for easy instances and 260% for harder ones.

To summarize our observations on the seven heuristics above, we formulate the following observations. If the solution quality is the only criterion to be taken into account and an efficient exact IP approach is not available, the use of LP-IP strategy or LPapprox-IP strategy for easy instances is advised, while for the harder instances, the use of DSATUR is advised. However, if both the solution quality and computational time are relevant, we strongly recommend the use of LP-intHeuristic, LPapprox-intHeuristic or DSATUR for easy instances and only LP-intHeuristic strategy or LPapprox-intHeuristic strategy for harder ones. Finally we advise not to use tabu search or Greedy for both types of instances unless no other choice is available.

V. Predicting Algorithm Performance Using Data Mining

There are two types of solution approaches proposed and used in this paper: exact and heuristic. Although heuristics can often find a near-optimal solution extremely fast compared to exact approaches, the performance of any heuristic is verydependent on the properties of the instance. Time and complexities can be minimized by knowing which approach to use for a particular instance rather than testing all approaches by trial and error. However, selecting the best algorithm on average across all instances does not help us pick the best one for a particular instance, as supported by Wolpert and Macready's [15] No-Free-Lunch (NFL) theorems. Even though we have the experimental results for each proposed approaches as well the other heuristics mentioned above, we need to understand more about how our proposed approaches perform on different subsets of test instances. We would like to know if there are particular features of instances that make some approaches perform well or fail. Based on the recent research of Smith-Miles and Lopes [9], exploring the featuresof graph coloring instances to determine if there is a particular class of instances where one approach dominates the others is the final contribution of the research. Part of this work, focused on the general purpose heuristics, has already been published [13], and in this paper we extend the work further to learn about the suitability of the exact and heuristics strategies based upon the MIS formulation of graph coloring. We explore different features of instances that have been described before, to determine which of our proposed approaches, LP-IP, LPapprox-IP, LP-intHeuristic and LPapprox-intHeuristic, as well as other general purpose heuristics, DSATUR, tabu search and Greedy, gives a near-optimal solution for particular classes of instances. Here, data mining can help us to gain a good understanding of the relationship between the features of instances and the performance of algorithms; in this case they are our proposed strategies and three general heuristics: DSATUR, tabu search and Greedy.

Graph features are the properties of a graph that depends on its abstract structures. In this research, graph features play important roles as they have correlation with the performance of graph coloring algorithms. The analysis of different features described below has been performed in our recent work [13] by using a Python interface to the open source software called igraph and an open source library of scientific tools called SciPy.

The 16 graph features we used are: the number of vertices in a graph (denoted by n); The number of edges in a graph (denoted by m); The diameter of a graph; The density of a graph; The average path length; The girth; Mean and standard deviation of node degree; The clustering coefficient; The betweenness centrality; The standard deviation of betweenness centrality; The mean of eigenvector centrality; The standard deviation of eigenvector; The algebraic connectivity; and the graph spectrum. Details can be found in [13].

A. Experimental Meta-Data

We present our experimental meta-data using the framework proposedby Rice in 1976 [16], and adapted to the study of optimization algorithm performance by Smith-Miles [12]. There are four essential components in our meta data that are derived from the GCP:

1. The problem space P: a set of 63 easy instances and 28 harder instances, of which 64 instances are DIMACS benchmark instances and 27 instances are geometric and random instances that are randomly generated using Networkx generator.
2. The feature space F: a set of 16 features or characteristics of graph as described above [13].
3. The algorithm space A: consists of four proposed heuristic approaches: LP-IP, LPapprox-IP, LP-intHeuristic and LPapprox-intHeuristic, and three other general purpose heuristics: DSATUR, tabu search and Greedy,
4. The performance space Y: the minimum number of colors found by an approach after a single run.

The Algorithm Selection Problem, defined by Rice [58], requires us to learn the relationship between the feature space F and the algorithm performance space Y for each algorithm in the algorithm space A, by using the instances in the problem space P as training data. We use a combination of supervised and unsupervised data mining techniques to learn these relationships and gain insights in the next section.

B. Visualizing Instance Space

We begin with a self-organizing feature map (SOFM) to visualize the high dimensional feature space as a two dimensional map of the instance space. Here, our instance space comprises a set of 91 instances, where each of these

instances is defined by 16 features that are related to GCP. We apply a preprocessing procedure which involves two transformations. Firstly, the 16 features are transformed using a logarithmic transformation to improve the normality of the distributions. Secondly, by using a scaling transformation, all features are normalized to [0,1]. We used the software package Viscovery SOMine [14] to generate the SOFM using a rectangular map of approximate ratio 100:77, trained with 42 cycles (complete presentations of all 91 instances). Figure 1 displays the distribution of five classes of instances where the 16 dimensional feature space is projected onto a two dimensional map. The size of each cluster is indicative of how many instances belong to each cluster. Two instances are close to each other in the two dimensional map if their 16 dimensional feature vectors is similar according to the Euclidean distance metric. In practice, we label each class shown in Figure 1 as Class 1, Class 2, Class 3, Class 4, and Class 5 for the blue region, pink region, yellow region, dark green region, and light green region, respectively. The distribution of the 91 instances can be seen in Figure 2.

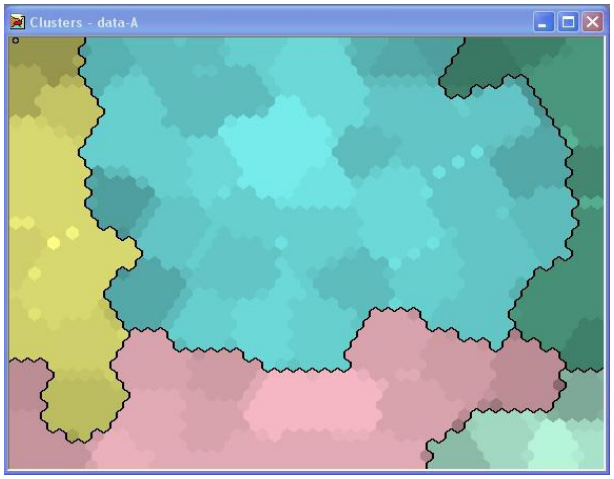


Fig. 1. The distribution of five clusters of instances across instance spaces.

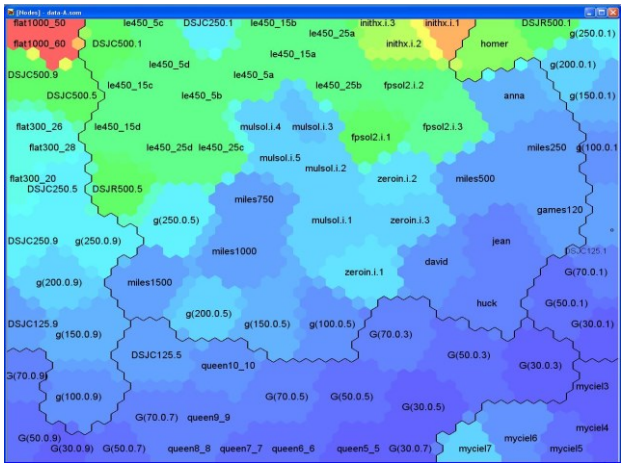


Fig. 2. The distribution of all labelled instances across instance space.

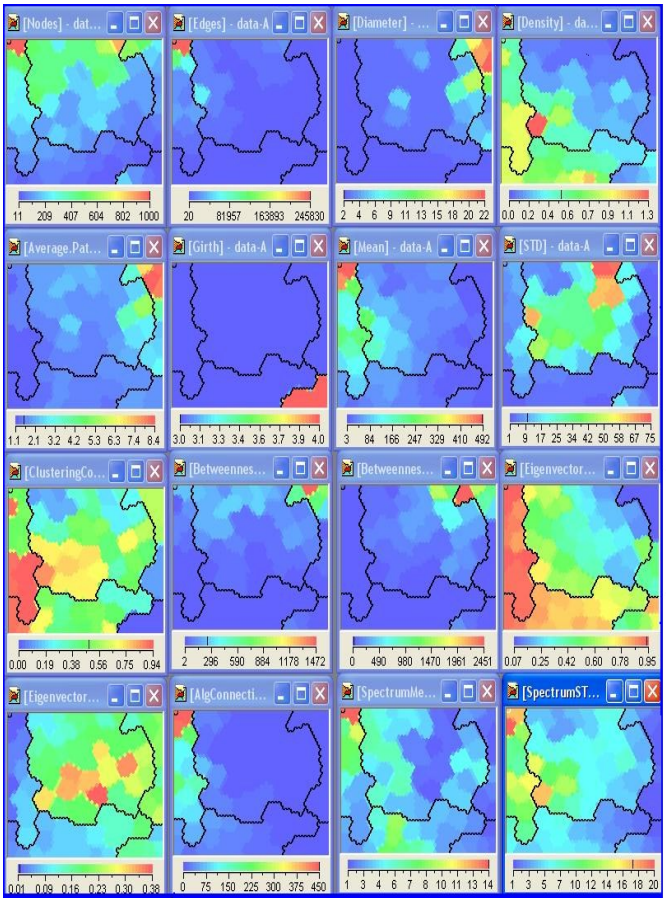


Fig. 3. The distribution of instances features across instace space.

We can determine specific features that are dominant in a class by looking at the distribution of features as seen in Figure 3. The values according to one variable are displayed by the color of the pixel; where warm colors (red, orange, and yellow) represent high values and cool colors (blue) represent low values.

We can see that Cluster 1 contains instances that have a high standard deviation of node degree, a low mean and high standard deviation of eigenvector centrality. Instances with a very low standard deviation of node degree, a low betweenness centrality mean and standard deviation are found in Cluster 2. Cluster 3 contains instances with a medium to high value of mean node degree, a low betweenness centrality standard deviation, high eigenvector centrality mean, a medium to high value of algebraic connectivity, and a medium to high value of spectrum standard deviation. Instances with a low density, a high average path length, and a low spectrum mean are clustered in Cluster 4. We can conclude that the instances that are in Cluster 4 are sparse and close to bipartite since the spectrum mean for bipartite graphs is zero due tothe symmetry of the eigenvalues $\pm\lambda_i$. Cluster 5 contains instances with a high girth and a very low clustering coefficient.

C. Footprints of Algorithm Performance

After the clusters or instance classes are identified, the next step is to examine the algorithm performances across

those instances. Here, the algorithm performances are the performances of our four proposed strategies as well DSATUR, tabu search and Greedy in producing the feasible integral solution, ie. number of colors. The solution gaps to the optimal solution or best known solution (as a percentage) for each algorithm is shown for all instances across the map in Figure 4.

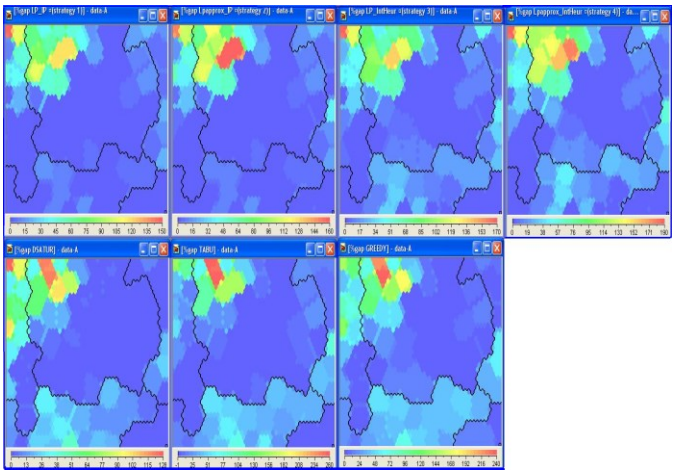


Fig. 4. The performance of each algortihm based on the gap to the optimal solution across instance space. Blue value represent a small optimality gap, while red values represent a poor solution with a high optimality gap.

In Figure 4, the performance of LP-IP, LPapprox-IP, LP-intHeuristic and LPapprox-intHeuristic, are shown in the top map, while the performance of DSATUR, tabu search and Greedy are shown in the bottom map. The warm colors (red, orange, and yellow) represent a high gap between the optimal solution and the solution obtained by each algorithm. Cool colors (blue) represent a low gap between the optimal solution and the solution obtained by each algorithm, thus a very good near-optimal solution. There are many interesting results from this SOFM from which we can draw some conclusions. From the first two maps at the top part, we can see that LP-IP and LPapprox-IP achieve similar performances across most of instances. Both algorithms perform poorly in the half subset of Cluster 3 and small subset of Cluster 1, which are on the top left corner of the maps, while the use of these algorithms are effective for rest of the instances in the map.

There is a small subset of instances from Cluster 1 close to the top left corner where LP-IP and LP-intHeuristic produce a better feasible integral solution rather than those obtained by the LPapprox-IP or the LPapprox-intHeuristic strategies. Among the four proposed heuristics, LP-intHeuristic gives the best feasible solution to those instances, even though the gap is still high. It means that for those instances, it is necessary to solve the LP relaxation exactly. However, DSATUR or tabu search will give the best feasible integral solution among all heuristics for these instances. LP-IP and LPapprox-IP are needed for all instances in Cluster 2 to solve the problem, since the use of LP-intHeuristic and LPapprox-intHeuristic gives slightly worse feasible integral solutions. It means that the final solution of the LP relaxation of these instances needs to be solved as an IP. Likewise DSATUR, tabu search and Greedy will give the worst performances in this instance class. All heuristics give a good performance on the instances in Cluster 5, except for a small pocket in this class where Greedy produced slightly worse integral solutions. If we compare to the Figure 3 in the previous section, we can conclude that all heuristics perform well on instances having small number of nodes, edges, standard deviation, clustering coefficient, betweenness mean, betweenness standard deviation, algebraic connectivity; and medium to high value of eigenvector centrality mean and high girth. DSATUR performs particularly well for the instances in Cluster 4 in the top right, which are close to bipartite, since it is known to be exact for bipartite graphs [1].

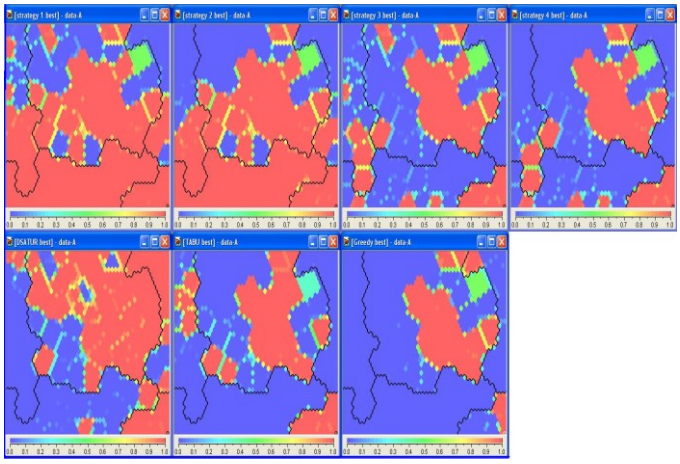


Fig. 5. The distribution of algorithm performance based on their ability to obtain the best solution among other heuristics.

If we convert the optimality gap results into a binary flag to indicate if a given algorithm achieved the best result from the set of 7 considered approaches, then we can show the distribution in Figure 5 of each algorithm's ability to achieve the best performance across the instance space. The warm region shows the algorithm produces the best feasible solution among other 6 algorithms, but it does not mean that this heuristic gave an optimal solution for the problem. Here, the best feasible solution among others is the minimum number of colors that all algorithms can obtain. Blue regions shows the algorithm does not obtain the best solution. The red regions of the instance space show the footprint of each algorithm, a term defined by Corne and Reynolds [16], to show where an algorithm can be expected to perform well. The footprints for LP-IP, LPapprox-IP, LP-intHeuristic and LPapprox-intHeuristic, are shown in the top map, while the footprints for DSATUR, tabu search and Greedy are shown in the bottom map. As seen in Figure 5, the LP-IP and the LPapprox-IP strategies, give the best solutions for most of the instances, where tabu search gives the least number of best solutions among other algorithms and has the smallest footprint. Most of instances in Class 2, also some instances in Class 1 and 3 obtain the best solution when we construct the integral solution exactly while the greedy heuristic is failed to give the best solution. If we take a look closer to Figure 3, those

instances have medium value of density and medium to high value of clustering coefficient. Thus, constructing the integral solution using an exact method is necessary when the graph's density is high as well as if it is well-clustered. We can also see the superiority of DSATUR in the top right corner where DSATUR can produce the best integral solution, while the others cannot. From Figure 3, we can conclude that DSATUR gives the best solution for instances with a high number of vertices, a high number of edges, a high mean, a high algebraic connectivity and a high spectrum mean. In other words, a large non-bipartite graph which is very well connected will obtain the best solution using DSATUR. There is also a small pocket at the bottom of Cluster 1 (in the middle of the map), where LP-IP and LPapprox-IP, tabu search obtains the best solution, while the other algorithms fail. Unfortunately, we cannot observe the specific instance features that make it different from others. For all instances in Cluster 5, any algorithms will give the best solution, except Greedy. Similar to with our previous work [13] focused on the general purpose heuristics; here SOFM is also able to show clear regions where the performance of algorithms is similar. It should be reiterated that we do not use any algorithm performance data as inputs to the clustering process of SOFM, but only superimpose performance data after the clusters have formed. The fact that two instances that are close to each other in feature space and also have similar algorithm performance results (shown by the continuity of the footprints) means that the features that we chose are well suited to characterize the similarities between the instances, and the properties of these features clearly affect the performance of algorithms.

ACKNOWLEDGMENT

The authors would like to thank to Dr. Leonardo Lopes and Brendan Wreford who implemented the codes to generate the instances features.

REFERENCES

[1] D. Brelaz, New methods to color the vertices of a graph, Communications of the ACM, vol. 22, no. 4, p. 251, 1979.

[2] E. Burke, J. Marecek, A. Parkes, and H. Rudova, A supernodal formulation of vertex colouring with its applications in course timetabling, Annals of operations research, vol. 179, no. 1, pp. 105-130, 2007.

[3] J. Culberson, Graph coloring page, URL: http://www. cs. ualberta.ca/ joe/Coloring, 2006.

[4] A. Hertz and D. Werra, Using tabu search techniques for graph coloring, Computing, vol. 39, no. 4, pp. 345-351, 1987.

[5] A. Mehrotra, A column generation approach for graph coloring, INFORMS Journal on Computing, vol. 8, no. 4, p. 344, 1996.

[6] A. Mehrotra and M. A. Trick, A branch and price approach for graph multicoloring, in Extending the Horizons: Advances in Computing, Optimization, and Decision Technologies, ser. Operations Research/Computer Science Interfaces Series, E. K. Baker, A. Joseph, A. Mehrotra, M. A.Trick, R. Sharda, and S. Vob, Eds. Springer US, 2007, vol. 37, pp.15-29.

[7] J. R. Rice, The algorithm selection problem, ser. Advances in Computers, M. Rubino and M. C. Yovits, Eds. Elsevier, 1976, vol. 15, pp. 65 - 118. [Online]. Available: http://www.sciencedirect.com/science/ article/pii/S0065245808605203

[8] K. Smith-Miles, R. James, J. Giffin, and Y. Tu, Understanding the relationship between scheduling problem structure and heuristic performance using knowledge discovery, lncs, Learning and Intelligent OptimizatioN, LION, vol. 3, 2009.

[9] K. Smith-Miles and L. Lopes, Measuring instance difficulty for combinatorial optimization problems, Computers & Operations Research, vol. 39, no. 5, pp. 875-889, 2012.

[10] K. Smith-Miles, J. van Hemert, and X. Lim, Understanding tsp diffculty by learning from evolved instances, in Learning and Intelligent Optimization: 4th International Conference, Lion 4, Venice, Italy, January 2010.Selected Papers, vol. 6073. Springer-Verlag New York Inc, 2010, p. 266.

[11] K. Smith-Miles, Towards insightful algorithm selection for optimisation using meta-learning concepts, in Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on. IEEE, 2008, pp. 4118-4124.

[12] K. A. Smith-Miles, Cross-disciplinary perspectives on meta-learning for algorithm selection, ACM Comput. Surv., vol. 41, no. 1, pp. 1-25, 2008.

[13] Smith-Miles, K.. B. Wreford, L. Lopes and N. Insani, Predicting metaheuristic performance on a graph coloring problems using data mining, in Hybrid Metaheuristics, E.G. Talbi, Ed. Springer, 2013.

[14] V. SOMine, Eudaptics software gmbh.

[15] D. H.Wolpert and W. G. Macready, No free lunch theorems for optimization, Evolutionary Computation, IEEE Transactions on, vol. 1, no. 1, pp. 67-82, 1997.

Semantic Interrelation in Distributed System Through Green Computing Ontology

Herlina Jayadianti, Lukito Edi Nugroho, Paulus Insap Santosa
Department of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
herlinajayadianti@gmail.com, Lukito@ugm.ac.id, insap@jteti.gadjahmada.edu

Carlos Alberta Baptista Sousa Pinto
Department of Information System
Universidade do Minho Guimaraes Portugal
csp@dsi.uminho.pt

Wahyu Widayat
Department Economic Development
Universitas Gadjah Mada
Yogyakarta, Indonesia
wahyu@mep.ugm.ac.id

***Abstract*—Green computing refers to the system that provides minimal impact on the environment. When we are talking about green computing we discuss about how much energy is used by the system, such as energy used by the system, time used for the search process, and how effective the system is. Related to that issue, trough this paper we want to proposes a new effort to achieve Green Computing in heterogeneous data in distributed system. The technology chosen to deal with them is Ontology. We try to generate a common ontology including a common set of terms, based on the several ontologies available, in order to make possible to share the common terminology (set of terms) that it implements, between different communities. If a very large amount of distributed data is not managed and distributed properly, user will need more time to do a search process. The longer the search is done, the more energy is used.**

Keywords—Ontology; Green Computing, Data Heterogeneity; Effectiveness;

I. INTRODUCTION

Support of computer systems have become part of the national infrastructure of each country. Almost the entire national infrastructure has been utilizing computers to support and offer essential and critical services either distributed or not distributed. Problem then appeared if the required data are scattered and are in a place that is different, then of course search process would become longer and takes a significant level of electrical power, thus contributing to increased fuel consumption be. The idea of the green computing has become popular in recent concern, especially if it is linked to the issue of energy crisis. Green computing focuses on how much energy the system is used and how they can make it more efficient. Related to that issue, through this paper we would like to propose a new effort to achieve an efficient search process for distributed heterogeneous data [2].

A distributed system [5] is a collection of autonomous computers linked by a computer network that appear to the users of the system as a single computer. Design issues that arise specifically from the distributed nature of the application are: (1) Transparency, (2) Heterogeneity, (3) Performance, (4) Security, and (5) Openness. In this paper we will focus only in heterogeneity problem, such as: Technical heterogeneity, data model heterogeneity and semantic heterogeneity. Semantic heterogeneity is a general term referring to disagreement about the meaning, interpretation or intended use of the same or related data. This problem is poorly a clear definition of the problem [4], [6], [7]. The importance of being aware of semantic heterogeneity and doing semantic reconciliation is to guarantee meaningful data sharing. The technology chosen to deal with semantic heterogeneity problem is Ontology. Ontologies [8], [9], [10], [11] is a model for determining semantic concepts used by various heterogeneous sources in a way that clearly defined. As more ontologies are built in a different domain, ontology heterogeneity is become another significant problem for the integration of information.

Through this paper we want to prove that through ontology, can make the distribution of the data becomes easier without reducing the semantic meaning. We also want to propose a better solution in searching process to support an energy efficient [3], [18]. The objectives of this paper are to make an easy sharing semantic meaning; and to make the system can understand the tag given by each user. We will show the result trough a small implementation project. This paper is organized as follows: (1) Introduction; (2) Ontology and Green Computation; (3) Interrelation of Semantic Heterogeneity; (4) Implementation; (5) Discussion; (6) Conclusions of work.

II. ONTOLOGY AND GREEN COMPUTATION

Knowledge [1], [15] is people's individual map of the world. Knowledge acquisition involves complex cognitive processes such as perception, communication, and reasoning. According to the knowledge differences, then it is possible for people have a different perception to attain awareness or understand

the environment or reality. There is a group of people in different communities managing data using different perceptions, different concepts, different terms (terminologies), and different semantics to represent the same reality. Perceptions are converted into data, and then saved into separate storage devices that are not connected to each other. Each user – belonging to different communities with large automatic environment - use different terminologies in collecting huge amount of data and as a consequence they also get different results of that exercise. It is not a problem if different results are used for each community. Problems occur if people need to take data from other communities and share it in a collaboration to get a better-quality solution. Ontology is a better technology to solve this problem. We can share different conceptualizations, different terminologies, and different meanings between systems using ontology. Ontology integration [14],[16],[17] and Ontology mapping [7],[8],[14] allows finding correspondences between the concepts of two or more ontologies. If two concepts correspond to each other, then they can also easily "understand" to each other. This is the main reason to bring ontology into green computing area. The effectiveness of search engine processing is closely associated with energy saving.

III. INTERRELATION OF SEMANTIC HETEROGENEITY

In this section, we will illustrate the problem of semantic heterogeneity and how to solve it. As an example there are different groups of people (different communities) have different set of term about object, example term "Hospital" in domain health. "Hospital" is a common term to refer a health facility where patients receive treatment or a medical institution where sick or injured people are given medical or surgical care, but it is not possible if people as an independent human being may use a different terms to refer term "hospital" such as "Infirmary", "medical building", "health facility", "healthcare facility", "clinic", "medical institution", or "extended care facility". (See Fig 1)

Even though each of people use a different "term" to refer "hospital" but in semantic have the same meaning (equivalent) e.g. term "Health facility" is equivalent to term "health care facility". People can easily argue and debate to each other about differences, but the ability of the human can not be replaced easily by computers. Computers need an artificial brain to understand the problems. It is a challenge that must be faced to make computers think like humans. (See Fig 2)

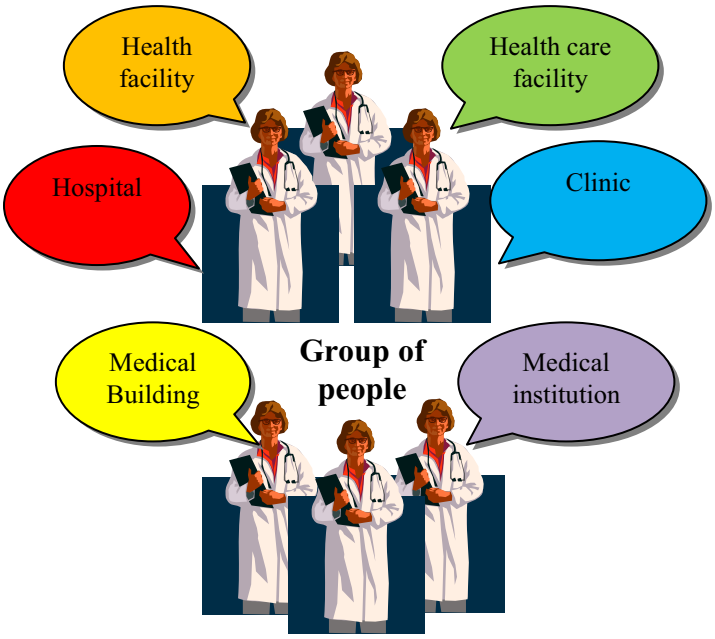


Fig 1. Group of people use different terms to refer the same object

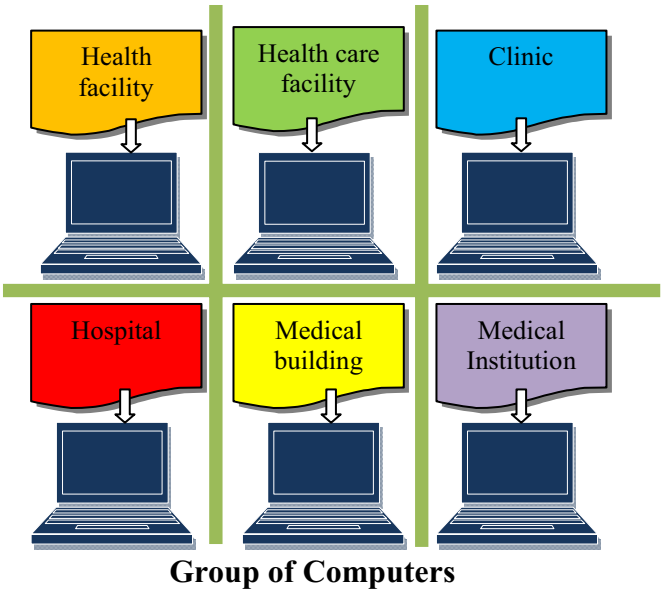


Fig 2. Different terms in different computers

Fig.2 represents several computers in different places that captured reality about the building where patients receive treatment. Let's make a scenario that one day a President as a user want to search data about Hospital in a country that he leads. In his knowledge and experience term "Hospital" is the only one term that he knows about the building where patients receive treatment. In reality each computer use different terms to refer the "building of treatment" (See Fig 2). As an example, one computer use term "Hospital" and another use term "Medical building". To be similar or not equal depend on several factors, such as the programmer's interpretation and the needs of the system itself. Users who

deal with computers has a very important role in controlling and changing the terminology and semantic of the data. Let's see again Fig 1, if a group of people faced differences they will easily discuss and communicate to each other, but what happened if this situation appears in a group of computers. Knowledge representation and ontology are the technologies choose to solve this problem. First step, any data in each computer (Fig 2) are needed to be transferred into ontology (See Fig 3). Second step is mapping terms between ontology.

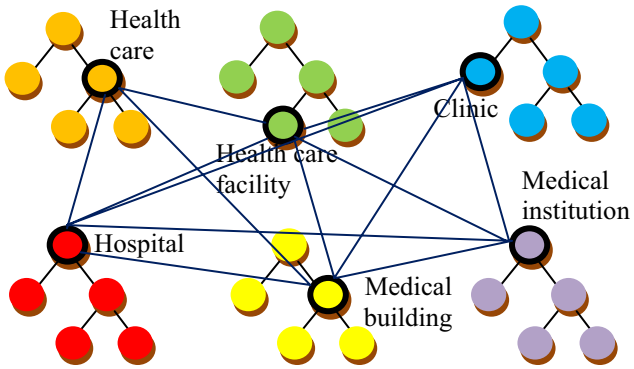


Fig 3. Semantic interrelation

Mapping [16] is the process of relating similar concepts or relations from different sources through some equivalence relation. Mapping allows finding correspondences between the concepts of two or more ontologies. If two concepts correspond, then they mean the same thing or closely related things e.g. "Hospital" from URI: `http://www.semanticweb.org/user/ontologies1/#Hospital` EquivalentTo "HealthCare_facility" from URI:`http://www.semanticweb.org/user/ontologies5#Healthcare_facility"/>`, not only that term `"Hospital"` also EquivalentTo `"Clinic"`, `"Medical institution"`, `"Medical building"` and `"HealthCare"`. See the code in RDF/OWL below:

```
<?xml version="1.0"?>
http://www.semanticweb.org/user/ontologies1/#Hospital
-->
<rdfs:label xml:lang="en">Hospital</rdfs:label>
<owl:equivalentClass
rdf:resource="http://www.semanticweb.org/user/ontolog
ies1#Hospital"/>
<owl:equivalentClass
rdf:resource="http://www.semanticweb.org/user/ontolog
ies2#Medical_institution"/>
<owl:equivalentClass
rdf:resource="http://www.semanticweb.org/user/ontolog
ies3#Infirmary"/>
<owl:equivalentClass
rdf:resource="http://www.semanticweb.org/user/ontolog
ies4#Medical_building"/>
<owl:equivalentClass
rdf:resource="http://www.semanticweb.org/user/ontolog
ies5#Healthcare_facility"/>
```

Ontology mapping allows finding correspondences between the concepts of two or more ontologies. Mapping illustration between ontologies is shows in Fig 4.

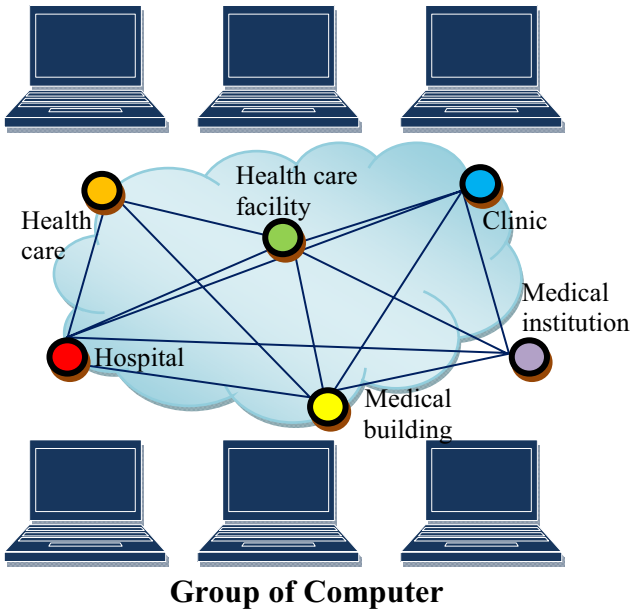


Fig 4. Mapping in the cloud

Ontology mapping can support systems to share different conceptualizations, different terminologies, and different meanings without any problem in differences. Through this reason, we can say that ontology is a good way to solve the problem of differences, especially in data heterogeneity to simplify the search process. The effectiveness of a search process is give a contibute to energy saving. Next section will describe more deatil about semantic integration process.

IV. IMPLEMENTATION

In this section we will discuss about how to implement semantic integration trough several ontologies with Protege[1]. There are five (5) ontologies in same domain – "Health", each ontology refer a "building of treatment" with different terms e.g. Ontology1 : Hospital, Ontology2 : Medical Institution, Ontology3 : Infirmary, Ontology4 : Medical_building and Ontology5 : Healthcare_facility. Using ontology mapping we can make correspond between classes (concepts) in two or more ontology. Fig 5 shows correspond between classes in OWL visualization. OWL is a language for create ontologies to the web. OWL was designed for processing information and to provide a common way to process the content of web information. See Fig 5. `Hospital` is corresponding to `Infirmary`, `Medical building`, `Health facility`, `Healthcare facility`, `clinic`, and `medical institution`; or it can be write that `Infirmary is-a hospital`, `medical building is-a hospital`, `health facility is-a hospital`, `healthcare facility is-a hospital`, `clinic is-a hospital`, `medical institution is-a hospital`, and `extended care facility is-a hospital`. Relation process should be implementing as a complete graph, if `Hospital` is corresponding to `Infirmary` so

[1] http://protegewiki.stanford.edu/wiki/Protege4GettingStarted

Infirmary is corresponding to Hospital, and Infirmary also corresponding to Medical building, Health facility, Healthcare facility, clinic, and medical institution;

```
SELECT ?Health_Building ?type
WHERE { ?Health_Building :Type ?value.
FILTER (?value= 'hospital'). }
```

or with

```
FILTER (?value= 'Medical building').}
```

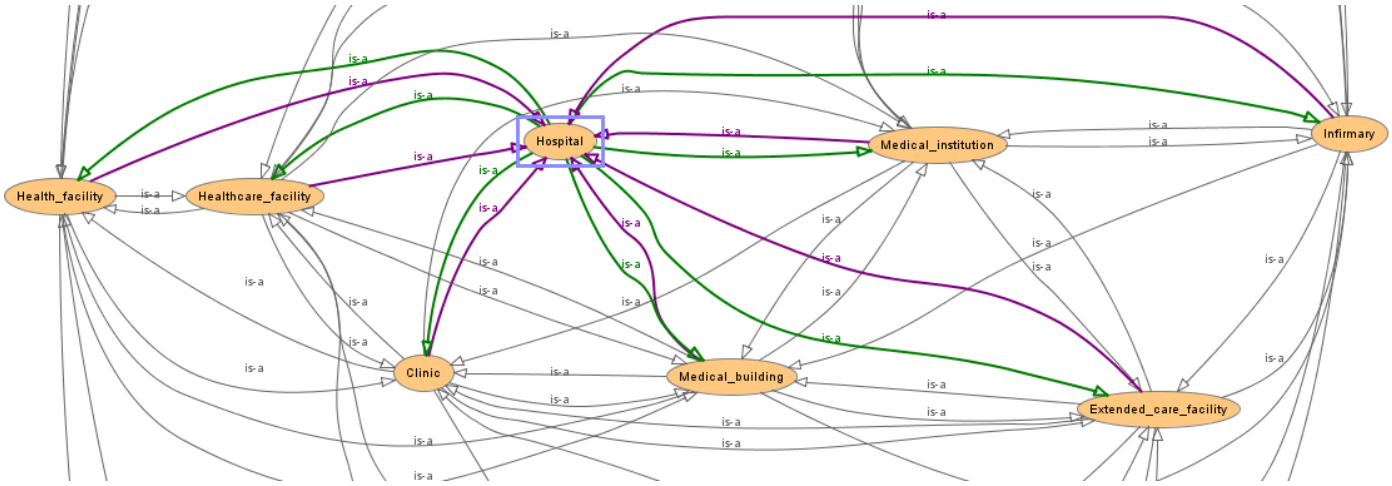


Fig 5. Ontology web language (OWL) visualization in protégé

Individual name of Hospital "Bethesda" shows in RDF syntax below:

```
<owl:NamedIndividual
rdf:about="&Ontologies1;Bethesda">
<rdf:type rdf:resource="&Ontologies1;Hospital"/>
<rdf:type
rdf:resource="Ontologies2;Medical_institution"/>
<rdf:type rdf:resource="&Ontology3;Infirmary"/>
<rdf:type
rdf:resource="&Ontology4;Medical_building"/>
<rdf:type
rdf:resource="&Ontology5;Health_facility"/>
<rdf:type <rdfs:label
xml:lang="en">Bethesda</rdfs:label>
<Name rdf:datatype="&xsd;string">Bethesda</Name>
    </owl:NamedIndividual>
```

Ontologies1 refer "Bethesda" as a hospital and Ontologiest5 refer "Bethesda" as a Medical Building. Hospital is corresponding to Medical building and also Medical building is corresponding to Hospital, so user will do searching with both perceptions.

We use a SPARQL[2] as a graph matching query language to express queries across different data sources.

```
Prefix:<http://www.semanticweb.org/user/ontologies1#
>
PREFIX    rdf:    <http://www.w3.org/1999/02/22-rdf-
syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

[2] http://www.w3.org/TR/rdf-sparql-query/

The SPARQL query will refer the same result : Panti rapih, Bethesda and Sarjito as a building : Hospital or as a Medical building. (See Fig 6).

Health_Building
Panti_Rapih
Sarjito
Bethesda

Fig 6 SPARQL result

Semantic engine process with ontology allows users to use their knowledge and perception in their searching activities. User can refer the same object with different term, such as: Bethesda is-a Hospital or Bethesda is-a Medical Building.

V. DISCUSSIONS

Green computing inspiring organizations to re-evaluate the use of IT resources. They need to find new ways to increase the energy efficiency. One of the efforts is to improve energy efficiency in the computing process especially search process. One of the major problems in the search process is to make the system can understand the perception that they are referring to. They should be able to make system can think like humans. In order to bring an expected result, they need to be organize data and information in a better way.

Various approaches like Ontology and Semantic Web has been created for better searching on the web. Ontology can be used to process knowledge properly. Semantic engine process with ontology allows users to use their knowledge and

perception in their searching activities. As discussed in the previous section; user can refer the same object with different term, such as: `Bethesda is-a Hospital` or `Bethesda is-a Medical Building`, and machines can understand user perceptions.

Base on this reason in this paper we try to propose an ontology for discovering user needs especially to realize a better search engine. Ontologies can help the machine (search engine) realize the definition of the perception in the metadata. Thing we have done is combine/ integrate several ontologies and achieve general views of the users. Ontologies are used in obtaining results according to the user query and reasoning used to meet the user's needs. Ontology can make a seaching process more efficient and effective.

We can say that ontology can support a green computing main objective especially related to the issue making an intelligent search engine. This project is just the beginning project of big ideas about the role of ontology in the area of green computing. Our future work is about green ontology to support a cost reduction in organizations.

VI. CONCLUSIONS

In this paper we try to managing knowledge by using an ontology integration as a process to create a new ontology (Global ontology - Common Ontology) to simplify the search process. On the other hand, green computing refers to the systems efficiently and effectively. Hence in this paper we try to collaborate between these two research areas, because if a very large amount of data are managed and distributed properly, user will need a short time to do a search process. The longer the search is done, the more energy is used.

ACKNOWLEDGMENT

We would like to acknowledge the support of the Erasmus Mundus EuroAsia program for the research foundation of this research, and also to acknowledge Universidade do Minho and Universitas Gadjah Mada for the collaboration

REFERENCES

[1] S. Murugesan, "Harnessing green IT: Principles and practices," *It Prof.*, vol. 10, no. 1, pp. 24–33, 2008.

[2] A. S. Tanenbaum and M. Van Steen, *Distributed systems*, vol. 2. Prentice Hall, 2002.

[3] D. Theodoratos, "Semantic integration and querying of heterogeneous data sources using a hypergraph data model," *Adv. Databases*, pp. 798–805, 2002.

[4] Y. Xue, H. Ghenniwa, and W. Shen, "Ontological view-driven semantic integration in collaborative networks," *Leveraging Knowl. Innov. Collab. Networks*, pp. 311–318, 2009.

[5] A. P. Sheth, "0 Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics," *Interoperating Geogr. Inf. Syst.*, pp. 5–29, 1998.

[6] N. Guarino, *Formal ontology in information systems*. Citeseer, 1998.

[7] J. F. Sowa, *Knowledge representation: logical, philosophical, and computational foundations*, vol. 594. MIT Press, 2000.

[8] C. Diamantini, D. Potena, and W. W. Smari, "Collaborative knowledge discovery in databases: A knowledge exchange perspective," in *Fall Symposium on Semantic Web for Collaborative Knowledge Acquisition. AAAI, Arlington, VA: IEEE Computer Society*, 2006, pp. 24–31.

[9] N. Choucri, *Mapping sustainability*. Springer, 2007.

[10] K. Knight and S. K. Luk, "Building a large-scale knowledge base for machine translation," in *Proceedings of the National Conference on Artificial Intelligence*, 1994, pp. 773–773.

[11] T. Kumazawa, O. Saito, K. Kozaki, T. Matsui, and R. Mizoguchi, "Toward knowledge structuring of sustainability science based on ontology engineering," *Sustain. Sci.*, vol. 4, no. 1, pp. 99–116, 2009.

[12] E. Jiménez-Ruiz, B. C. Grau, I. Horrocks, and R. Berlanga, "Ontology integration using mappings: Towards getting the right logical consequences," in *The Semantic Web: Research and Applications*, Springer, 2009, pp. 173–187.

[13] B. C. Grau, B. Parsia, and E. Sirin, "Ontology integration using ε-connections," in *Modular Ontologies*, Springer, 2009, pp. 293–320.

[14] H. Wache, T. Voegele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. H\übner, "Ontology-based integration of information-a survey of existing approaches," in *IJCAI-01 workshop: ontologies and information sharing*, 2001, vol. 2001, pp. 108–117.

[15] N. F. Noy and M. A. Musen, "Algorithm and tool for automated ontology merging and alignment," in *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-00). Available as SMI technical report SMI-2000-0831*, 2000.

[16] N. F. Noy and M. A. Musen, "The PROMPT suite: interactive tools for ontology merging and mapping," *Int. J. Hum.-Comput. Stud.*, vol. 59, no. 6, pp. 983–1024, 2003.

[17] W. Binder and N. Suri, "Green computing: Energy consumption optimized service hosting," in *SOFSEM 2009: Theory and Practice of Computer Science*, Springer, 2009, pp. 117–128.

[18] "Sustainable Computing: It's Not Easy Being Green," *Computerworld*, 20-Mar-2006. [Online]. Available: http://www.computerworld.com/s/article/109629/Sustainable_Computing_It_s_Not_Easy_Being_Green. [Accessed: 16-Jun-2013].

Student Classification for Academic Performance Prediction using Neuro Fuzzy in a Conventional Classroom

Indriana Hidayah, Adhistya Erna Permanasari, Ning Ratwastuti
Electrical Engineering and Information Techology, Gadjah Mada University
Yogyakarta, Indonesia
indriana.h@ugm.ac.id, adhistya@ugm.ac.id, ratwastuti@gmail.com

Abstract— **Conventional classroom is still the main learning method applied in undergraduate program of Electrical Engineering and Information Technology Department, Gadjah Mada University. There are several problems in this method, such as large amount of students and limited number of meetings making difficult to understand each student. Student classification is a way to solve the problem by mapping the condition of each student based on certain parameters. Many methods have been applied to classify students that are based on IF-THEN rules and pattern recognition. However, many studies were done on intelligent tutoring systems and e-learning systems, not in a conventional classroom. Moreover, there are no researches that measure basic values by considering intelligence and non-intelligence performances. In this work, a student classification model was developed by applying neuro fuzzy concept; a combination of fuzzy's IF-THEN rules and neural network's ability to learn, so this method has the ability to learn from the generated rules to produce the best classification model. The model can be used to predict students' academic performance. Data were processed using ANFIS Editor-Matlab Fuzzy Logic. The results showed that combination of three parameter values -interest, talent, and motivation- is the best model for students classification, which has training RMSE value 0.12301 and testing average RMSE value 0.25611.**

Keywords-neuro fuzzy; student classification; students' academic performance; ANFIS; Fuzzy Logic Toolbox

I. INTRODUCTION

Education is the main pillar of a country [1]. Formal education is an example of education that can be reached through the school. Conventional class as the main learning media in undergraduate program of Electrical Engineering and Information Technology Department, Gadjah Mada University, has several drawbacks, such as large amount of students and limited number of meetings. As a result, it is hard to understand each student. Usually, lecturers held a pretest at the beginning of a course in each semester. Pretest can be used to determine student's prior knowledge on the concept they are going to explain. However, questions in the pretest usually measures intelligent variables, meaning that the test was conducted to determine student's intelligence level. Intelligence is an important aspect that influences the success of learning, but there are other factors that also influence a student's ability to learn something.

Basic value is the value that indicates student's readiness in following a learning process. It is not only determined by the intelligence factor, but also several other contributing factors. There are two factors that influence the success of learning, which are internal and external factors. Internal factors of a student are including physiological aspects (physical state, eyes, and ears) and psychological aspects (intelligence, attitude, talent, interest, and motivation). External factors, on the other hand, are factors coming from student's environment, such as family, teachers, community, friend, school, and equipment [2].

Using basic values, including intelligence and non-intelligence, to classify students has not been done to undergraduate student in the Department of Electrical Engineering and Information Technology (JTETI) Universitas Gadjah Mada (UGM). Student classification by considering these parameters can help teachers to understand and determine appropriate treatment for the student.

II. RELATED WORKS

Student classification have been done with a variety of models to generate classification with minimal error rate. Naïve Bayesian classification is used to predict student classification based on achievement and demographic data to determine student's department [3]. Several decision tree algorithms such as C4.5, IDE (Investigational Device Exemption), and CART (Classification and Regression Trees) can be used to predict test results and identify students who are likely to fail using students' demographic and past performance data [4]. Student classification based on a variety factors, including socio-demographic (age, gender, ethnic, education, employment status) and learning environment can also be done with a tree classification method, namely CHAID (Chi-squared Automatic Interaction Detector), exhaustive CHAID, and QUEST (Quick, Unbiased, Efficient Statistical Tree) [5]. Fuzzy logic and multi layer perceptron (MLP) also have been applied for student classification, such as in [6, 7, 8, 9].

Several studies were conducted to compare the use of classical methods (mathematical rules) and fuzzy logic for student's achievement classification, and found that fuzzy logic provide a more flexible models than the classical methods [10]. Romero et.al stated that student classification using methods based on IF-THEN rules such as decision tree, rule induction, and fuzzy rules are easier to use and interpret, while the method based on pattern recognition such as statistical and neural network methods are highly dependent on data's condition [11].

Furthermore, combination of modeling technique could be applied to obtain higher accuracy and produce

better classification results, such as neuro fuzzy method. Neuro fuzzy is the combination of neural network and fuzzy logic. Neural network was able to recognize the system through a learning process to improve the adaptive parameters, however, it also has a complex structure. On the other hand, fuzzy system has a simpler concept that is similar to human thinking. Thus, combination of the two would be complementary [12].

Based on previous research conducted, there are no research that measure basic values by considering a combination of intelligence and non-intelligence (interests, talents, and motivations) performance. This research determined the student classification based on basic value with a combination of input parameters that can be used as predictors of student's academic performance. Furthermore, most studies were done on intelligent tutoring systems and e-learning systems, student classifications that were done in conventional classroom were very rare. ANFIS (Adaptive Neuro Fuzzy Inference System) Editor on Matlab's Fuzzy Logic Toolbox is used for analyzing and processing data**.** RMSE (Root Mean Squared Error) values are used to measure system's performance based on different combinations of inputs. A full description of RMSE will be delivered in the next section. Classification model that has the smallest RMSE value is the best model for classification.

III. CLASSIFICATION

Classification is the process of finding models (or functions) that describe and distinguish data classes or concepts, in order to be able to use the model to predict an object whose class label is unknown. The model is based on an analysis of training data set in a data object whose class is known. Classification can be used for forecasting or prediction, by identifying the distribution trends based on available data. Classification and prediction may need to be preceded by relevance analysis is used to identify the attributes that do not contribute to the classification or prediction process. These attributes can then be excluded [13].

The classifier as a predictor can be evaluated by several methods, such as mean absolute error and mean squared error. When a system takes the square root of the mean squared error, the resulting error measure is called the root mean squared error (RMSE). This is useful in that it allows the error measured to be of the same magnitude as the quantity being predicted or classified [13].

RMSE values were evaluated to obtain a combination of inputs that became the best classification model. RMSE value indicates the difference in the average data of the squared differences in the estimated values with observed values of a variable, as shown in equation (1),

$$\text{RMSE} = \sqrt{\frac{\sum|X_t - \hat{X}_t|^2}{n}} \quad (1)$$

where X_t states the actual output data, $\hat{X}_t$ is data processed by system, and *n* as the number of data. Smaller number of RMSE indicates models or variables is valid [14].

IV. NEURO FUZZY

Neuro Fuzzy is a combination of two systems, neural network and fuzzy rules. Neuro fuzzy based on the fuzzy inference system is trained using a learning algorithm derived from neural network system. Based on its ability to learn the neuro fuzzy system is often referred to as ANFIS (Adaptive Neuro Fuzzy Inference System) [15]. ANFIS continue to be developed with a variety of types, such as ANFIS type 3 that use Sugeno method [16]. ANFIS structure is shown in Fig. 1.

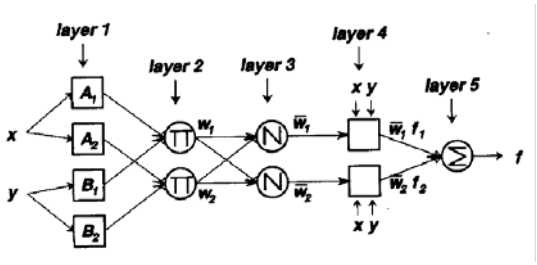


Fig. 1 ANFIS Structure [16]

Circle node on the ANFIS layer represents the non adaptive node, while the square represents adaptive node, which means the value of the parameter can be changed according to the learning provided. A full description of the structure of ANFIS the layers are as follows [16].

a. The first layer is the adaptive fuzzy rules, each parameter can be changed according to the function node, as shown in equation (2).

$$O_{1,i} = \mu_{A_i}(x) \quad (2)$$

Equation (2) shows that *x* is the input to node *i*, and A_i is the linguistic label. O_i^1 is a membership function that determines the degree of membership in accordance with the given *x*. There are several ways to determine membership function, for example with a trapezoidal function.

b. The second layer is a nonadaptive layer, the function of this layer is to multiply all incoming values to produce an output value, as shown in equation (3).

$$O_{2,i} = w_i = \mu_{A_i}(x).\mu_{B_i}(y), i = 1,2 \quad (3)$$

Each output node represents the firing strength of a rule.

c. The third layer is a nonadaptive layer, use to normalize the firing-strength. Normalization is done as shown in equation (4).

$$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1 + w_2}, i = 1,2 \quad (4)$$

d. The fourth layer is an adaptive layer, which value can be customized by consequent parameters. The function of this layer shown in equation (5),

$$O_{4,i} = \overline{w_i} f_i = \overline{w_i}\,(p_i x + q_i y + r_i) \quad (5)$$

where $\overline{w_i}$ is the output of third layer, and $\{p_i, q_i, r_i\}$ is the parameter set.

e. The fifth layer consists of a nonadaptive node that sum of all inputs, the computations shown in equation (6).

$$O_{5,i} = \sum \overline{w_i} f_i = \frac{\sum_i w_i . f_i}{\sum_i w_i} \quad (6)$$

Training set is used to fix the parameters of the adaptive layer. Training in ANFIS uses backpropagation algorithm. On the backpropagation learning, the consequent parameters are made permanent. Errors that occur between adaptive network output and the actual output propagated back using gradient descent to fix the parameters of the premise. The learning phase is called epoch [17].

V. METHODOLOGY

The main step in the research is the collection and processing of data, student classification, and an evaluation of the classification model. Data collection was conducted by using questionnaire and quiz on Entrepreneurship class in JTETI UGM. The use of one class in this research aimed at getting the same assumptions about the calculation of quantitative value and also class condition.

Classification begins with analysis of questionnaire result then distribute data into four categories (poor, satisfactory, good, and very good). The categorized data then divided into three independent sets according to 3-fold cross validation method. Two thirds of data used for training process, while the rest is used for testing process. Iteration lasts three times with variations of the training data and testing using a combination of three subset of data. Each iteration produces an error rate, and the average value from the three iteration is the error of the whole system [18].

Training data is processed by ANFIS Editor, by generating Sugeno fuzzy type and split the membership function of each parameter into three categories: low, medium, and high. Data are trained with backpropagation method. Number of training and error tolerance is set to get optimal value. The process of training and testing produced RMSE values and performed on 15 different combinations of input from four parameters -intelligence, interests, talents, and motivations. Combination model with the smallest RMSE value can be determined as the best classification model that can be used as predictors of students' academic performance.

VI. RESULTS AND DISCUSSIONS

Data collection was conducted using questionnaire and quiz on 71 respondents in conventional classroom of Entrepreneurship. The respondents are undergraduate student in Department of Electrical Engineering and Information Technology (JTETI) Universitas Gadjah Mada (UGM). This process is done at the beginning of the semester. Questionnaire and quiz consisting of 13 questions are used to determine the level of intelligence, interests, talents, and student motivation.

Data collected from the questionnaire result and quizzes are then assessed quantitatively according to the four parameters to be assessed. The maximum value for each parameter is 10, so the sum of the four parameters will have maximum value 40. This assessment results is used as the basic of students' grades.

Data processing begins with dividing data into four student categories: poor, satisfactory, good, and very good. The threshold for each category is defined by interviewing the lecturer of entrepreneurship course and literature study. Table I shows the boundary of each student category; x represents the percentage of total score collected by each student. Suppose, a student with total score 30, then x= (30/40) * 100% = 75%, thus the students will be categorized as very good.

TABLE I. STUDENT CATEGORY IN PERCENTAGE OF GAINED TOTAL SCORE.

Category	Percentage Value
Poor	$x \leq 60\%$
Satisfactory	$60\% < x \leq 65\%$
Good	$65\% < x \leq 70\%$
Very Good	$x > 70\%$

After the categorization step, data is divided for ANFIS training and testing process. The data is divided into three subsets for 3-fold cross validation. Experiments were done by using different number of inputs into ANFIS, i.e. 1, 2, 3, and 4 input parameters. As a result, an optimum composition of inputs to the classification process will be achieved, which is 3 input parameters. Due to limited space in this paper, we only discuss in detail the use of three parameters, including intelligence, interest, and talent, to be fed into ANFIS to produce one student classification model.

ANFIS Editor on Fuzzy Logic Toolbox-Matlab has the ability to generate FIS (Fuzzy Inference System), for example, in the generation of FIS from three input values, a classification model can be produced, as shown in Fig. 2.

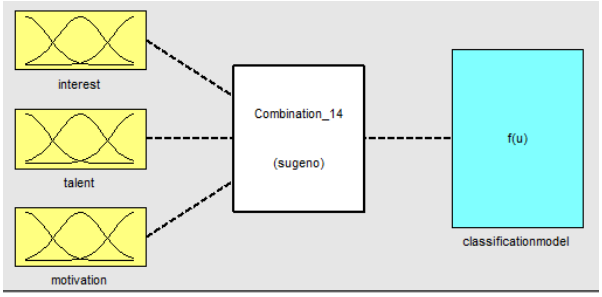


Fig. 2 FIS Structure with Three Inputs and One Output

The results of the first iteration are shown in Fig. 3, x-axis represents number of epoch and y-axis shows the error. The training process used backpropagation algorithm, with amount of training (epoch) 30, error tolerance limit of 0, and produces RMSE value of 0.10658 in the first iteration. The second iteration produces RMSE value 0.13033, while the third iteration produces 0.13213. Average of the three training iterations is 0.12301.

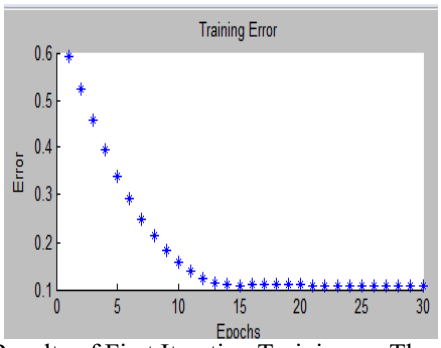


Fig. 3 Results of First Iteration Training on Three Input Combination (Interests, Talents, and Motivation)

Results of the first iteration of testing were shown in Fig. 4.

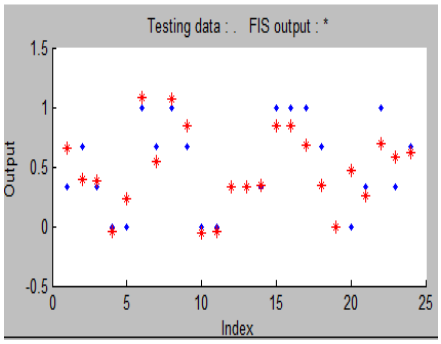


Fig. 4 Results of First Iteration Testing on Three Input Combination (Interests, Talents, and Motivation)

Fig. 4 shows the value that is generated by the system and the expected value, represented by red dots and blue dots respectively. *X*-axis in Fig. 4 represents the number of testing data and *y*-axis represents the student category ranged from 0 to 1 as output. Poor category is represented by 0, very good category as the highest student's category for academic performance is represented by 1, while both of middle categories are represented in number between 0 and 1 : 0.3 for satisfactory and 0.6 for good student category. Difference of both values, i.e. value generated by system and the expected values, produces RMSE values. The average RMSE values generated in testing data is equal to 0.19803 in the first iteration. Value generated in the second iteration 0.32323, and 0.24707 at the third iteration. Average of the three iterations of testing is 0.25611.

FIS generation, training, and testing were conducted using the same rules for other models, which were using different input combinations. Student classification results with 15 different models of input combinations shown in Table II.

Table II shows that neuro fuzzy method can be used to classify students, shown by the small RMSE values generated in the process of training and testing. The more parameters used in the training process makes the smaller the RMSE value. The use of additional parameter/s will add variety to the data that were accommodated in the training process, thus, making the system easier to classify students into four categories.

TABLE II. COMPARISON OF CLASSIFICATION RESULTS.

Number of Input	Parameter combinations	Average of RMSE	
		Training	Testing
1	Intelligence	0.35724	0.39632
	Interest	0.26647	0.29569
	Talent	0.31518	0.31016
	Motivation	0.26888	0.27496
2	Intelligence - Interests	0.26332	0.33435
	Intelligence - Talent	0.29722	0.35949
	Intelligence - Motivation	0.23809	0.32461
	Interests - Talent	0.23355	0.31093
	Interests - Motivation	0.19405	0.26080
	Flair - Motivation	0.1968	0.26040
3	Intelligence - Interests - Talent	0.21523	0.33513
	Intelligence - Interests - Motivation	0.1516	0.30555
	Intelligence - Talent - Motivation	0.1634	0.25737
	Interests - Talent - Motivation	0.12301	**0.25611**
4	Intelligence - Interests - Talent - Motivation	0.35724	0.39632

Combination of three parameters -interest, talent, and motivation- produces the smallest error rate in testing process when compared with 14 other combinations. This suggests that the combination of these three parameters is the best combination to classify students, so the classification results can be used to predict students' academic performance. This classification can be used in other courses, with a standard adjustment for measuring the value of intelligence, talent, interest, and motivation of students.

VII. CONCLUSION

Results showed that the more parameters used in input combination, the resulting error rate on the training data was likely to decline. From four parameters used in the experiment, motivation parameter was the best parameter that can be used in classification by producing error value 0.27496 in the testing process, then followed by interest parameter with an error rate of 0.29569, talent parameter with an error rate of 0.31016, and intelligence parameter that produce 0.39632.

Combination of three inputs (interest, talent, and motivation) is the best classification model to predict students' academic performance, with RMSE values in the training process is 0.12301 and the average value 0.25611 in the testing process.

Student classification can be done in other courses with adjustment parameter measurements to assess the interests, talents, and motivations value of the students.

Suggestions for the development of this research include the addition of parameters and the amount of data for the input so that the system is expected to identify patterns and variance in data classification more easily. Furthermore, by testing with different algorithm one can compare the results and the classification error rate between neuro fuzzy and other algorithms that may increase the confidence of the result.

REFERENCES

[1] Dewantara, Ki Hajar. (1977). Karya Ki Hadjar Dewantara: Bagian Pertama Pendidikan. Yogyakarta: MLPTS.

[2] Syah, M. (2006). *Psikologi Belajar*. Jakarta: PT RajaGrafindo Persada.

[3] Bhardwaj, B. K., Pal, S. (2011). *Data Mining: A Prediction for Performance Improvement Using Classification*. International Journal of Computer Science and Information Security, Vol. 9, No. 4.

[4] Yadav, S. K., Pal, S. (2012). *Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification*. World of Computer Science and Information Technology Journal, Vol. 2, No. 2.

[5] Kovačić, Z. J. (2010). *Early Prediction of Student Success: Mining Students Enrolment Data*. Proceedings of Informing Science & IT Education Conference.

[6] Yadav, R. S., Singh, V. P. (2011). *Modeling Academic Performance Evaluation Using Soft Computing Techniques: A Fuzzy Logic Approach*. International Journal on Computer Science and Engineering, Vol. 3, No. 2.

[7] Mustafidah, H., Suwarsito. (2012). Student Learning Achievement Prediction Based on Motivation, Interest, and Discipline Using Fuzzy Inference System. International Conference – Green World, Business, and Technology.

[8] Naik, B., Ragothaman, S. (2004). *Using Neural Network to Predict MBA Student Success*. College Student Journal.

[9] Tasarov, A., Kopytov, E. (2006). *Students' Classifications Based on Study Results*. RESEARCH and TECHNOLOGY – STEP into the FUTURE, Vol. 1, No. 1.

[10] Gokmen, G., Akinci, T. C., Tektaş, M., Onat, N., Kocyigit, G., Tektaş, N. (2010). *Evaluation of Student Performance in Laboratory Applications using Fuzzy Logic*. World Conference on Educational Science.

[11] Romero, C., Ventura, S., Espejo, P. G., Hervás, C. (2008). *Data Mining Algorithms to Classify Students*. Proceedings of Educational Data Mining.

[12] Aria, M. (2010). *Sistem Neuro Fuzzy*, http://www.elib.unikom.ac.id/download.php?id=107544. Diakses 1 Maret 2013.

[13] Han, J. Kamber, M. (2006). *Data Mining : Concepts and Techniques*. San Fransisco: Morgan Kaufmann Publisher.

[14] Pindyck, R. S., Rubinfeld, D. I. (1998). *Econometric Models and Economic Forecats*. New York: McGraw-Hill.

[15] Jang, J. S. R., Sun, C.T., Mizutani, E. (1997). *Neuro Fuzzy and Soft Computing*. New Jersey: Prentice-Hall International.

[16] Jang, J. S. R. (1993). *ANFIS: Adaptive-Network-Based Fuzzy Inference System*. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 23, No. 3.

[17] Kurniawan, H. (2008). *Optimasi Pemilihan Produk Operator GSM dan CDMA Menggunakan ANFIS*. http://repository.ipb.ac.id/handle/123456789/17675. Diakses 27 Februari 2013.

[18] Martin, P. (2012). Dispel Tutorial: Case Studies, http://homepages.inf.ed.ac.uk/pmartin/tutorial/case_studies.html, diakses 1 Maret 2013.

Tech Review: Game Platform for Upgrading Counting Ability on Preschool Children

Endah Sudarmilah, Ridi Ferdiana, Lukito Edi Nugroho, Adhi Susanto
Department of Electrical Engineering and Information Technology
Universitas Gadjah Mada (UGM)
Yogyakarta, Indonesia
endah_s3_12@ugm.ac.id, ridi@ugm.ac.id, lukito@mti.ugm.ac.id, susanto@te.ugm.ac.id

Neila Ramdhani
Faculty of Psychology
Universitas Gadjah Mada (UGM)
Yogyakarta, Indonesia
neila_psi@ugm.ac.id

***Abstract*—One of difficulties in learning mathematics (counting) can be overcome by providing a relaxed and fun learning for preschoolers. Games can be used as an alternative solution. This study was conducted as a pilot project for reviewing the Kodu, Unity 3D and Construct 2 game platform for this purpose. The method used is the classification, review/evaluation, prototyping and analysis. Tech Review of these game platforms will be discussed as a result in this paper.**

Keywords— game platform, preeschool, mathematics, Kodu, Unity 3D, Construct 2

I. INTRODUCTION

Children usually do not like to learn mathematics. Because it's difficult to learn, irrelevant to their lives, and boring [1], it raises mathematics anxiety. Mathematics anxiety and aptitude/achievement measures are inversely correlated. The higher mathematics anxiety is, the lower mathematics performance will be [2]. Many children, including those with and without disabilities, as well as adults, do not feel confident in their ability to do math [3].

Computer games, on the other hand, can be highly effective in increasing children's learning and enjoyment of mathematics. Its effectiveness, however, depends on many things including details of the software design such as interface styles and scaffolding, teacher and student expectations, the level of integration with other learning activities, and the setting and usage pattern [4]. Symbol manipulation (interpreted by something in reality) is the essential characteristic of preschool children [5]. This has been the research basis in improving cognitive abilities, especially for preschool children numeracy skills through the game. This is expected to be able to maximize the symbol manipulation by children.

One of the ways to assist in reducing math anxiety for children is by developing calm/positive ways to deal with their math anxiety, included visualization, positive messages, relaxation techniques, frustration breaks [3] that can be handled by games.

Learning math courseware used is known as playtime learning. It consists of play games and fun activities [6] which covers two modules. First module facilitates to recognize numbers and its sequence and the second one provides training of number addition and subtraction on the lowest difficulty level. In this study, these were used with different game genres like adventure, racing or shooting. These games are a combination of multimedia elements such as graphic and animations facilitating learners to experience mathematics [7].

This study was purposed a comparison between three game platforms that suitable for counting games. It will improve cognitive abilities of preschoolers. Three gaming platforms used were Kodu, Unity 3D and Construct 2. The outcome of this study was choosing the best platform for its application.

II. LITERATURE REVIEW

A. Related Works

Children are one of the oriented population sectors who will obtain game treatment and education [8]. It is so undeniable that there are some works dedicated to this goal for example: the story-based math video game has captivated the interest of students and it has been beneficial in improving their performance in an assessment test. Most notably, the improvement was higher for students who used to have poor performance in mathematics [9]. Other work mentioned that a simulation-based "game" environment that included a sequence of challenges would further motivate the student and expand their learning abilities [10]. It's also stated that their observations and findings were on children's psychology of learning mathematics in the context of computer-based mathematical game environments and some elements of computer games that satisfy children's learning needs and motivate them to learn mathematics [11].

Review study in terms of computer game engines stated that several applications of commercially available computer game engines were implemented on virtual education and training environments. While these systems are still in the early stages of their development, they have already provided us with glimpses into their tremendous potential for creating effective learning and training experiences in various fields, including education and student laboratories, medicine as well as disaster response and military training [12].

B. Game Engine Review

Game Engine is specialized software that used to design and to develop a video game. It is also an important core of a video game with the functional components that can be reused (such as graphics rendering, audio output, physics modeling, game, logic, basic artificial intelligence, interaction users, as well as multi-user network) and usually accompanied by a software development kit (SDK) [12]. Detailed explanation of the features can be seen in Fig. 1. It shows about the features that should be present in a platform game engine namely input support, sound support, graphical support, AI support and physics support [13].

1) Kodu

Kodu is software support a creation of child-friendly games (suitable for children at any age) and the programming environment that enables non-technical users to create complete games. It is including behavioral AI (Artificial Intelligence) and many reusable components. Kodu includes graphical programming environment based on a system of rules and it can also be controlled by a 10 year old child. Kodu can be used to implement various game genres, design terrain and complete level. This game platform uses Xbox 360 or Windows PC [14].

Unlike the other game engine, Kodu language uses "when-do" algorithms. Although simple, this model is identified in 346 Kodu programs created by users that show the flow of control and sophistication using complex Boolean logic [15]. It can be seen in Fig. 2.

2) Unity 3D

Unity is an integrated tool to create games, simulations and architectural buildings. It can be applied for PC games and online games using the Unity web browser [16].

The programming language used varies, ranging from JavaScript, C #, and Boo. Unity is a Multi platform game that can be deployed on PC, Mac, Wii, iPhone, iPad, Android and browsers. Many things can be done using unity. There are audio feature reverb zone, particle effects, Sky Box, as well as features rendering, lighting, sound effect, and game physics. Unity can also directly edit the texture from editor [17]. Fig. 3 shows about it.

3) Construct 2

Construct 2 is an application for creating 2-dimensional games based on HTML-5. It does not require programming skills [18].

The construct 2 display is easy to understand. Its features includes Powerful Event System focusing on logic; Flexible Behaviors (physical properties, movement, plat former, etc.), Instant Preview; Stunning Visual Effects; Multiplatform Export applied in games that can be published to a variety of platforms, namely, Chrome Web Store, Facebook, Kongregate, Newgrounds, Firefox Marketplace, Arcade Sencha while to get the same performance as the previous, PC, Linux, Mac are as good as mobile platforms such as iOS, Android, Blackberry [19].

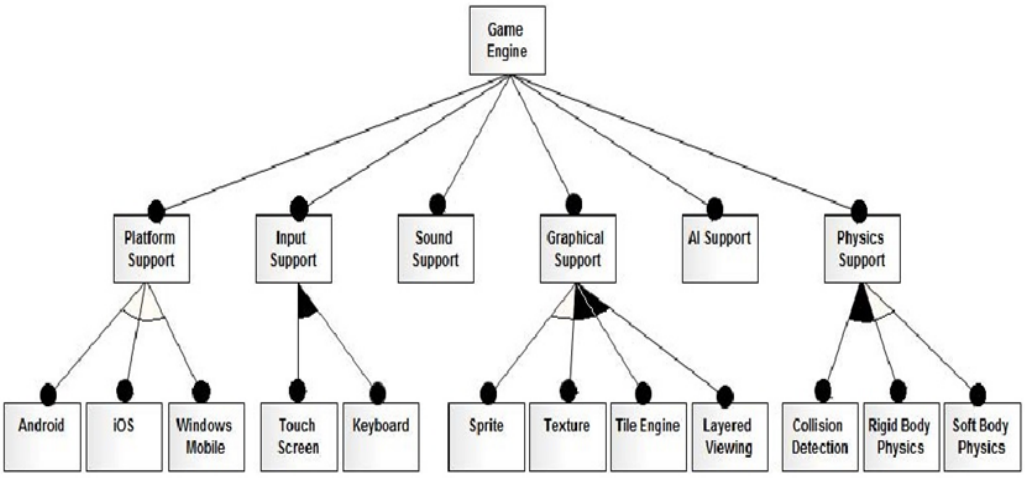


Fig. 1. A Game Engine Features Diagram [13]

Fig. 2. Complex Boolean Logic on Kodu

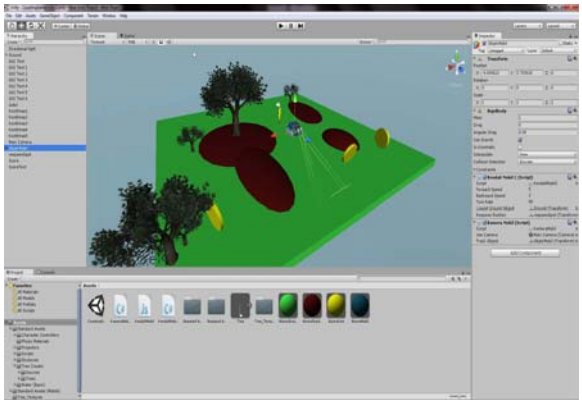


Fig. 3. Texture Editor on Unity 3D

III. METHOD

The method conducted in this study was a modified linear sequential model. This method consists of four software development steps that begin at the system level and progresses through analysis, design, coding, testing, and support [20]. The first step of this research was game engine classification and continued with review and evaluation using part of analysis module. Subsequently, prototyping and analysis (tech review) were the next steps. Research method flowchart can be described as in Fig. 4. It shows that research method contain for steps which are game engine classification, review and evaluation, prototyping, and analysis.

A. Game Engine Classification

In this phase, we categorized game engines based on their features and characteristics to be compared and collaborated with software requirements analysis [21]. In this research, it was needed 2D and 3D game engines to support the whole applications. Several game engines were thus classified into these criteria.

B. Review and Evaluation

Review and evaluation were carried out by technological observation during the installation process, configuration, programming, interfaces and support tools on each of the game engine. Evaluation methodology will select ideal engines based on the technical requirements of which an associated link to the driving pedagogic factors [22]. In this step, several game engines were reviewed and evaluated such as jMonkeyEngine, CryEngine 3, Unity3D, Kodu and Blender for 3D game engines; Adobe Flash Professional, Construct 2, Game Of Engines, GameMaker Studio and AndEngine. The technical requirements for making applications of counting games improves cognitive abilities of preschoolers, support and feature game engines; indeed for the programming easiness, three game engines were selected to be prototyped.

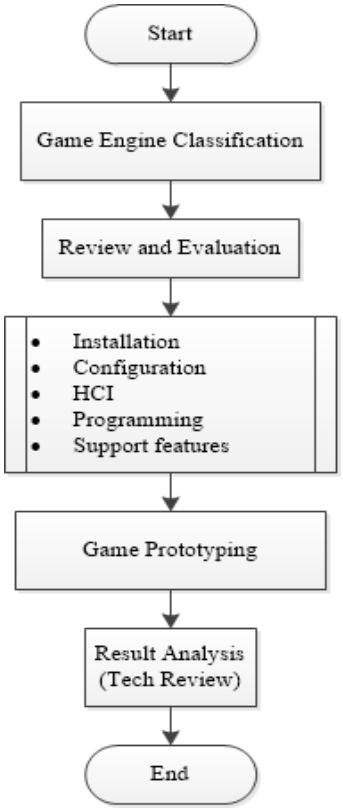


Fig. 4. Research Method Flowchart

C. Prototyping

The prototyping paradigm is begun with requirements gathering. It is continued to identify any requirements and outlining a "quick design". The quick design focused on software aspect representations that would be visible to the customer/user (e.g., input approaches and output formats) [20].

Prototyping was conducted by designing and building a counting game for preschool children using game engines that were selected, namely Kodu, Unity 3D and Construct 2.

D. Analysis

Lastly, we observed prototype details, collated and compiled the comparison result as a tech review. In this step, we used personal review to get data on using evaluation, programming experience and support features on the gaming platforms which were analyzed.

IV. Result And Discussion

A. HCI (Human and Computer Interaction) evaluation

All games created with this platform still use keyboard control system in which each one uses different genre but having the same goal i.e. to improve the cognitive abilities of children in numeracy.

The colors are attractive and the sound stimuli giving game instructions is also used to increase the children interest in learning to count by using the game. Children will be more interested and fun in playing the game [23]. Interface 3D platform games have a tendency to be more attractive than 2D because children will be able to feel the real environment like in the real world using 3D games visualization facilities [24]. Screenshot HCI on each platform game can be seen in Fig. 5, 6, and 7.

B. Programming Experience

Programming process undertaken to develop the game in the third platform has its own advantages and disadvantages.

Fig. 5. Screenshot HCI Counting Game with Kodu

Fig. 6. Screenshot HCI Counting Game with Unity 3D

Fig. 7. Screenshot HCI Counting Game with Construct 2

Kodu game development process does not require complex scripting for all events in the game are controlled by algorithms "when-do" using logic Boolean. Therefore, the character of which will be used depending on the character in the game provided the platform without having to create their own character [15] [25].

Unity 3D game scripting is done through Mono which is an open source implementation of the NET Framework, thus the programmer making it will be more flexible to use JavaScript, C# or Boo (which has Python syntax). Unity 3D game character is able to use self-made character or import it from already available one in the Unity Asset. Although using unity, it still needs higher scripting capabilities than Kodu because events that occur in all games must be declared through the scripts used. This would be workflow module development flexibility and a more automated navigation [26].

On Construct 2, a game is created just by dragging and dropping object from scripting through event sheet which is the ability of the core logic requires strong logic games can run well. In the scripting process, Construct 2 also provides the function of default behaviors which has been provided as physical, movement, plat former, etc [18].

C. Supports

Documentation supports for Kodu, Unity 3D and Construct 2 can be said to be very good with each having its official web page for a basic introduction, detailed information about the function, tutorial and download source [18][25][26].

In addition to support the official website, users and developers community of all platforms are also quite good with a fairly lively forum [27][28][29].

D. Technology Review

The differences in technical criteria of each game engine are essentially influenced by the intended use of the game engine, as shown in Table 1. Kodu and Unity 3D have some similarities such as the use of graphics API and the environment as both are aimed to construct a 3D game. Kodu and Unity 3D however, also have a very significant difference due to the different target user goal. Kodu is used more to learn programming while the Unity 3D game is more applicable to the professionals or industrial users [30].

TABLE I. TECHNOLOGY REVIEW LIST

Criteria	Kodu	Unity 3D	Construct 2
Engine Type	3D Visual game editor	3D Visual game editor	2D Visual game editor
Target platform	Windows, Xbox	PC, Mac, Android, iOS, Xbox, Wii	PC, Mac, Android, iOS, Blackberry
Graphics API	DX 9 Shader Model 2.0	DX 9 OpenGL	2D context, WebGL context, based on OpenGL
Script runtime environment	.NET Framework 3.5 XNA Framework 3.1	Mono (open source implementation of .NET)	HTML 5
Physics	N/A	PhysX	Behaviors
3D scene editor	No	Yes	No
3D model editor	No	No	No
Terrain editor	Yes	Yes	No
Particle system editor	No	Yes	Yes
Asset importing	No	Yes	Yes
Runtime Engine	.NET Framework 3.5	Mono 2.0	HTML 5
Video support	Yes	Yes	No
Sound	DirectX	DirectSound3D	Web Audio API

Construct 2 has a more striking difference with other game engines because it leads to the development of 2D games that do not require rendering realistic environment [24]. It is because the player only controls the game character in a 2D environment [31].

Furthermore, Construct 2 using the HTML 5 web base is different from the others but also supports the Unity 3D web browser to use platform with a virtual world viewer [16].

V. CONCLUSION

Game engine can be used based on the system making purpose or market goals. In making application, it should consider the needs analysis and technology criteria owned by a game engine.

Based on the results of tree game engine reviews, they were used to create counting game application aiming to improve the cognitive abilities of preschoolers. Kodu and Unity 3D were more suited for games that emphasized the advantages of three dimensional environments. Although Kodu was faster for prototyping process, Unity 3D had advantages in adding character and properties.

Furthermore, 2D constructs were suitable for creating 2D games with simpler scripting process only by the "drag and drop" and using simple logic in the Event Sheet.

ACKNOWLEDGMENT

This work was supported by Directorate General of Higher Education (DIKTI) of Indonesia and Informatics Department of Universitas Muhammadiyah Surakarta (UMS).

REFERENCES

[1] J. M. Furner and B. T. Berman, “Review of Research: Math Anxiety: Overcoming a Major Obstacle to the Improvement of Student Math Performance,” Child. Educ., vol. 79, no. 3, pp. 170–174, 2003.

[2] R. Hembree, “The Nature, Effects, and Relief of Mathematics Anxiety,” J. Res. Math. Educ., vol. 21, no. 1, pp. 33–46, Jan. 1990.

[3] J. M. Furner and M. L. Duffy, “Equity for all students in the new millenium: Disabling math anxiety,” Interv. Sch. Clin., vol. 38, no. 2, pp. 67–74, 2002.

[4] M. M. Klawe, “When Does The Use Of Computer Games And Other Interactive Multimedia Software Help Students Learn Mathematics?(DRAFT May 29, 1998).”

[5] J. . Santrock and S. R. Yussen, Child Development. Dubuque, IA, Wm, C.Brown., 1992.

[6] W. F. W. Ahmad, E. A. P. Akhir, and S. Azmee, “Game-based learning courseware for children with learning disabilities,” in Information Technology (ITSim), 2010 International Symposium in, 2010, vol. 1, pp. 1–4.

[7] Z. Muda and R. E. K. Mohamed, “Multimedia Design And Development In Mathematics Learning Courseware For Preschool Education,” in International Conference on Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, 2005, vol. 2, pp. 514–517.

[8] N. A. Bartolome, A. M. Zorrilla, and B. G. Zapirain, “Can game-based therapies be trusted? Is game-based education effective? A systematic review of the Serious Games for health and education,” in 2012 17th International Conference on Computer Games (CGAMES), Los Alamitos, CA, USA, 2011, vol. 0, pp. 275–282.

[9] M. Giannakos, K. Chorianopoulos, and L. Jaccheri, “Math Is Not Only for Science Geeks: Design and Assessment of a Storytelling Serious Video Game,” in Advanced Learning Technologies, IEEE International Conference on, Los Alamitos, CA, USA, 2012, vol. 0, pp. 418–419.

[10] J. Tan and G. Biswas, “Simulation-Based Game Learning Environments: Building and Sustaining a Fish Tank,” in The First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning, 2007. DIGITEL ’07, 2007, pp. 73–80.

[11] K. Sedighian and A. Sedighian, “Can educational computer games help educators learn about the psychology of learning mathematics in children,” in 18th Annual Meeting of the International Group for the Psychology of Mathematics Education, 1996.

[12] F. Arango, E.-S. Aziz, S. K. Esche, and C. Chassapis, “A review of applications of computer games in education and training,” in 2012 Frontiers in Education Conference Proceedings, Los Alamitos, CA, USA, 2008, vol. 0, pp. T4A–1–T4A–6.

[13] A. G. Peker and T. Can, “A design goal and design pattern based approach for development of game engines for mobile platforms,” in 2011 16th International Conference on Computer Games (CGAMES), 2011, pp. 114–120.

[14] M. MacLaurin, “Kodu: end-user programming and design for games,” in Proceedings of the 4th International Conference on Foundations of Digital Games, New York, NY, USA, 2009, p. 2:xviii–2:xix.

[15] K. T. Stolee and T. Fristoe, “Expressing computer science concepts through Kodu game lab,” in Proceedings of the 42nd ACM technical symposium on Computer science education, New York, NY, USA, 2011, pp. 99–104.

[16] N. Katz, T. Cook, and R. Smart, “Extending Web Browsers with a Unity 3D-Based Virtual Worlds Viewer,” Ieee Internet Comput., vol. 15, no. 5, pp. 15–21, 2011.

[17] W. Goldstone, Unity Game Development Essentials. Packt Publishing Ltd, 2009.

[18] “Create Games with Construct 2 - Scirra.com.” [Online]. Available: https://www.scirra.com/construct2. [Accessed: 27-May-2013].

[19] D. Bigelow, Construct Game Development Beginners Guide. Packt Publishing Ltd, 2012.

[20] R. S. Pressman, Software engineering: a practitioner’s approach. McGraw-Hill Higher Education, 2010.

[21] B. Chao, X. Zhu, Q. Li, and A. Huang, "Reliability Management in Software Requirement Analysis," in 2006 IEEE International Conference on Management of Innovation and Technology, 2006, vol. 2, pp. 1104–1107.

[22] P. Petridis, I. Dunwell, S. de Freitas, and D. Panzoli, "An Engine Selection Methodology for High Fidelity Serious Games," in 2010 Second International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES), 2010, pp. 27–34.

[23] S. Jumisko-Pyykko, S.-M. Hellsten, M. Weitzel, and I. Rakkolainen, "Children'S Game Experiences in Different Settings," in 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008, 2008, pp. 377–380.

[24] M. Yang, Z. Wang, and S. Xiao, "Research on of 3D game design and development technology," in 2010 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), 2010, vol. 9, pp. 762–765.

[25] "Kodu | About." [Online]. Available: http://www.kodugamelab.com/About. [Accessed: 29-May-2013].

[26] "UNITY: Game Development Tool." [Online]. Available: http://unity3d.com/. [Accessed: 30-May-2013].

[27] "Kodu | Forum > Home." [Online]. Available: http://koduwebdnn.cloudapp.net/. [Accessed: 30-May-2013].

[28] "Unity | Home." [Online]. Available: http://udn.unity3d.com/. [Accessed: 30-May-2013].

[29] "Scirra Forum." [Online]. Available: https://www.scirra.com/forum/. [Accessed: 30-May-2013].

[30] M. Hubbell and J. Kepner, "Large scale network situational awareness via 3D gaming technology," in 2012 IEEE Conference on High Performance Extreme Computing (HPEC), 2012, pp. 1–5.

[31] D. Fu, R. Jensen, and E. Hinkelman, "Evaluating Game Technologies for Training," in 2008 IEEE Aerospace Conference, 2008, pp. 1–10.

The Effect of Transformation on Anisotropic Semivariogram Model

Rr. Kurnia Novita Sari
Bandung Institute of Technology
Bandung, Indonesia
kurnia@math.itb.ac.id

Udjianna S. Pasaribu
Bandung Institute of Technology
Bandung, Indonesia
udjianna@math.itb.ac.id

Oki Neswan
Bandung Institute of Technology
Bandung, Indonesia
oneswan@math.itb.ac.id

***Abstract*—Usually there are two anisotropic semivariogram model. Those are geometric anisotropy and zonal anisotropy that depend on distance and direction between pairs of observations. Here, we study more the geometric anisotropy model because this model use transformation of any coordinate of locations. The transformations include translation, rotation and scaling. Translation and rotation does not change the distance that is an important element in semivariogram models. While scaling change the distance that also certainly change semivariogram model.**

***Index Terms*—Anisotropic, rotation, scalling, semivariogram, translation.**

I. INTRODUCTION

Many phenomena in our life describes a pattern. The phenomena can be considered as sequence of random variables that is depend on an index parameter. The index parameter are time, location, or time and location. For example, the number of victims of the SARS (Severe Acute Respiratory Syndrome) virus in Hongkong city will affect into the number of victims in the surrounding towns. That effect is depends on the distance and direction between the pair of locations. That is show the the weight of spatial relationship between a location to other locations [1].

The above example show that spatial correlation between pair of locations are affected by h where h is distance between two locations. The spatial correlation can be measured with semivariogram, covariogram, or correlogram. In this article, we discuss an anisotropic semivariogram as one measure of spatial correlation. The anisotropic semivariogram consider the the distance and direction between the pair of locations.

Rostaman [2] is conducted the study on the distribution of Bradysia ocellaris (BO) insects at 21 locations Oyster mushroom cultivation. That research use isotropic semivariogram analysis that is considered only the distance between pair of locations. The location of Oyster mushroom is made regularly with the distance between locations is 2.75 meters. Sari and Pasaribu [3] consider the distance and direction between the locations. The direction factor can be considered in this case because some information from the data are follow: (a) The BO insect always fly towards bright light, (b) The BO insects can fly lowly only 30 centimeters, so that the insects only flying around the nearby locations and (c) the location of the entrance where all the insects can enter into the room research. With consideration of directions, Sari and Pasaribu [3] established a model of anisotropic semivariogram which divides the area into three sectors based on the angle that is formed between locations. This anisotropic model is better for describing the distribustion of insects and have the subjectivity of the data.

II. ANISOTROPIC SEMIVARIOGRAM

Let a random variable $\{Z(s), s \in D \subseteq R^d\}$ with a positive integer d, based on the set of locations $s=\{s_1, s_2, \ldots, s_n\}$, will be observed a pattern of spatial relationship between random variables. Let a pairs of random variables $Z(s_i)$ and $Z(s_j)$ with $i \neq j$. We select the pairs of random variables that are separated or the same as far as h, and write that pairs as $Z(s_i)$ and $Z(s_i+h)$. Define a random variable $B_i(h)$ that specifying the difference between pairs of random variables $Z(s_i)$ and $Z(s_i+h)$ for $i = 1,2,...,n$. Suppose that h is formed from pairs of random variables have m pieces, say h_k for $k = 1,2,...,m$ with k is the number of h. Then there are m random variable $B_i(h_k)$ that obtained. Variance of $B_i(h)$ or $Z(s_i)$–$Z(s_i+h_k)$ can be written as $\gamma(Z(s_i)-Z(s_i+h_k))$. γ is called semivariogram. In general, that semivariogram can be formulated as a function of a norm matrix or written as $\gamma(\|A(s_i-(s_i+h_k))\|) = \gamma(\|A(h)\|)$ with **h** is a the distance vector and matrix A is a function of **h**.

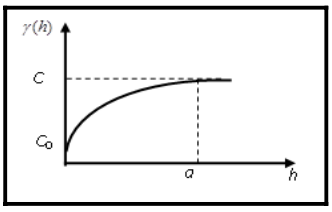


Fig.1. The example of semivariogram models with three parameters. There are nugget effect (C_0), sill (C) and range (a).

From the semivariogram that calculated for each pair of location, there are several candidates model which can be fitted on the semivariogram. Every model has three parameters to be estimated. There are nugget effect (C_0), sill (C) and range (a). Nugget effect is the variance of the

difference between pairs of observation at h = 0. Sill is the semivariogram value that constant for a given h where there is no correlation between pairs of observations. While range is the maximum distance between a pair of locations that have a spatial correlation.

The relationship between pair of locations can depend on distance, direction, or both of distance and direction. However, many spatial phenomena show the relationship between pair of locations that depend on both of distance and direction. The phenomena that only depend on distance can be analyzed by an isotropic semivariogram. The phenomena that depend on distance and direction can be analyzed by an anisotropic semivariogram. So, the anisotropic semivariogram is model to descrieb the relationship between pair of locations that depend on distance and direction. First, the anisotropic semivariogram is developed by Isaaks and Srivastava (1989). Then, this research is continued by Eriksson and Siska (2000), Ecker and Gelfand (2003) and Budrikaitė and Dučinskas (2005).

Eriksson and Siska [4] explain that the anisotropic semivariogram is divided into three types. There are based on the parameter range, sill and nugget effect. There are geometric or range anisotropic, zonal or sill anisotropic and nugget anisotropic. The geometric anisotropic occurs because the difference of ranges for difference directions. The zonal anisotropic show the difference of sill for difference directions. And the nugget anisotropic show the difference of nugget effect for difference directions. However, the geometric anisotropic and zonal anisotropic are often discussed by researchers. The difference between two anisotropic models is shown in Figure 2.

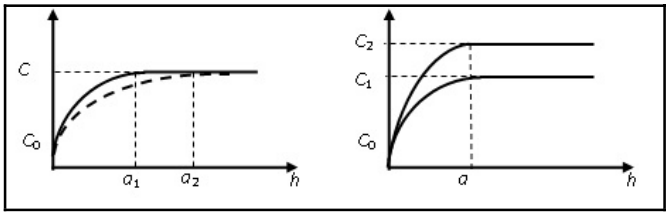


Fig.2. Two types of anisotropic semivariogram models are geometric anisotropic model (left) and zonal anisotropic model (right). The geometric anisotropic model have difference of range, a_1 and a_2, for difference directions. The zonal anisotropic model have difference of sill, C_1 and C_2, for difference directions.

Ecker and Gelfand [5] introduce the anisotropic semivariogram as a function of the distance between pair of locations that expressed as a vector **h**. Semivariogram formulated as follows $\gamma(s_i - s_j) = \gamma(\| h_{ij}^* \|) = \gamma(\| \mathrm{A}h_{ij} \|)$ where s_i and s_j are two location that separated by h_{ij}^*, $\mathrm{h_{ij}}$ is distance vector between s_i and s_j, and A is nonsingular matrix.

Budrikaitė and Dučinskas [6] describes the geometric anisotropic through the geometric transformation of each coordinates of locations. The geometric transformations are translation, rotation, and scaling. While, the zonal anisotropic more complex than the geometric anisotropic. The zonal anisotropic phenomena usually occurs coincide with anisotropic geometry. Thus, the formulation of zonal anisotropic can be expressed as the sum of isotropic and geometric anisotropic models.

III. GEOMETRIC ANISOTROPIC

The geometric anisotropic shows the strong correlation in a particular direction than the other directions. This model is derived from the isotropic model with the transformation of each coordinate locations. For the case of two dimensions, semivariogram is formulated as follows:

$$\gamma(Z(s_i) - Z(s_j)) = \gamma(\| \mathrm{A}(s_i - s_j) \|). \quad (1)$$

Semivariogram is defined as function from the norm of matrix A with A is constructed from pairs of locations. Matrix norm is a vector norm on K that size m x n. If $\|\mathrm{A}\|$ is a matrix norm, then apply:

a. $\|\mathrm{A}\| \geq 0$

b. $\|\mathrm{A}\| = 0$ if and only if $\mathrm{A} = 0$

c. $\|c\mathrm{A}\| = |c|\|\mathrm{A}\|, c \in \mathrm{K}$ and $\mathrm{A} \in \mathrm{K}^{m\,x\,n}$

d. $\|\mathrm{A} + \mathrm{B}\| \leq \|\mathrm{A}\| + \|\mathrm{B}\|$, with B is a matrix

e. $\|\mathrm{AB}\| \leq \|\mathrm{A}\|\|\mathrm{B}\|$

where K is a vector space. a–d are the property of vector norm and e show relationship the size of matrix A, B and AB.

In geometric anisotropic, matrix A is composition function of matrix R and S where R is rotation matrix and S is scalling matrix. So, matrix A can be written as A = SR with R and S is shown as follows

$$\begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \text{and} \begin{pmatrix} 1/a_{\max} & 0 \\ 0 & 1/a_{\min} \end{pmatrix} \quad (2)$$

where φ is angle who formed between vector **h** which is the distance of a pair of locations with the vertical axis, while $a_{\min}$ and $a_{\max}$ are minimum and maximum range respectively. Scalling is the way for standardization the range so that direction of semivariogram can be reduced become two directions. That directions show the strong influence among random variables.

In general, if we use the spherical model in semivariogram, then it is defined as follows:

$$\gamma(|\,\mathrm{h}\,|) = \gamma(r, \beta) = \Theta_1 \left(\frac{3}{2}\left(\frac{|\,\mathrm{h}\,|}{\Theta_2(\beta)} \right) - \frac{1}{2}\left(\frac{|\,\mathrm{h}\,|}{\Theta_2(\beta)} \right)^3 \right),$$

where β is argument of **h**, r is the length of **h**, Θ_1 is sill and Θ_2 is range that is defined as follows:

$$\Theta_2(\beta) = a_{\min} \sqrt{\frac{1}{\cos^2(\varphi - \beta)k^2 + \sin^2(\varphi - \beta)}}$$

where $a_{\min}$ is a minor axis of the ellipse anisotropic, $a_{\max}$ is major axis of the ellipse anisotropic, and k is an anisotropic ratio between $a_{\min}$ and $a_{\max}$. For example, Budrikaitė and Dučinskas [6] show the spatial relationship among locations in the Baltic Sea which have salinity are modeled with spherical anisotropic model as follows:

$$\gamma(|\,h\,|) = 0.08\left(\frac{3}{2}\left(\frac{|\,h\,|}{\Theta_2(\beta)}\right) - \frac{1}{2}\left(\frac{|\,h\,|}{\Theta_2(\beta)}\right)^3\right)$$

where $\Theta_2(\beta) = 2097.15\sqrt{\frac{1}{\cos^2(\frac{\pi}{4}-\beta)0.15^2 + \sin^2(\frac{\pi}{4}-\beta)}}$.

From equation (1), the geometric transformations that is used in anisotropic are rotation and scaling. Before we do the rotation and scalling, we often do the translation to move the coordinate axis to the centre of data. In this case, translation is included in transformation.

Definition. A mapping T of Euclidean space E^2 onto E^2 is said to be an isometry if for any X and Y in E^2, apply that the distance between TX and TY is equal to the distance between X and Y or written $d(TX, TY) = d(X,Y)$ [7].

Translation and rotation is an isometry which both are preserves the distance. In Theorem 1.22 [7], if T and R that respectively translation and rotation matrix are isometries so are TR or RT. T^{-1} and R^{-1} are also isometries. Moreover, translation and rotation can be avowed as the the multiplication from any reflections which is an isometry that is preserve shape and size.

A. Translation

Definition. Let ℓ be any line, and let m and n be perpendicular to ℓ. The transformation T_mT_n is called a translation along ℓ.

In the that definition, T_m and T_n are respectively translation matrix along the line l that is perpendicular to line m and n where m and n are parallel. Translation is an affine transformation from the Eulidean space where every point is moved or shifted by a certain distance and the same of direction. Translation is the multiplication of two reflections towards two parallel lines. Translation is an isometry that preserve the distance.

B. Rotation

Definition. If m and n are lines passing through a point P, the isometry R_mR_n is called a rotation about P.

Rotation is a transformation that is formed by rotating an object towards a certain point which is the center of rotation. In that definition, P is a centre point. R_m and R_n are respectively the rotation matrix to the lines m and n, where m and n are two lines that intersect in a point P.

Rotation is adjusted by certain angle that rotates clockwise or counter-clockwise. Rotation is the multiplication of two reflections towards two intersected line with two lines intersect is the center point of rotation.

C. Scaling

Scaling is a linear transformation that zoom in or out the object. Scaling is divided into two types. There are uniform scaling and directional scalling. Uniform scaling have the same scale for all directions so it is called homothety or dilatation. So that, the object that is resulted by scaling will be the same as the original object. However, the scale can different for each axis direction and that is called directional scaling as in equation (2). So that, directional scaling will change the shape and size of the object.

IV. CASE STUDY AND RESULTS

To determine the effect of each geometric transformation that has been described in Section III, we are used data with nine locations that is depicted in Figure 3.

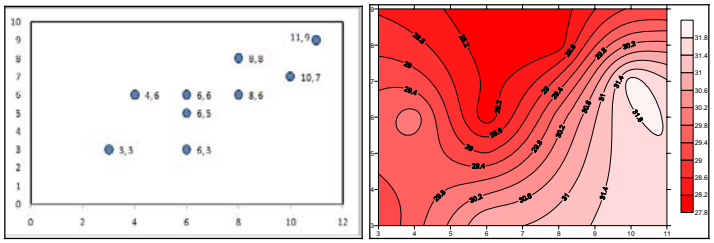


Fig.3. The plot of nine locations with pair of coordinate cartesian (left) and the contour of value from nine locations (right).

The above contour is constructed from each location coordinates and the observation value at each location. The contours show that the locations that have the similar observations value will be grouped in one line contours. So the gradation of color show the changes of observations value. From north to south, we can see that there is an increase in the intensity of observation value.

From the above data, transformations is performed on each coordinate of locations. There are translation, rotation, and scaling. The transformation process is follows:

1. Translation is done by moving or shifting each location coordinates, so the centre coordinates is located at centre of data. The translation operator is determined by the selection of centre of data that is obtained by finding the line that shows the farthest distance between pair of locations. So that the center point is the midpoint of that line. For this case, each location coordinates will be shifted by $\begin{bmatrix}-7\\-6\end{bmatrix}$. Suppose a new location coordinate is written as (x', y').
2. Each location coordinates that are translated, is rotated through an angle α that is counter-clockwise. Transformation matrix that represent the rotation is $R=\begin{bmatrix}\cos\alpha & \sin\alpha\\ -\sin\alpha & \cos\alpha\end{bmatrix}$ where α is determined by finding the direction or angle where the distance between the data center and a location are the farthest distance. For this case, the angle α is $\pi/3$. So that, the new coordinate is written as (x'', y'').
3. Scaling each location coordinates is done by multiplying the coordinates of each point with a diagonal matrix which has a different scale factor. Transformation matrix

that represent scaling is $S = \begin{bmatrix} 1/r_j & 0 \\ 0 & 1/r_k \end{bmatrix}$ where r_j is the longest radius and r_k is the shortest radius of ellipse that indicate the maximum range (a_{max}) and the range minimum (a_{min}).

For example, any coordinates that are transformed through translation, rotation, and scaling are presented in Table 1.

TABLE I. THE COORDINATE LOCATIONS AND THE RESULT OF TRANSFORMATION OF COORDINATE LOCATIONS

No Obs.	Trans-formation	Location	Realization
1	Original	(3,3)	29
	Translation	(-4,-3)	
	Rotation	(-5,0)	
	Scaling	(-1,0)	
2	Original	(4,6)	30
	Translation	(-3,0)	
	Rotation	(-2.4,1.8)	
	Scaling	(-0.48,1)	
3	Original	(6,6)	28
	Translation	(-1,0)	
	Rotation	(-0.8,0.6)	
	Scaling	(-0.16,0.33)	
4	Original	(6,5)	29
	Translation	(-1,-1)	
	Rotation	(-1.4,-0.2)	
	Scaling	(-0.28,-0.11)	
5	Original	(6,3)	31
	Translation	(-1,-3)	
	Rotation	(-2.6,-1.8)	
	Scaling	(-0.52,-1)	
6	Original	(8,8)	28
	Translation	(1,2)	
	Rotation	(2,1)	
	Scaling	(0.40,0.56)	
7	Original	(8,6)	30
	Translation	(1,0)	
	Rotation	(0.8,-0.6)	
	Scaling	(0.16,-0.33)	
8	Original	(10,7)	32
	Translation	(3,1)	
	Rotation	(3,-1)	
	Scaling	(0.6,-0.56)	
9	Original	(11,9)	29
	Translation	(4,3)	
	Rotation	(5,0)	
	Scaling	(1,0)	

This is the graph of coordinates of each point and their transformations.

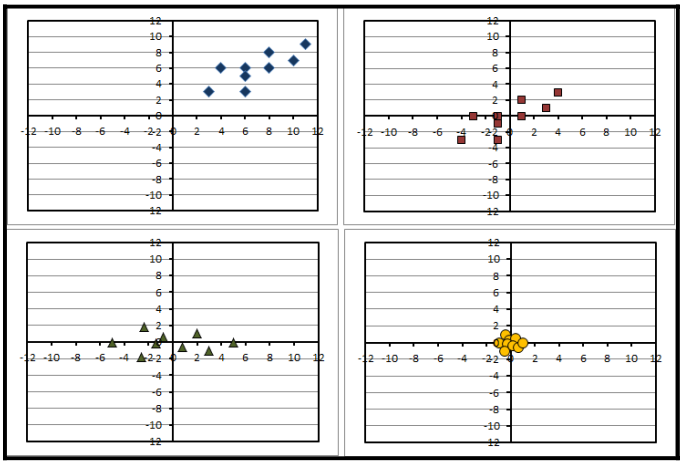


Fig. 4. The graph of nine location coordinates (◊) and the changes of location coordinates that is the result of translation (□), rotation (Δ) and scaling (o).

From that figure, each transformation has a purpose, as described in Section III. Translation and rotation are isometries which preserve the distances of pair of location and not change the shape and size of the distribution of data

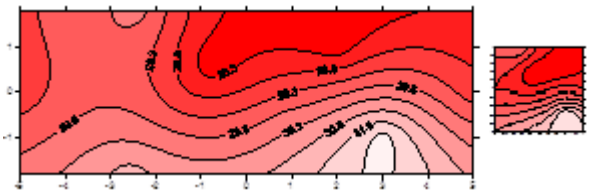


Fig.5. The comparison of contour location coordinates after rotation (left) and contour locatiion coordinates after scaling (right) with the same scale.

From Fig. 4 and 5 are obtained some information. The results of the original contour and contour in Fig.5 (left) show no significant change of the data structure. There are only rotate the coordinates axis. The structure of contour, that is parallel with angle $\pi/3$ against the horizontal axis, rotate into a horizontal direction. While from Fig.5 (right), location coordinates change into one unit at an interval [-1,1]. So that, the contours become more dense than the distance of original data.

V. CONCLUSIONS

In the geometric anisotropic semivariogram, the geometry transformations that involve three types of transformations are translation, rotation, and scaling. Translation and rotation are isometries which preserve the distance between the pair of locations. While the scaling does not preserve the distance between the pair of locations but the distance is made with the standard distance in the interval [-1,1]. So that, each transformation have purpose in changing the location coordinates.

REFERENCES

[1] W. Shi, *Principles of Modelling Uncertainties in Spatial Data and Spatial* Analyses, Taylor and Francis Group, 2010.

[2] Rostaman, "The Dynamics of Population of Sciarid Flies (Diptera: Sciaridae) on Oyster Mushroom Cultivation", Ph.D. Disertation, Bandung Institute of Technology, 2005.

[3] K.N. Sari and U.S. Pasaribu, The Comparison of Isotropic and Anisotropic Semivariogram for Gauss Model, *American of International Proceeding* from ICMNS (International Conference of Mathematics and Natural Science), ITB, 2012. (accepted)

[4] M. Eriksson and P.P. Siska, Understanding Anisotropy Computation, *Mathematical Geology*, 32, 6, 2003, pp.683-700.

[5] M.D. Ecker and A.E. Gelfand, Spatial modeling and prediction under stationary non-geometric range anisotropy, *Environmental and Ecological Statistics*, vol. 10, 2003, pp. 165-178.

[6] L. Budrikaitė and K. Dučinskas, Modelling of Geometric Anisotropic Spatial Variation, *Mathematical Modelling and Analysis, Proceeding of the 10th International Conference MMA2005 & CMAM2*, 2005, pp. 361-366.

[7] P.J. Ryan, "Euclidean and Non-Euclidean Geometry", Cambridge University Press, Australia, 1997.

[8] E.H. Isaaks and R.M. Srivastava, "Applied Geostatistics", Oxford University Press, New York, 1989.

[9] K.N. Sari, U.S. Pasaribu and O. Neswan, The Effect of Transformation for Anisotropic Semivariogram Model, *WAJS*

from ICMS (International on Computing, Mathematics, and Statistics), UTM Malaysia, 2013. (accepted)

Two-Level Feature Selection for Naive Bayes with Kernel Density Estimation in Question Classification based on Bloom's Cognitive Levels

Catur Supriyanto[1], Norazah Yusof[2], Bowo Nurhadiono[1], Sukardi[3]
[1]Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia
[2]Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Malaysia
[3]Program of Informatics Engineering, College of Computing and Informatics Adhi Guna, Palu, Indonesia
caturs@research.dinus.ac.id, norazah@utm.my, bnurha@yahoo.com, sukardi@stmikadhiguna.ac.id

***Abstract*— This paper proposes a two-level feature selection to improves Naïve Bayes with kernel density estimation. The performance of the proposed feature selection is evaluated on question item set based on Bloom's cognitive levels. This two-level feature selection contains of filter and wrapper based feature selection. This paper uses chi square and information gain as the filter based feature selection and forward feature selection and backward feature elimination as the wrapper based feature selection. The result shows that the two-level feature selection improves the Naïve Bayes with kernel density estimation. The combination of chi square and backward feature elimination give more optimal quality than the other combination.**

Keywords—bloom's cognitive level; naïve bayes; kernel density estimation; filter based feature selection; wrapper based feature selection

I. INTRODUCTION

Bloom's taxonomy has been developed by Benjamin Samuel Bloom [1] and widely used to categorize the question item set based on student's deep understanding. Six levels of Bloom's taxonomy are knowledge, comprehension, application, analysis, synthesis, and evaluation. Higher level of bloom's taxonomy indicates the deeper level of student's knowledge [2]. Bloom's taxonomy is used for designing and assessing learning objectives. Teacher or lecturer needs to design the learning objective and question item set to know the student understanding. In fact, it is difficult to design the quality of test items based on bloom's taxonomy. Therefore, intelligence system helps the teacher for designing the test items.

Automatic questions classification into Bloom's cognitive levels using Support Vector Machines (SVM) has been studied by Yahya and Osman [3]. In their research, SVM produced good performance, but the complexity of SVM is very high [4]. Another research on Bloom's cognitive levels using machine learning algorithm was done by Norazah Yusof and Chai Jing Hui [5] [6]. They used Backpropagation Neural Network to determine item set question into six levels of Bloom's taxonomy. They proposed Document Frequency (DF) and Category Frequency-Document Frequency Method (CF-DF) as feature selection to reduce the complexity of Backpropagation algorithm. The result show that DF feature reduction method can be considered as a more effective feature reduction method than the whole feature set or the CF-DF feature reduction method. However, execution time of training Backpropagation algorithm is very slow [7]. Then, the accuracy of question classification still can be improved.

In our work, Naïve Bayes with kernel density estimation using filter and wrapper based feature selection is proposed. Naïve Bayes has good accuracy and speed for the large training dataset. However, Naïve Bayes cannot overcome small training dataset [8]. Therefore, this paper proposed two-level feature selections to select the important features to improve the performance of Naïve Bayes on small dataset.

Based on classification criterion, feature selection can be classified into filter and wrapper based feature election [9]. Filter based feature selection selects the informative features by ranking them according to a criterion function [10]. The wrapper method takes feature selection and pattern classification as a whole and evaluates feature subsets based on classification results directly [11].

This paper use Chi Square and Information Gain to select the best individual features. There are several commonly used feature selection methods such as Document Frequency (DF), Information Gain (IG), Mutual Information (MI), Chi-square Statistic (CHI), Expected Cross Entropy Weight of Evidence text. Yang and Pederson [12] pointed out that IG and CHI are the best in text categorization. Next, Wrapper methods select the best feature from filter based feature selection based on classification results. This paper used forward feature selection (FFS) and backward feature elimination (BFE) as wrapper based method. The difference between forward selection and backward elimination is forward selection starts with the empty set of features, and vice versa.

Wrapper methods are widely recognized as a superior alternative in supervised learning problems, but on account of the number of executions that the search process requires results in a high computational cost than filters methods [13]. This paper expects the using of filter based feature selection

can reduce the computational time of wrapper based feature selection.

The remainder of this paper is organized as follows. Section 2 introduces the theoretical background related to question classification based on Bloom's cognitive levels and feature selection. Section 3 presents our conducted experiment and the result. Section 4 is devoted for conclusion and future work.

II. THEORETICAL BACKGROUND

A. Cognitive Level of Bloom's Taxonomy

Educational objectives in the Bloom's Taxonomy contains of three domains: cognitive, affective and psychomotor. Cognitive domain involves the intellectual skills, while the acts as the emotional and attitudinal are affective domain component and the psychomotor domain involves the physical skills [14]. This paper focused on cognitive domain of bloom's taxonomy. Table I shows the keyword relate to cognitive domain of bloom's taxonomy.

TABLE I. KEYWORDS USED IN KOGNITIVE DOMAIN [15]

Category	Keywords
Knowledge	Defines, describes, identifies, knows, labels, lists, matches, names, outlines, recalls, recognizes, reproduces, selects, states.
Comprehension	Comprehends converts, defends, distinguishes, estimates, explains, extends, generalizes, gives examples, infers, interprets, paraphrases, predicts, rewrites, summarizes, translates.
Application	Applies, changes, computes, constructs, demonstrates, discovers, manipulates, modifies, operates, predicts, prepares, produces, relates, shows, solves, uses.
Analysis	Analyzes, breaks down, compares, contrasts, diagrams, deconstructs, differentiates, discriminates, distinguishes, identifies, illustrates, infers, outlines, relates, selects, separates.
Synthesis	Categorizes, combines, compiles, composes, creates, devises, designs, explains, generates, modifies, organizes, plans, rearranges, reconstructs, relates, reorganizes, revises, rewrites, summarizes, tells, writes.
Evaluation	Appraises, compares, concludes, contrasts, criticizes, critiques, defends, describes, discriminates, evaluates, explains, interprets, justifies, relates, summarizes, supports.

B. Naïve Bayes Classifier

Naïve Bayes is a simple probabilistic machine learning algorithm [16]. Naïve Bayes is one of classification algorithm that needs training data to predict the unknown data. Naïve bayes classification is computed based on Bayesian theory.

$$P(C \mid d) = \frac{P(C)P(d \mid C)}{P(d)} \quad (1)$$

$P(C \mid d)$ is the probability of class given a document, or the probability that a given document d belongs to a given class C. $P(d)$ is the probability of a document. $P(d)$ can be ignored since $P(d)$ is a Constance divider to every calculation. $P(C)$ is the probability of a class, we can compute it from the number of documents in the category divided by documents number in all categories. $P(d \mid C)$ represents the probability of document given class, and documents can be modeled as sets of words, thus the $P(d \mid C)$ can be written like:

$$P(d \mid C) = \prod P(w_i \mid C) \quad (2)$$

$$P(C \mid d) = P(C)\prod P(w_i \mid C) \quad (3)$$

$P(w_i \mid C)$ is the probability that the i^{th} word of a given document occurs in a document from class C, and this can be computed as follows:

$$P(w_i \mid C) = \frac{T_c + \lambda}{M + \lambda N} \quad (4)$$

Where T_c is the number of times the word w_i that occur in class C, M is the number of words in category C, N is The size of the vocabulary table, λ is the positive constant, usually 1, or 0.5 to avoid zero probability.

C. Naïve Bayes with Kernel Density Estimation

Kernel Density Estimation (KDE) can manipulate quantitative attributes for naïve-Bayes [17]. To deal with the quantitative data, naïve Bayes use normal Gaussian distribution.

$$f = g(x, \mu_i, \sigma_c) = \frac{1}{\sqrt{2\pi\sigma_c}} e^{-\frac{(x-\mu_i)^2}{2\sigma_c^2}} \quad (5)$$

$$P(D = d \mid C = c) = \frac{1}{n}\sum_i g(x, \mu_i, \sigma_c) \quad (6)$$

Where i is the range of training data for the attribute x in class C, $\mu_i = x_i$ and $\sigma_c = \frac{1}{\sqrt{n_c}}$, n_c is the number of document in class C.

D. Filter Based Feature Selection

Filter based feature selection evaluate the features by simply ranking the features based on the score of each feature. This feature selection is more fast compare to wrapper based feature selection [18]. Two the best of filter based feature selections are Chi Square (CHI) and Information Gain (IG). The equation of CHI and IG is shown in (7) and (9), respectively.

$$\chi^2(t, C_i) = \frac{N(A \times D - B \times C)^2}{(A+B)\times(C+D)\times(A+C)\times(B+D)} \quad (7)$$

$$\chi_{avg}{}^2 = \sum_{i=1}^{m} P(C_i)\chi^2(t,c) \quad (8)$$

Where A is the number of documents in category C_i and containing t, B is the number of documents not in category C_i and containing t, C is the number of documents in category C_i and not containing t, D is the number of documents not in category C_i and not containing t, N is the total number of documents.

$$IG(t) = -\sum_{i=1}^{n} P(C_i)\log P(C_i) + P(t)\sum_{i=1}^{n} P(C_i \mid t)\log P(C_i \mid t)$$

$$+P(\bar{t})\sum_{i=1}^{n} P(C_i \mid \bar{t})\log P(C_i \mid \bar{t}) \quad (9)$$

Where, $P(C_i)$ is the ratio of the number of C_i category documents to the number of all training documents, $P(t)$ is the ratio of the number of documents which include term t to the number of all training documents; $P(C_i \mid t)$ is the ratio of the number of documents which include term t and belong to C_i category to the number of documents which include term t in all training samples; $P(\bar{t})$ is the ratio of the number of documents which don't include term t to the number of all training documents; $P(C_i \mid \bar{t})$ is the ratio of the number of documents which don't include term t but belong to C_i category to the number of documents which don't include term t in all training samples;

E. Wrapper based Feature Selection

Wrapper based feature selection utilizes classifier algorithm to evaluate the feature subset by measuring the cross-validation. Commonly wrapper based feature selection are forward feature selection (FFS) and backward feature elimination (BFE). The FFS algorithm starts with an empty feature subset. In each iteration, adding one feature each forward step until a predefined number of features is reached. For one step, each candidate feature is separately added to the current subset and then evaluated. The feature that induced the highest improvement is included in the resulting subset.

1. Forward selection start from the empty set $F_k = \{\emptyset\}$
2. Iterate
   a. Select the next best feature j^+ to add to F_k with most significant cost reduction
   $j^+ = \arg\max J(F_k + j)$
   b. Update $F_{k+1} = F_k + j^+$, $k = k+1$

Fig. 1. The Algorithm of Forward Feature Selection

BFE is difference from FFS. BFE starts with the full features subset. The feature is removed when the feature decrease the accuracy of classifier. Fig. 1 and Fig. 2 show the algorithm of FFS and BFE, respectively.

1. Start with the full set $F_k = \{1,...,n\}$ and $k = n$
2. Iterate
   a. Remove the worst feature
   $j^- = \arg\max J(F_k - j)$
   b. Update $F_{k-1} = F_k - j^-$, $k = k-1$

Fig. 2. The Algorithm of Backward Feature Elimination

III. Experiment And Result

A. Question Dataset

Dataset of this paper was collected from Chai Jing Hui [6]. The question item dataset contains 274 items belongs to six bloom taxonomy levels. The distribution of the question item dataset is shown in table II. In order to have good performance of the experiment, *k*-fold cross-validation has been used. We used 10-fold cross-validation using stratified sampling in the experiment.

TABLE II. Distribution of Question Dataset

Name of Category	Number of Question
Knowledge	28
Comprehension	44
Application	41
Analysis	48
Synthesis	59
Evaluation	54

B. Performance Measures

The performance evaluation of this study was measured using confusion matrix (Table III). Confusion matrix can produces accuracy.

TABLE III. Confusion Matrix

	Actual Positive	Actual Negative
Predicted Positive	*TP*	*FP*
Predicted Negative	*FN*	*TN*

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

C. Experiment results

This subsection presents the result of experiment. Our experiment used tokenization, stopword removal, and stemming for preprocessing of question item set document.

This paper used stratified 10-fold validation to test the classification result. Cross validation is standard evaluation technique in pattern classification, in which the dataset is split into n parts (folds) of equal size, $n-1$ folds are used to train the classifier, and the *n-th* fold that was held out is then used to test it [19]. Experiments were conducted with RapidMiner 5.3.005.

Several conditions were observed by Naïve Bayes performance: Naïve Bayes Kernel Density Estimation without feature selection (NBK), Naïve Bayes Kernel Density Estimation using filter based feature selection, Naïve Bayes Kernel Density Estimation using wrapper based feature selection, and Naïve Bayes Kernel Density Estimation using the combination of filter and wrapper based feature selection. This study used 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% for size of important feature selection (left terms). The using of 10% for size of important feature selection means that the classification process only used 10% of the total terms. This 10% of selected terms automatically generated from filter based feature selection. Table IV shows the performance of FFS and BFE on Naïve Bayes with Kernel Density Estimation. Based on table IV, FFS produces better accuracy and computation time than BFE.

TABLE IV. ACCURACY OF FORWARD FEATURE SELECTION AND BACKWARD FEATURE ELIMINATION ON NAÏVE BAYES WITH KERNEL DENSITY ESTIMATION

	Accuracy	Computation Time (seconds)
NBK	57.61% +/- 5.64%	2
NBK+FFS	67.18% +/- 7.77%	748.8
NBK+BFE	60.9% +/- 11.95%	1861.8

TABLE V. ACCURACY OF TWO-LEVEL FEATURE SELECTION ON NAÏVE BAYES WITH KERNEL DENSITY ESTIMATION

Left Terms (%)	CHI	CHI+ FFS	CHI+ BFE	IG	IG+ FFS	IG+ BFE
10	73,28	67,61	75,17	73,69	65,71	75,94
20	76,59	72,29	**77,8**	73,33	73,32	76,28
30	71,49	62,87	74,48	68,28	68,69	72,59
40	71,1	71,26	74,5	60,19	73,43	63,8
50	71,48	66,43	73,07	60,52	68,32	63,1
60	67,46	74,11	70,09	60,86	70,11	63,48
70	66,4	54,42	70,11	57,92	72,63	59,81
80	64,58	70,79	66,79	56,85	67,12	59,22
90	65,3	72,69	68,23	56,11	67,53	58,78

The accuracy of two-level feature selection is shown in table V. It can be seen that the combination of CHI and BFE has the best accuracy. Our approach improves the performance of Chai Jing Hui approach [5]. The accuracy of our proposed method reaches 77.8%. The using of CHI and IG also reduce the computational time of FFS and BFE. For the best accuracy in table V (77.8%), CHI reduces the computational time of BFE from 1861.8 seconds to 138 seconds. Detailed computational time of our proposed question classification technique is shown in table VI. Fig. 3 demonstrated the performance of two-level feature selection on Naïve Bayes with Kernel Density Estimation.

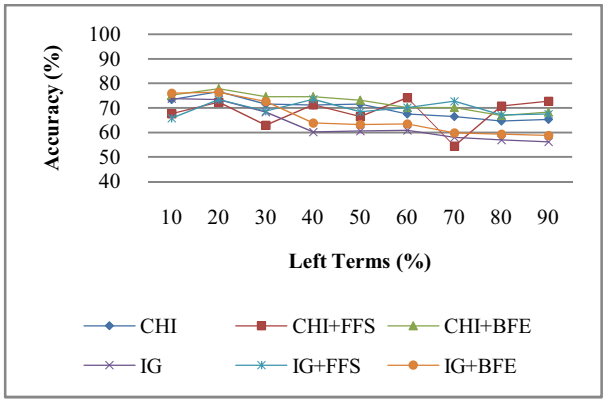


Fig. 3. Accuracy of two-level feature selection on Naïve Bayes with Kernel Density Estimation

TABLE VI. COMPUTATIONAL TIME OF TWO-LEVEL FEATURE SELECTION ON NAÏVE BAYES WITH KERNEL DENSITY ESTIMATION (IN SECONDS)

Left Terms (%)	CHI	CHI+FFS	CHI+BFE	IG	IG+FFS	IG+BFE
10	1	39	43	3	53	21
20	1	73.8	**138**	2	139.2	72
30	2	258	213	3	301.2	153
40	1	270	431.4	3	392.4	307.8
50	2	333	913.8	2	624.6	490.8
60	2	272.4	922.2	3	670	1054.8
70	2	554.4	1212	2	812.4	1042.8
80	2	670	1764.6	2	860	1400
90	2	1021.8	1999	3	954	1993.2

IV. CONCLUSION

This paper studies the filter and wrapper based feature selection on Naïve Bayes with kernel density estimation to work on question classification according to cognitive level of bloom's taxonomy. We proposed the combination of filter and wrapper based feature selection to improve the performance of Naïve Bayes with Kernel Density Estimation. The result shows that our approach improves the performance of Naïve Bayes with Kernel Density Estimation. The combination of CHI and BFE shows the best accuracy compared to other combination of filter and wrapper based feature selection. Our result also shows that the using of filter based feature selection reduce the computational time of wrapper based feature selection.

In the future, feature selection based on evolutionary computation technique, such as Genetic Algrotihm and Particle Swarm Optimization should be experimented and combined

with filter or wrapper based feature selection in classifying exam question into Bloom's cognitive levels.

REFERENCES

[1] Benjamin Samuel Bloom, *Taxonomy of Educational Objectives, Handbook I. The Cognitive Domain.* New York: David McKay Co Inc., 1956.

[2] Timothy Highley and Anne E. Edlin, "Discrete Mathematics Assessment Using Learning Objectives Based on Bloom's Taxonomy," in *Proceedings of the 39th IEEE international conference on Frontiers in education conference*, 2009.

[3] Anwar Ali Yahya and Addin Osman, "Automatic classification of questions into Bloom's cognitive levels using support vector machines ," in *The international arab conference on information technology*, 2011.

[4] M. Janaki Meena and K. R. Chandran, "Naïve Bayes Text Classification with Positive Features Selected by Statistical Method," in *First International Conference on Advanced Computing*, 2009.

[5] Norazah Yusof and Chai Jing Hui, "Determination of Bloom's Cognitive Level of Question Items using Artificial Neural Network," in *International Conference on Intelligent Systems Design and Applications*, 2010, pp. 866-870.

[6] Chai Jing Hui, "Feature Reduction for Neural Network in Determining the Bloom's Cognitive Level of Question Items," Universiti Teknologi Malaysia, Computer Science Master's Project Report 2009.

[7] Ji Peirong, Wang Peng, Zhao Qin, and Zhao Li, "A new parallel back-propagation algorithm for neural networks," in *IEEE International Conference on Grey Systems and Intelligent Services*, 2011, pp. 807-810.

[8] Li Lei, Huang Yu-guang, and Liu Zhong-wan, "Chinese text classification for small sample set," *The Journal of China University of Posts and Communications*, vol. 18, no. 1, pp. 83–89, 2011.

[9] R.Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1-2, pp. 273–324, 1997.

[10] Yuh-Jye Lee, Chien-Chung Chang, and Chia-Huang Chao, "Incremental Forward Feature Selection with Application to Microarray Gene Expression Data," *Journal of Biopharmaceutical Statistics*, vol. 18, no. 5, pp. 827-840, 2008.

[11] K. Z. Mao, "Orthogonal Forward Selection and Backward Elimination Algorithms for Feature Subset Selection," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 34, no. 1, pp. 629-634, February 2004.

[12] Yiming Yang and Jan O.Pedersen, "A Comparative Study on Feature Selection in Text Categorization," in *Proceedings of ICML-97*, 1997, pp. 410-420.

[13] Ranjit Abraham, Jay B. Simha, and S. Sitharama Iyengar, "Effective Discretization and Hybrid feature selection using Naïve Bayesian classifier for Medical datamining," *International Journal of Computational Intelligence Research*, vol. 5, no. 2 , pp. 116–129, 2009.

[14] David Andrich, "Framework relating outcomes based education and the taxonomy of educational objectives. ," *Studies in educational Evaluation*, vol. 28, pp. 35-59, 2002.

[15] Clark. C, "Learning Domains of Bloom's Taxonomy: The Three Types of Learning," , 2001.

[16] Fadi Thabtah, Mohammad Ali H. Eljinini, Mannam Zamzeer, and Wa'el Musa Hadi, "Naïve Bayesian Based on Chi Square to Categorize Arabic Data," *Communications of the IBIMA*, vol. 10, no. 20, pp. 158-163, 2009.

[17] Jingli Lu, Ying Yang, and Geoffrey I. Webb, "Incremental discretization for Naïve-Bayes classifier," in *Proceedings of the Second international conference on Advanced Data Mining and Applications*, 2006, pp. 223-238.

[18] Amir Navot, "On the Role of Feature Selection in Machine Learning," Hebrew University, PhD Thesis 2006.

[19] D.M.Chandwadkar and Dr. M.S.Sutaone, "Role of Features and Classifiers on Accuracy of Identification of Musical Instruments," in *CISP2012*, 2012.

Using Estimated Arithmetic Means of Accuracies to Select Features for Face-based Gender Classification

Ivanna K. Timotius[1], Iwan Setyawan[2]
Department of Electronic Engineering
Satya Wacana Christian University
Salatiga, Indonesia
Email: [1]ivanna.timotius@ieee.org, [2]iwan.setyawan@ieee.org,

***Abstract*—Selecting the appropriate features is essential in building a good classifier. This paper aims to use the approach of estimating the arithmetic means of accuracies (ameans) in selecting the features used in a face-based gender classification. In a face-based gender classification, there are many pixels of the input image that may not aid the classification process, such as those belonging to the background. The experiments show that this approach outperforms the approach based on mean difference especially on the data having relatively high variance by up to 2.14%. Compared to the classifier which does not use any feature selection approach, implementing the feature selection approach based on ameans estimation in a gender classification problem increases the accuracy by up to 7.86%. The experiments also show that the face-based gender classifications rely on the presence of long hair on subjects in the images to make their decision.**

Keywords*—Feature Selection, Arithmetic Means of Accuracies, Gender Classification

I. Introduction

Classifiers make their final decisions based on the values of the features that are passed to them. Good features will help the job of the classifiers. We might obtain several values of features from an object to be classified. If the number of the values of features is n, then these values can be represented by an n-dimensional feature vector.

The performance of a classifier depends on the features that are chosen to represent the data. A non-discriminative feature might lower the classifier's performance. On the other hand, if the number of features is too small, the classifier might not have enough information. An approach to eliminate the non-discriminative features while maintaining the discriminative features can help the classifier yield a better performance.

There are several approaches used to select the features based on their discriminating power. One possible approach is based on the difference between the class means of the values of the features [1]. The approach in [1] is implemented on an abnormality detection of endoscopy images. Another approach is based on the values of feature means and the values of feature variances between classes [2]. The means and variances are used to estimate the maximum arithmetic means of accuracies (ameans) that can be obtained from each feature.

This paper implements the approach of estimating ameans in the feature selection of a face-based gender classification. Face-based gender classification is a challenging subject that has been widely investigated in the literatures [3], [4], [5], [6]. The choice of face-based gender classification as a base with which to evaluate our approach is suitable. This is due to the fact that in this application there are many pixels in the input image that may not aid the performance of the classifier, such as those belonging to the background. This implementation aims to investigate how big the classifier performance improvement would be if the features are selected properly. This paper also aims to discover the features which are important in classifying gender based on the face images.

The remaining parts of this paper are organized as follows. Section II discusses the approach of estimating the arithmetic means. Section III explains the experiment scheme used in building the gender classifier. Section IV gives the experiment results and analysis. Finally, Section V presents our conclusions.

II. Arithmetic Means of Accuracies Estimation for Feature Selection

Arithmetic means of accuracies can be defined as the arithmetic means between the true positive rate and the true negative rate [7]. For a balanced testing dataset, the value of ameans is the same as the value of accuracy. The approach of estimating the ameans between two features has been introduced by [2].

The gender classification implemented in this paper use all pixels of the input image as an initial feature vector. The features are then selected based on their statistical properties. Two approaches to achieve this are compared to do this task, i.e. the approach based on the value of feature mean differences and the approach based on estimated ameans.

For the approach of selecting features based on the estimated ameans value, the densities of the values of features for each class, $f_N(x)$ and $f_P(x)$, are first estimated [8]. A threshold, t, for each feature is picked between the means of $f_N(x)$ and $f_P(x)$. This threshold should be chosen at the point where the two densities intersect. In the event that such a point is absent, the point having the closest differences of the densities as shown in Equation 1 is chosen. The value of ameans is then estimated using Equation 2 where $F_N(x)$ and $F_P(x)$ are the cumulative density functions of $f_N(x)$ and $f_P(x)$, respectively.

$$t = \arg\min_x |f_N(x) - f_P(x)| \tag{1}$$

Fig. 1: Examples of the images from the first dataset

Fig. 2: Examples of the images from the second dataset

$$\text{Estimated Ameans} = \frac{1}{2}\{F_N(t) + (1 - F_P(t))\} \qquad (2)$$

III. Experiment

The experiment is conducted using two different datasets. The first dataset contains the frontal face images from the VISiO lab dataset [9]. This dataset contains 50 male and 50 female 64×64 grayscale images with controlled lighting condition. The second dataset contains 70 male and 70 female grayscale images with resolution of 64×96. These images are taken from various sources. Furthermore, unlike the first dataset, most face images of this dataset are not fully frontal. Some examples of the images in the datasets are shown in Figure 1 and Figure 2.

Since many images in the second dataset are not fully frontal face images, the images in this dataset are first flipped along the vertical axis, then the flipped images are averaged with the original images. The results of this process are shown in Figure 3.

The feature selection implemented in this paper is conducted based on the class mean difference as used in [1] and the estimated ameans introduced by [2]. Figure 4 and Figure 5 show the average input image of each class, the mean difference, and the estimated arithmetic means for the two datasets.

In the mean difference approach experiment, the features having mean difference above a threshold t_m are passed to a k-nearest neighbor classifiers, with $k = 1$ and $k = 3$. Whereas in the ameans estimating approach experiment, the features having estimated ameans above a threshold t_a are passed to a k-nearest neighbor classifier. The experiments are conducted using leave-one-out cross validation.

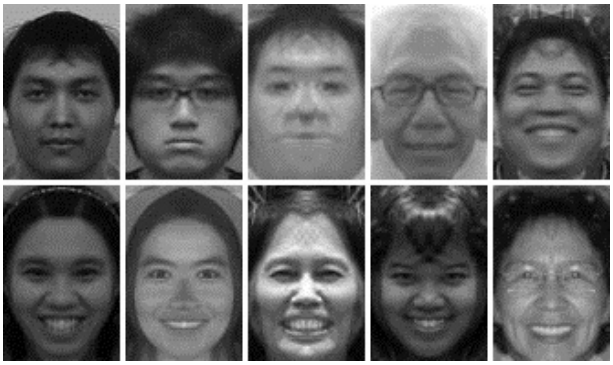
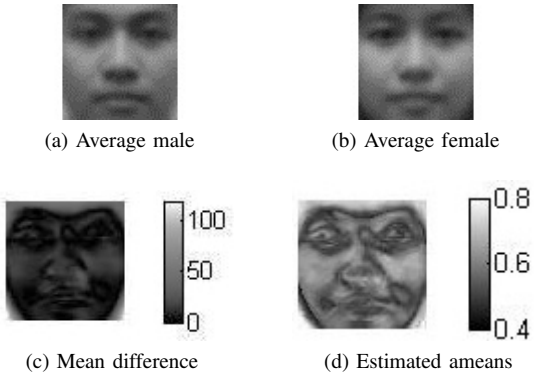


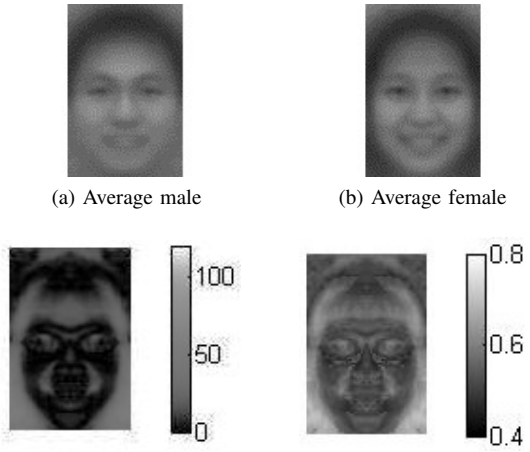
Fig. 3: Flipped and averaged images in the second dataset



(a) Average male (b) Average female

(c) Mean difference (d) Estimated ameans

Fig. 4: The first dataset properties



(a) Average male (b) Average female

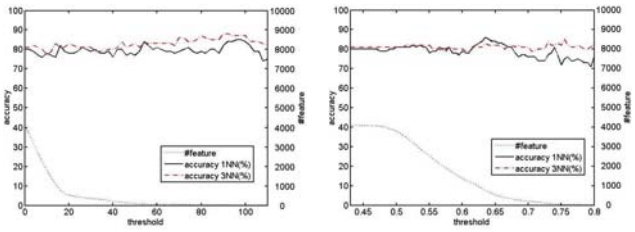
(c) Mean difference (d) Estimated ameans

Fig. 5: The flipped second dataset properties

IV. Results and Analysis

A. Results from the first dataset

The classifier accuracies resulting from the experiments using different values of t_m and t_a are depicted in Figure 6. If we use all 4096 image pixels as features, we will achieve 80% classification accuracy for 1-nearest neighbor classifier and 81% classification accuracy for 3-nearest neighbor classifier.



(a) Mean difference (b) Estimated ameans

Fig. 6: Experimental results from the first dataset

(a) $t_m = 98$ (b) $t_m = 100$ (c) $t_a = 0.635$ (d) $t_a = 0.755$

Fig. 7: The image pixels used as features

(a) male (b) male (c) female (d) female (e) female

Fig. 8: Misclassified images by all approaches

Using the approach based on the mean difference and 1-nearest neighbor classifier, the classification accuracy increases to 85% when $t_m = 98$ using 22 features. If 3-nearest neighbor classifier is used instead of 1-nearest neighbor classifier, the classification accuracy increases to 87% when $t_m = 100$ using 20 features.

The approach of selecting features based on estimating ameans increases the accuracy to 86% when $t_a = 0.635$ using 763 features for 1-nearest neighbor classifier and 85% when $t_a = 0.755$ using 39 features for 3-nearest neighbor classifier.

The image pixels used to achieve the best classification results are depicted in Figure 7. In this Figure, the eliminated pixels are represented by black pixels and the selected features are represented by the average value of the feature from all images in the dataset.

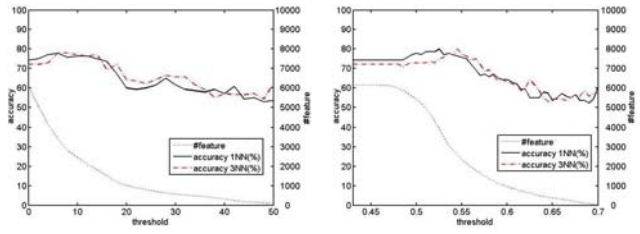
Figure 8 shows the images that are misclassified by all approaches and classifiers. Figure 9 shows the images that are misclassified by the approach based on mean difference using 1-nearest neighbor classifier, but are correctly classified by the approach based on ameans estimation. Figure 10 shows the images that are misclassified by the approach based on ameans estimation using 1-nearest neighbor classifier, but are correctly classified by the approach based on mean difference.

(a) male (b) male (c) female (d) female (e) female

Fig. 9: Some of the misclassified images by the approach based on mean difference

(a) male (b) male (c) female (d) female (e) female

Fig. 10: Some of the misclassified images by the approach based on ameans estimation



(a) Mean difference (b) Estimated ameans

Fig. 11: Experimental results from the second dataset

B. Results from the second dataset

Compared with the first dataset, the second dataset has lower between-class mean difference and higher within-class standard deviation. The classifier accuracies resulting from the experiment using different values of t_m and t_a are depicted in Figure 11. Using all 6144 image pixels as features, we obtain 74.29% classification accuracy by using 1-nearest neighbor classifier and 72.14% classification accuracy by using 3-nearest neighbor classifier.

The highest classification accuracy using the mean difference approach for 1-nearest neighbor classifier is 77.86% with $t_m = 6$ using 3412 features. The highest classification accuracy for 3-nearest neighbor classifier is also 77.86% with $t_m = 8$ using 2802 features.

For the ameans estimation approach, the classification accuracy increases to 80% with $t_a = 0.525$ using 3934 features for 1-nearest neighbor classifier. The 3-nearest neighbor classifier also give the same maximum classification accuracy with $t_a = 0.545$ using 2588 features. Note that since the images in the second dataset are flipped and averaged, the pixels at the left part of the images will be exactly the same as the pixels in the right part. Therefore by using half of the features (1967 for 1-nearest neighbor classifier and 1294 for 3-nearest neighbor classifier) we will have the same classifier performance.

Figure 12 depicts the features used to obtain these max-

(a) $t_m = 6$ (b) $t_m = 8$ (c) $t_a = 0.525$ (d) $t_a = 0.545$

Fig. 12: The image pixels used as features

(a) male (b) male (c) male (d) female (e) female

Fig. 13: Some of the misclassified images by all approaches

(a) male (b) male (c) female (d) female (e) female

Fig. 14: Some of the misclassified images by the approach based on mean difference

imum classification accuracies. The unused pixels are represented by black pixels.

From this dataset there are many images that are misclassified by both approaches. Some of these images are shown in Figure 13. Some of the images that are misclassified by the approach based on mean difference but correctly classified by the approach based on ameans estimation and 1-nearest neighbor classifier are shown in Figure 14. The images that are misclassified by the approach based on ameans estimation but correctly classified by the approach based on mean difference and 1-nearest neighbor classifier are shown in Figure 15.

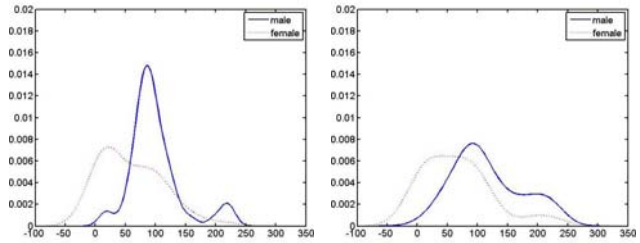
C. Analysis

Eliminating the non-discriminative features could help classifiers perform better. It is desired to have features having high between-class mean difference and low within-class variance.

All of the feature selection approaches consider the presence of black pixels at the bottom of the images as an important information. These black pixels usually represent the presence of long hair in the images. Because of this, it is very difficult to correctly classify the female face images when the individuals

(a) male (b) female (c) female

Fig. 15: The misclassified images by the approach based on ameans estimation



(a) First feature density (b) Second feature density

Fig. 16: Estimated feature densities from the first dataset

have short hair or when the hair is tied. The same problem exists for male face images with dark background.

However, the estimated ameans approach suggests relatively high ameans values on the pixels around the cheek. This make the estimated ameans approach able to correctly classify some of the female face images with short or tied hair. To show this, two estimated feature densities from the first dataset are plotted in Figure 16. The properties of these densities are shown in Table I. As can be seen from Figure 16b, the standard deviation of this density is larger than the one of Figure 16a. Therefore, although the mean difference of the first density is larger (which is supposed to yield a larger feature discriminability), the estimated ameans is smaller. This shows that estimated ameans is capable of giving a better prediction of feature discriminability.

TABLE I: Feature Density Properties

	Density 1	Density 2
Male mean	99.02	116.24
Male stdev	42.91	56.17
Female mean	59.52	66.22
Female stdev	46.69	54.94
Estimated ameans	71.16%	66.50%
Mean difference	39.50	50.02

Both feature selection approaches could help the classifier to have higher accuracies, since the non-discriminative features are eliminated. Compared to the approach based on mean difference, the approach of ameans estimation yields a better classification results on dataset having relatively high variation. This is because the estimation performed in this approach does not rely only on the mean difference, but also based on the data variation. For low variation dataset such as the first dataset the

maximum classification accuracies using both feature selection approaches are similar, since the dataset variation only has a little impact to the ameans estimation.

V. Conclusion

In this paper, we propose a feature selection approach based on ameans estimation. This approach is evaluated in a face-based gender classification system. Our experiments show that the approach based on ameans estimation outperform the approach based on the mean difference in terms of classification accuracy, particularly when the approaches are applied on dataset with relatively high data variance.

References

[1] M. T. Coimbra and J. P. S. Cunha, "MPEG-7 visual descriptor contribution for automated feature extractor in capsule endoscopy," IEEE Trans. on Circuit and Systems for Video Technology, vol. 16, no. 2, pp. 628–637, 2006.

[2] I. K. Timotius, "Predicting Receiver Operating Characteristic Curve, Area Under Curve, and Arithmetic Means of Accuracies based on the Distribution of Data Samples," Proc of IEEE Conf. on Control, Systems, and Industrial Informatics, 2013.

[3] Z. Ji, X. C. Lian, and B. L. Lu, "Gender classification by information fusion of hair and face," in State of the art in face recognition, M. I. Chacon M, Ed., Croatia: I-Tech Education and Publishing KG, 2009.

[4] B. Moghaddam and M. H. Yang, "Learning gender with support faces," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 707–711, 2002.

[5] E. Mäkinen and R. Raisamo, "Evaluation of gender classification methods with automatically detected and aligned faces," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, no. 3, pp. 541–547, 2008.

[6] J. Bekios-Calva, J. M. Buenaposada and L. Baumela, "Revisiting linear discriminant techniques in gender recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 33, no. 4, pp. 858–864, 2011.

[7] I. K. Timotius and S. G. Miaou, "Arithmetic means of accuracies: a classifier performance measurement for imbalanced data set," Proc. of Int. Conf. on Audio, Language, and Image Processing, 2010.

[8] A. W. Bowman and A. Azzalini, Applied Smoothing Techniques for Data Analysis, New York: Oxford University Press, 1997.

[9] I. K. Timotius, I. Setyawan, and A. A. Febrianto, "Face recognition between two person using Kernel Principal Component Analysis and Support Vector Machines," Int. J. Electrical Engineering and Informatics, vol. 2, no. 1, pp. 53-61, 2010.

Session 2

Wireless Communications, Networking and Vehicular Technology

A Fixed Backoff-time Switching Method for Wireless Mesh Networks: Design and Linux Implementation

Sritrusta Sukaridhoto*, Nobuo Funabiki†, Dadet Pramudihanto* and Zainal Arief*
*Electronics Engineering Polytechnic Institute Surabaya, Indonesia
†Graduate School of Natural Science and Technology, Okayama University, Japan
Email: {dhoto, dadet, zar}@eepis-its.edu, funabiki@cne.okayama-u.ac.jp

***Abstract*—As a flexible and cost-efficient scalable Internet access network, we have studied architectures, protocols, and design optimizations of the *Wireless Internet-access Mesh NETwork (WIMNET)*. WIMNET is composed of wirelessly connected access points (APs), where any host can basically access to the Internet through multihop communications between APs with IEEE 802.11 standard. In WIMNET, undesirable situations can often happen such that some links dominate the bandwidth while others become insufficient due to the limited shared bandwidth. However, the contention resolution mechanism using a random backoff-time in the CSMA/CA protocol of 802.11 standards is not sufficient for handling real-time traffic in multihop wireless communications. Previously, we have proposed the concept of the CSMA-based Fixed Backoff time Switching (CSMA-FBS) method for WIMNET to improve the performance by giving necessary link activation chances for multi-hop communications. We implemented our proposal on the QualNet simulator, and verify its effectiveness through simulations. In this paper, we present an implementation of the FBS method in Linux kernel to show its practicality and investigate the performance in a real network. Our design consists of implementations or modifications of the five programs: *Kernel configuration*, *Debugfs*, *Minstrel*, *iw*, and *FBSdaemon*.**

***Keywords*—Wireless mesh network, fixed backoff-time switching, CSMA-FBS, Linux, implementation**

I. Introduction

Recently, a *wireless mesh network* has been extensively studied as a promising network technology for a flexible and cost-efficient solution to expand the communication service area by distributing wireless mesh routers on a network field [1], [2], [3]. The mesh routers are connected with each other through multihop wireless communication links using *IEEE 802.11 standards*, in addition to wireless links between client hosts and routers. Then, as a scalable Internet access network based on this technology, we have studied architectures, protocols, and design optimizations of the *Wireless Internet access Mesh NETwork (WIMNET)* [3]. For a simple architecture, WIMNET is composed of only *access points (APs)* as mesh routers as shown in Figure 1. At least one AP acts as a *GateWay (GW)* to the Internet. Any host in WIMNET can be connected to the Internet through multihop communications between APs and the GW after associated with one neighbor AP.

WIMNET adopts the commonly used *CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) protocol* of the *IEEE802.11 MAC (Media Access Control)* for the shared communication media access by resolving contentions among interfered wireless links [4]. As illustrated in Figure 2, in this protocol, any node holding a transmission packet is on standby for a constant *DIFS* period and a random time called the *backoff-time* before starting the data frame transmission, to avoid frame collisions among contending nodes while providing their fairness. At each transmission chance, a random value within a size called the *Contention Window (CW)* is selected for the backoff-time. When a node fails in the transmission, the CW size is doubled to reduce the probability of the collision occurrence in the retransmission, which is called the *binary exponential backoff*. When the node succeeds in a transmission, it resets the CW size to the initial one.

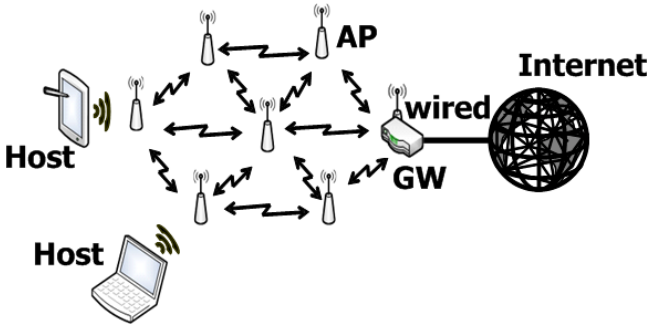


Fig. 1: Outline of WIMNET.

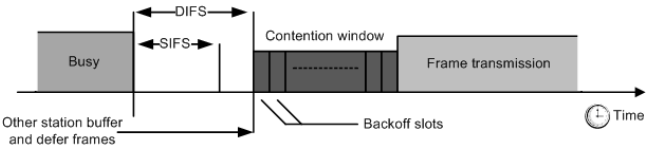


Fig. 2: Timing chart for data frame transmission.

Unfortunately, this conventional CSMA/CA protocol is not sufficient for multihop communications in WIMNET. Firstly, heavy congestions of links around the GW can be bottlenecks of whole communications in WIMNET, because these links have to handle a lot of packets to/from the GW for the Internet access. Thus, they should be activated with much higher priorities than other links. Secondly, interferences among these

congested links may not be resolved by a random backoff-time in the CSMA/CA protocol because of the limited CW size. Here, we note that the initial CW size is small, and even the maximum CW size is limited. Then, multiple conflicting links can be activated simultaneously by generating the same or similar backoff-times at their source nodes. As a result, any link cannot complete the packet transmission successfully, and needs a retransmission that may cause further conflicts. Hence, using the conventional CSMA/CA protocol, WIMNET may cause a lot of packet losses and intolerable delays, which cannot afford real-time applications such as IP-phones and IP-TVs, although their popularity has been increased with advancements of digital communication technologies.

In order to the abovementioned problem, we have proposed the *Fixed Backoff-time Switching (FBS) method* for the CSMA/CA protocol, and shown its implementation on *QualNet* [5], [6], [7], [8], [9], [10]. *QualNet* [11] adopts a more realistic physical model than other network simulators such as *ns-2* [12].

In the FBS method, the three constant parameters, namely the *target link activation rate*, the *active backoff-time*, and the *passive backoff-time*, must be assigned to every link before starting communications. Here, the *link activation* means that the source node of the corresponding wireless link sends out a signal for a frame transmission. The *target link activation rate* represents the rate of activating the corresponding link that is necessary to handle the link traffics properly. The *active backoff-time* represents a shorter waiting time for the link to be activated preferentially when it holds packets for transmissions. The *passive backoff-time* represents a longer waiting time for the link to be activated only if the contending links using the active backoff-time are not activated, where a larger value than any active backoff-time is used. Besides, for any backoff-time of any link, a different value is assigned from each other to avoid simultaneous link activations as best as possible, and the magnitude follows the descending order of expected traffic loads of links so that congested links can be activated more frequently.

During communications, the *actual link activation rate* is observed by counting the numbers of link activation chances and actually activated times for each link, and taking their fraction. If this value is smaller than the target activation rate, the active backoff-time is selected for the preferential activation of the link. Otherwise, the passive backoff-time is selected. Because different values are assigned to them, contentions among interfered links are expected to be resolved.

In this paper, we present an implementation design of the FBS method on a Linux PC to evaluate the performance in real networks. As an open source operating system, Linux has been used as a platform to implement new protocols, methods, and devices for advancements of wireless networks including wireless mesh networks [13]-[16]. Our implementation design consists of implementations or modifications of the five programs: *Kernel configuration*, *Debugfs*, *Minstrel*, *iw*, and *FBSdaemon*.

The rest of this paper is organized as follows: Section II reviews the FBS method. Section III presents our Linux implementation design of the FBS method. Section IV shows our experiment using Linux PC. Section V concludes this paper with some future works.

II. Review of FBS Method

In this section, we briefly review the FBS method for WIMNET.

A. Overview of FBS Method

The FBS method uses the *active backoff-time* and the *passive backoff-time* for each link, and selects either of them as a backoff-time at a frame transmission by comparing the *target link activation rate* and the *actual link activation rate*. Any backoff-time is assigned a different fixed value from each other so that no pair of the conflicting links may be activated simultaneously. Besides, the backoff-time for a link with larger traffic is assigned a smaller value than that for a link with smaller one, so that congested links can be activated preferentially. Furthermore, any active backoff-time is assigned a smaller value than a passive one, so that links using active ones have higher priorities in activations than links using passive ones.

During communications, every time a node holding packets detects that the channel for transmissions becomes free, it updates both the target activation rate and the actual activation rate. If the actual one is smaller than the target one, it selects the active backoff-time to let the link be activated, because the current activation rate of the link is not sufficient to handle its traffic. On the other hand, if it is larger, it selects the passive backoff-time to let other links with active backoff-times be activated with higher priorities. A link with the passive backoff-time can be activated only if any conflicting link with the active backoff-time does not hold packets. The following subsections describe how to calculate the three parameters in the FBS method.

B. Target Link Activation Rate

For a wireless link l_{ij} transmitting packets from AP_i to AP_j for $i = 1, \cdots, N$ and $j = 1, \cdots, N$, the target link activation rate rt_{ij} can be calculated by:

$$rt_{ij} = \frac{tn_{ij}}{an_{ij}} \tag{1}$$

where tn_{ij} represents the target number of activating link l_{ij} per second, and an_{ij} does the average number of link activations per second. tn_{ij} can be given from the requested bit rate by:

$$tn_{ij} = \frac{rb_{ij}}{fb_{ij}} \times (1 + fe_{ij}) \tag{2}$$

where rb_{ij} represents the number of bits per second that link l_{ij} needs to be transmitted, fb_{ij} does the average number of bits in one transmitted frame, and fe_{ij} does the rate of causing the frame transmission error. an_{ij} can be given by:

$$an_{ij} = \frac{1}{ft_{ij}} \tag{3}$$

where ft_{ij} represents the average duration time of one frame transmission.

Among the parameters for the target link activation rate, rb_{ij} should be calculated by taking the summation of the bit rates requested by the applications using link l_{ij} in the routing path of WIMNET. The others, fb_{ij}, fe_{ij}, and ft_{ij}, should be updated during communications by the following equations:

$$fb_{ij} = \frac{sb_{ij}}{sf_{ij}} \tag{4}$$

$$fe_{ij} = \frac{ff_{ij}}{sf_{ij} + ff_{ij}} \tag{5}$$

$$ft_{ij} = \frac{t}{sf_{ij} + ff_{ij} + of_{ij}} \tag{6}$$

where sb_{ij}, sf_{ij}, ff_{ij}, and of_{ij} represent the total number of successfully transmitted bits by link l_{ij}, the total number of successfully transmitted frames, the total number of failed frames, and the total number of transmitted frames of the interfered links with link l_{ij}, when t seconds have passed since the communication started in WIMNET, respectively.

C. Actual Link Activation Rate

The *actual link activation rate* ra_{ij} for link l_{ij} is obtained by dividing the number of successfully transmitted frames with the number of possibly activating chances for the link:

$$ra_{ij} = \frac{sf_{ij}}{ac_{ij}} \tag{7}$$

where ac_{ij} represents the number of possibly activating chances of link l_{ij}.

In the CSMA/CA protocol, ac_{ij} is hard to be obtained. Unlike the TDMA protocol where the link activations are synchronized by a single clock, the timing of counting the number of activating chances is not clear in the CSMA/CA protocol. Besides, the link activation chances resulting in transmission failures must be considered. In this paper, ac_{ij} is counted every time AP_i detects that the channel becomes free.

D. Active/Passive Backoff-time

The *active backoff-time* ta_{ij}^m and the *passive backoff-time* tp_{ij}^m for link l_{ij} are calculated by the following procedure, where m represents the number of consecutively failed transmissions (or retry counter) due to heavy traffics and is saturated by 6. These backoff-times are updated every time the routing path is changed due to the topology change by adding a new AP or removing an existing AP and the host distribution change by the host join or leave to WIMNET. Then, they are fixed during communications.

1) Calculate the number of bits to be transmitted per second rb_{ij} for link l_{ij} by taking the summation of the bit rates for all the communication requests by the hosts using l_{ij}:

$$rb_{ij} = \sum_{k \in H_{ij}} hr_k \tag{8}$$

where H_{ij} represents the set of the host indices using link l_{ij} in the routing path, and hr_k does the requested bit rate (bps) of host k.

2) Sort every link in descending order of rb_{ij}, where the tiebreak is resolved by the number of hosts using this link for the routing path.
3) Set this sorted order to the link priority p_{ij} for l_{ij}.
4) Calculate the active/passive backoff-times for l_{ij}:

$$\begin{aligned} tamin_{ij}^m &= CW_{\min} \cdot \left(2^{m-1} + 2^{m-2} \cdot \tfrac{p_{ij}-1}{P}\right), \\ tamax_{ij}^m &= CW_{\min} \cdot \left(2^{m-1} + 2^{m-2} \cdot \tfrac{p_{ij}}{P}\right), \\ ta_{ij}^m &= rand\left[tamin_{ij}^m, tamax_{ij}^m\right], \end{aligned} \tag{9}$$

where $tamin_{ij}^m$ and $tamax_{ij}^m$ represent the minimum and maximum values for the active backoff-time for l_{ij} when the retry counter is m, $CW_{\min}$ does the initial CW size, and P does the largest priority among the links. In our simulations, $CW_{\min} = 31$ is used in any case.

$$\begin{aligned} tpmin_{ij}^m &= CW_{\min} \cdot \left(2^{m-1} + 2^{m-2} \cdot \tfrac{P+p_{ij}-1}{P}\right), \\ tpmax_{ij}^m &= CW_{\min} \cdot \left(2^{m-1} + 2^{m-2} \cdot \tfrac{P+p_{ij}}{P}\right), \\ tp_{ij}^m &= rand\left[tpmin_{ij}^m, tpmax_{ij}^m\right]. \end{aligned} \tag{10}$$

where $tpmin_{ij}^m$ and $tpmax_{ij}^m$ represent the minimum and maximum values for the passive backoff-time for l_{ij} when the retry counter is m.

III. Design for Linux Implementation of FBS Method

In this section, we present our design for Linux implementation of the FBS method. For convenience, we call a Linux PC implementing the FBS method a *Linux-FBS* in this paper.

A. Overview

Basically, in this design for a *Linux-FBS*, we collect the necessary information from the statistics in the devices, to calculate the fixed back-off time in the FBS method, and assign its calculated value to $AIFS$ for use as the actual backoff-time in the network device with $CW_{min} = CW_{max} = 0$, as shown in Figure 3.

For our implementation of the FBS method in Linux kernel, we have considered implementations or modifications of the following five programs.

- *Kernel configuration* is modified to activate *Debugfs and* Minstrel.
- *Debugfs* is used to obtain the necessary information in the kernel space at the user space through *Minstrel*.
- *Minstrel* is used to obtain the necessary information for the FBS method.
- *iw* is modified to allow the assignment of a specified value (fixed backoff-time) to CW_{min}.
- *FBSdaemon* is newly implemented as a daemon application to calculate the target/active link activation rates and select the fixed back-off time by comparing them as the main procedure of the FBS method.

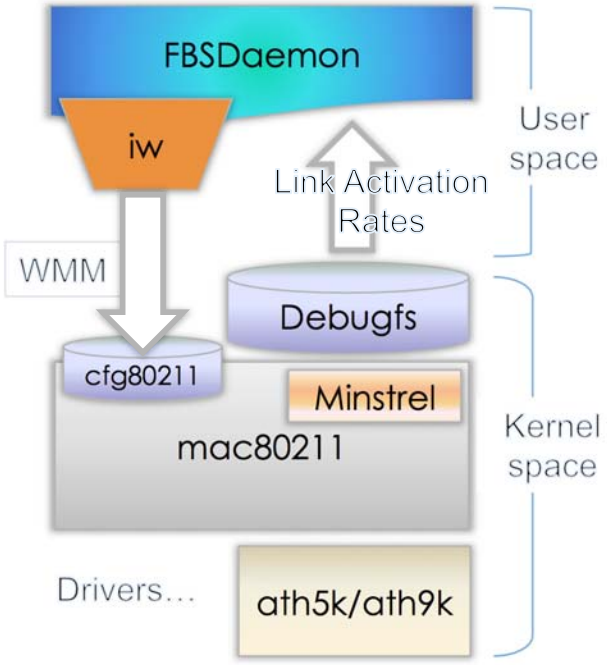


Fig. 3: Data Flow for FBS Method in Linux Implementation.

B. Kernel configuration

For our *Linux-FBS*, we need to activate some features in Linux kernel configurations such as *Debugfs* [17] and *Minstrel* [18] that are used for wireless networks. Therefore, we set up the configuration of the Linux kernel as follows:

```
CONFIG_DEBUG_FS=y
CONFIG_DEBUG_KERNEL=y
CONFIG_WIRELESS=y
CONFIG_CFG8011=m
CONFIG_CFG80211_DEBUGFS=y
CONFIG_LIB80211=m
CONFIG_LIB80211_DEBUG=y
CONFIG_MAC80211=m
CONFIG_MAC80211_RC_MINSTREL=y
CONFIG_MAC80211_RC_MINSTREL_HT=y
CONFIG_MAC80211_RC_DEFAULT_MINSTREL=y
CONFIG_MAC80211_RC_DEFAULT="minstrel_ht"
CONFIG_MAC80211_DEBUGFS=y
```

For the wireless drivers for our implementation, we set up the configuration of the Linux kernel as follows:

```
CONFIG_ATH_COMMON=m
CONFIG_ATH_DEBUG=y
CONFIG_ATH5K_DEBUG=y
CONFIG_ATH9K_DEBUGFS=y
CONFIG_ATH9K_HTC_DEBUGFS=y
```

C. Debugfs

Debugfs is a special file system available in a Linux kernel. It is technically referred as a *kernel space-user-space interface*, and is a simple RAM-based file system that is designed for debugging the kernel. *Debugfs* allows a kernel developer to make information in the kernel space available in the user space. To compile a Linux kernel with *Debugfs*, we need to set *CONFIG_DEBUG_FS* option *yes*. Then, we need to mount *Debugfs* with the following command:

```
mount -t debugfs none /sys/kernel/debug
```

D. Minstrel

Minstrel is a *mac80211* rate control algorithm ported over from *MadWifi* that supports multiple rate retries. *Minstrel* has been claimed to be one of the best rate control algorithms. *Minstrel* provides the success/failure information, the actual data rate communication, and the status of interface.

After mounting *Debugfs*, we can use *Minstrel* from a subdirectory of *Debugfs*. Inside the directory of */sys/kernel/debug/ieee80211/phy0/netdev:wlan0/stations*, subdirectories exist where each subdirectory corresponds to each wireless node in the network such as a host (client PC) that is associated with the *Linux-FBS*. The name of a subdirectory is the MAC address of the associated node. For example, */sys/kernel/debug/ieee80211/phy0/netdev:wlan0/stations/ 00:22:cf:72:21:22/* represents a subdirectory corresponding to a node whose *mac address* is *00:22:cf:72:21:22*. Inside of this subdirectory, we can find the files of the minstrel information for this node.

For the FBS method, we use the following files from *Minstrel*: *rc_stats*, *tx_bytes*, *tx_packets*, *tx_retry_count*, and *tx_retry_failed*.

From the *tx_packets* file, we can get the value for sf_{ij} (the total number of successfully transmitted frames of link l_{ij}). From the *rc_stats* file, we can get the value for ac_{ij} (the number of possibly activating chances) from the *attemp* value. Then, we can calculate the value for ra_{ij} (actual link activation rate). Also, from this file, we can get the value for rb_{ij} (the number of bits to be transmitted per second for link l_{ij}) from the *throughput* value.

Then, we can get the value for sb_{ij} (the total number of successfully transmitted bits by link l_{ij}) from the *tx_bytes* file, the value for of_{ij} (the total number of transmitted frames of the interfered links with link l_{ij}) from the *tx_retry_count* file, and the value for ff_{ij} (the total number of failed frames) from the *tx_retry_failed* file, respectively. Then, we can calculate the value for ta_{ij} (target link activation rate).

1) Modification of iw*: iw*[19] is a new *nl80211* based *CLI (Command Line Interface)* configuration utility for wireless devices. *nl80211* is a new IEEE 802.11 netlink interface public header. *iw* supports most of the new drivers that have been recently added to the Linux kernel. In our *Linux-FBS* implementation, we use *iw* to assign the fixed backoff-time in the FBS method by changing the values of the variables for *Wi-Fi Multimedia (WMM)* in *IEEE802.11e*, namely CW_{min}, CW_{max}, $AIFS$, and $TXOP$.

However, a default application of *iw* cannot access to or modify the values for them. Thus, we modified the source code of *iw* so that it is possible. In this source code modification, we use a function in the *hostapd* application so that we can change the values for CW_{min}, CW_{max}, $AIFS$, and $TXOP$. Actually, we add the *handle_txq_params* function in the *phy.c* file to access to TXQ_PARAMS in wireless Linux kernel parameters.

Using the *iw phy0 set txq_params 0 0 0 0 10* command, we set the values of the *WMM* variables for *class 0* (Best Effort), such that $CW_{min} = CW_{max} = TXOP = 0$, and $AIFS = 10$ if the selected fixed backoff-time in the FBS method is 10 for this link.

Listing 1 shows our modification of the source code for *iw*.

```
static int handled_txq_params(struct nl80211_state *
    state, struct nl_cb *cb, struct nl_msg *msg, int
     argc, char **argv)
{
    __u8 queue, aifs;
    __u16 cwmin,cwmax,txop;

    struct nlattr *txq;

    //Sanity checking
    ...

    queue = strtoul(argv[0], NULL, 10);
    cwmin = strtoul(argv[1], NULL, 10);
    cwmax = strtoul(argv[2], NULL, 10);
    txop = strtoul(argv[3], NULL, 10);
    aifs = strtoul(argv[4], NULL, 10);

    printf("Set TXQ_PARAMS for class[%d] : cwmin=%d
        cwmax=%d txop=%d aifs=%d\n", queue, cwmin,
        cwmax, txop, aifs);

    //Range checking for the access class param
    ...

    txq = nla_nest_start(msg,
                NL80211_ATTR_WIPHY_TXQ_PARAMS);
    if (!txq)
        return -ENOBUFS;

        struct nlattr *tx = nla_nest_start(msg,queue
            );

        NLA_PUT_U8(msg, NL80211_TXQ_ATTR_QUEUE,
            queue);
        NLA_PUT_U16(msg, NL80211_TXQ_ATTR_CWMIN,
            cwmin);
        NLA_PUT_U16(msg, NL80211_TXQ_ATTR_CWMAX,
            cwmax);
        NLA_PUT_U16(msg, NL80211_TXQ_ATTR_TXOP, txop
            );
        NLA_PUT_U8(msg, NL80211_TXQ_ATTR_AIFS, aifs)
            ;

        nla_nest_end(msg,tx);

    nla_nest_end(msg, txq);

   return 0;

nla_put_failure:
    return -ENOBUFS;
}
COMMAND(set, txq_params, "<access_class> <cwmin> <
    cwmax> <txop> <aifs>", NL80211_CMD_SET_WIPHY, 0,
     CIB_PHY, handle_txq_params, "Set TXQ_PARAMS
    with Queue, CWmin, CWmax, TXOP, AIFS\n");
```

Listing 1: "IW modification in phy.c"

E. FBSdaemon

We implement the procedure for the FBS method by generating a *daemon* application using *Perl*. In this paper, we call this application *FBSdaemon*.

The main cycle for the backoff-time control for the FBS method in *FBSdaemon* consists of the four steps: 1) reading the necessary files from *Minstrel*, 2) calculating both the target and active link activation rates, 3) selecting the fixed back-off time by comparing the both rates, and 4) assigning the selected fixed back-off time to *AIFS* by using the syntax *system* and calling the modified *iw* application. Besides, *FBSdaemon* can give a log report, and can run in the background.

Algorithm 1 shows this procedure in *FBSdaemon*.

input : Minstrel files: *rc_stats*, *tx_bytes*, *tx_packets*, *tx_retry_count*, *tx_retry_failed*
BO file
output: AIFS

Perl initialization for Daemon, Log, Files;
begin
  Log start;
  Daemonize;
  for *(;;)* **do**
    $ac_{ij} \leftarrow$ read(*rc_stats*, attemp);
    $sf_{ij} \leftarrow$ read(*rc_stats*, success);
    $sb_{ij} \leftarrow$ read(*tx_bytes*);
    $sf_{ij} \leftarrow$ read(*tx_packets*);
    $of_{ij} \leftarrow$ read(*tx_retry_count*);
    $ff_{ij} \leftarrow$ read(*tx_retry_failed*);
    active $\leftarrow$ read(*bofile*, active);
    passive $\leftarrow$ read(*bofile*, passive);
    Calculate ra_{ij}, ta_{ij};
    if $ra_{ij} < ta_{ij}$ **then**
      AIFS $\leftarrow$ *active*
    else
      AIFS $\leftarrow$ *passive*
    end
    system("*iw* phy0 set *txq_params* 0 0 0 0 *AIFS*");
    wait(300s);
  end
end

Algorithm 1: FBS Daemon

IV. Experiment using Linux PC

In this section, we show implementation result using Linux PC to evaluate the CSMA-FBS protocol.

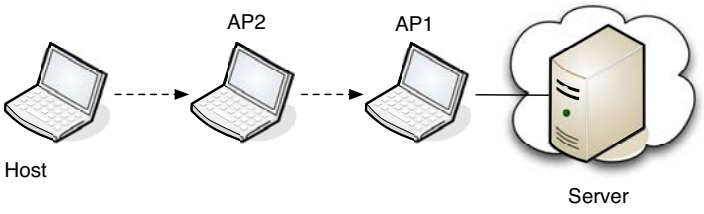


Fig. 4: Topology of Experiment.

Three Linux PCs with adhoc mode connected as shown in Figure 2. We sent 2Mb TCP packets data by using *iperf*[20]

program from host to server to make network fully loaded, we monitor the throughput performance by using *nload*[21] application. The simulation environment is summarized in Table I.

TABLE I: Simulation environment.

Parameter	Value
Proc	Intel i5
Interface	Atheros AR9285(bgn)
OS	Debian GNU/Linux
Kernel	2.6.39 Custom Debugfs and minstrel
User space	iw 0.19 (custom)
Applications	perl 5.10 Iperf, nload

We conduct the experiment for 50 times and compare the throughput between two protocols. The average throughput result for CSMA protocol is 17.33 Mbps and for CSMA-FBS protocol is 22.23 Mbps. This results indicates that CSMA-FBS throughput is better than CSMA protocol. From the simulation[8], CSMA-FBS shows around 40% better rather than CSMA protocol but in Linux implementation, the effectiveness is only around 28%. This is because in Linux implementation the range of topology is different.

V. Conclusion

In this paper, we presented a Linux implementation for the *Fixed Backoff-time Switching (FBS) method* for the CSMA/CA protocol in the *Wireless Internet-access Mesh Network (WIMNET)*. Our implementation consists of implementations or modifications of the five programs: *Kernel configuration*, *Debugfs*, *Minstrel*, *iw*, and *FBSdaemon*. In this Linux implementation, our CSMA-FBS gives better performance rather then CSMA protocol. In our future works, we will refine the codes of the implementation, generate a testbed with multiple *Linux-APs* implementing the FBS method, and investigate the performance of our method in a real network topology.

Acknowledgment

This work is partially supported by KAKENHI (22500059).

References

[1] I. F. Akyildiz, X. Wang, and W. Wang, Wireless mesh networks: a survey, Comput. Netw. ISDN Syst., vol. 47, no. 4, pp. 445-487, March 2005.

[2] Y. Zhang, J. Luo, and H. Hu, Wireless mesh networking: architectures, protocols and standards, Auerbach Pub., 2006.

[3] N. Funabiki edited, Wireless mesh networks, InTech - Open Access Pub., 2011, http://www.intechopen.com/books/show/title/wireless-mesh-networks.

[4] Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, IEEE Std. 802.11, 1999.

[5] S. Tajima, N. Funabiki, and T. Higashino, "A proposal of fixed backoff-time switching method by link activation rate for wireless mesh networks," Proc. Int. Conf. Complex, Intell., Software Intensive Sys. (CISIS), pp. 647-652, 2011.

[6] N. Funabiki, S. Sukaridhoto, Z. Wang, T. Nakanishi, K. Watanabe, and S. Tajima, "An implementation of fixed backoff-time switching method on IEEE 802.11 MAC protocol for wireless Internet-access mesh network," Proc. Int. Work. Smart Info-Media Sys. Asia (SISA 2011), pp. 67-72, Oct. 2011.

[7] S. Sukaridhoto, N. Funabiki, T. Nakanishi, and K. Watanabe, "A proposal of CSMA fixed backoff-time switching protocol and its Implementation on QualNet simulator for wireless mesh networks", Proc. WAINA, March 2012.

[8] S. Sukaridhoto, N. Funabiki, T. Nakanishi, K. Watanabe and S. Tajima, "A Fixed Backoff-Time Switching Method for CSMA/CA Protocol in Wireless Mesh Networks", IEICE TRANSACTIOINS on Communications, vol. 96, no. 4, pp. 1019-1029, Apr. 2013.

[9] S. Sukaridhoto, N. Funabiki, T. Nakanishi, K. Watanabe and S. Tajima, "A Linux Implementation Design of Fixed Backoff-time Switching Method for Wireless Mesh Networks", IEICE Tech. Rep., vol. 112, NS2012-67, pp.83-88, Sept. 2012.

[10] S. Sukaridhoto, N. Funabiki, T. Nakanishi, K. Watanabe and S. Tajima, "An Idea of Linux Implementation of Fixed Backoff-time Switching Method for Wireless Mesh Networks", Proceedings of the Society Conference of IEICE, 2012.

[11] QualNet simulator, Scalable network tech., http://www.scalable-networks.com.

[12] G. A. D. Caro, Analysis of simulation environments for mobile ad hoc networks, Tech. Rep., no. IDSIA-24-03, Dec. 2003.

[13] M. Vipin and S. Srikanth, Analysis of open source drivers for IEEE 802.11 WLANs, Proc. ICWCSC 2010, 2010.

[14] K. Chebrolu and B. Raman, FRACTEL: a fresh perspective on (rural) mesh networks, Proc. ACM SIGCOMM NSDR, Aug. 2007.

[15] A. Sharma and E. M. Belding, FreeMAC: implementing a multi-channel TDMA MAC on 802.11 hardware, http://moment.cs.ucsb.edu/~asharma/freemac-mobisys.pdf.

[16] P. Djukic and P. Mohapatra, Soft-TDMAC: a software TDMA-based MAC over commodity 802.11 hardware, Proc. INFOCOM, April 2009.

[17] Debug - Linux Wireless http://linuxwireless.org/en/users/Drivers/ath9k/debug.

[18] Minstrel - Linux Wireless, http://linuxwireless.org/en/developers/Documentation/mac80211/RateControl/minstrel/.

[19] iw - Linux Wireless http://wireless.kernel.org/en/users/Documentation/iw.

[20] iperf - TCP and UDP bandwidth performance measurement tool https://code.google.com/p/iperf/.

[21] nload - a console application which monitors network traffic and bandwidth usage in realtime. http://sourceforge.net/projects/nload/.

A Game-Theoretic Approach for Dynamic Spectrum Sharing in Cognitive Radio Networks

Komang Wahyu Trisna, I Wayan Mustika, Widyawan, and Selo Sulistyo
Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada
Jalan Grafika No. 2, Yogyakarta, 55281 Indonesia
Email: trisna_s2te12@mail.ugm.ac.id, {wmustika, widyawan, selo}@ugm.ac.id

***Abstract*—In a spectrum sharing system, lower-priority users are allowed to spatially reuse the spectrum allocated to higher-priority users as long as they do not disrupt communications of the latter. Therefore, to improve spectrum utilization, an important requirement for the former users is to manage the interference and ensure that the latter users can maintain reliable communications. This paper presents a game theoretic framework to model the dynamic spectrum sharing in cognitive radio networks. First, a utility function that captures the selfish and cooperative behavior of the lower-priority users to manage the interference by selecting the best channel with minimal intra- and inter-system interference is defined. Next, based on the defined utility function, the proposed framework can be formulated as a potential game; thus, the convergence to a Nash equilibrium point is ensured as long as the best response dynamic is adopted. At the equilibrium point, power allocation algorithm is proposed such that the interference to higher-priority users can be maintained below the maximum allowable level. The simulation results show the convergence of the proposed potential game and the performance improvement of higher-priority users in terms of SINR and outage probability.**

I. Introduction

The dramatic development of wireless communication system have remarkably increases the demand for wireless spectrum. With the static spectrum allocation policy, a part of the spectrum band is permanently assigned to the licensed wireless systems. This situation may lead to serious spectrum scarcity for new emerging wireless technologies and services. In contrast, it is evident that the static spectrum usages are far from efficient due to under-utilized the spectrum both in spatial and temporal way [1]. Therefore, a flexible spectrum usage through a dynamic spectrum sharing becomes a promising approach to achieve efficiency enhancement of spectrum usage. In the dynamic spectrum sharing, the unlicensed wireless users (lower-priority users) dynamically access the spectrum of licensed wireless users (higher-priority users) on an opportunistic or a negotiated basis [2], [3].

Cognitive radio along with software defined radio is the key enabling technology that can facilitate the dynamic spectrum sharing by providing wireless devices the ability to sense the spectral environment over a wide spectrum band and autonomously adapt their transmission parameters according to the changing environment [4]. Thus, intelligent decision making algorithms are necessary for cognitive radios in order to achieve the desired objective, e.g. maximize the spectrum utilization. In a cognitive radio network where a higher- and lower-priority system coexist, the dynamic spectrum sharing can be realized. The former system consists of higher-priority users or primary users (PUs), which are allocated with licensed radio spectrum. The utilization of such spectrum could be enhanced by allowing the lower-priority users or secondary users (SUs), which have cognitive radio capabilities, to access the spectrum [5]. However, it is a challenging task in constructing the dynamic spectrum sharing due to the requirement of enabling "peaceful" co-existence of both primary and secondary users, in which SUs are allowed to spatially reuse the spectrum as long as they can avoid harmful interference to the PUs. In other words, the PUs can operate effectively as long as the received signal-to-interference and noise power ratio (SINR) exceeds an established threshold or the interference power is less than the maximum allowable level.

Game theory provides mathematical tools for modeling and analyzing the situation where autonomous decision makers have conflicting interests with each other. Using the game-theoretic approach, the outcome resulted from the interaction among nodes with conflicting interests can be predicted through equilibrium analysis [6]. Since cognitive radios can make intelligent decisions based on their interaction through radio environment, there has been an increasing interest in analyzing their behavior from a game theoretic perspective. Recently, game theory has been extensively applied to problems in wireless communication, particularly in areas related to distributed resource allocation, e.g., power control, resource management, medium access control and topology control [7]–[12]. In [9], the authors proposed a game theoretic approach for adaptive channel allocation in cognitive radio networks. In the network, N transmitter-receiver pairs of SUs compete for a single channel among C available channels for transmission. However, since the PUs were not taken into consideration, and therefore, they were not explicitly protected from excessive interference due to spectrum access of the SUs. In addition, Mustika *et al.* proposed a game-theoretic approach for spectrum sharing in cognitive radio networks [10]. In such a scheme, SUs select the appropriate combination of channel and transmit power level while aware of the interference generated to the PUs. However, about 10% of PUs performance is still below the required SINR target due to severe interference from nearby SUs.

In the present paper, we proposed a potential game approach

for channel allocation combined with power control to enable a dynamic spectrum sharing in cognitive radio networks. We first propose a utility function for the game learning that captures the selfish and cooperative behavior of each SU in controlling the interference to PUs and other SUs. We then formulate a potential function for the proposed game, which can be seen as the global objective of the game. Thus, finding a (local) optimal solution to the proposed potential function is equivalent to finding a solution of the game, called Nash equilibrium (NE). Interestingly, when the best strategy of a player to the other players' strategy is applied, the improvement path of each player converges to a pure strategy NE [13]. While our previous work uses the combination of channel and transmit power level as a strategy during the game playing [10], this work consider the channel as a strategy in evaluating the utility function to achieve a NE. In addition, we then propose a power allocation algorithm for SUs at the NE to ensure that the SINR performance at PUs are maintained above a certain threshold.

The rest of this paper is structured as follows. In Section II, the system model is described. Section III presents the game-theoretic framework and power allocation algorithm for dynamic spectrum sharing. The simulation results are shown in Section IV. Finally, Section V summarizes the conclusions of the paper.

II. System Model

The cognitive radio network under consideration consists of a set of N transmitting-receiving pairs of SUs and M transmitting-receiving pairs of PUs, uniformly distributed in a two dimensional square region. The network topology is assumed to be remain fixed over the time required to establish strategies. We assume that the number of frequency bands or channel is equal to the number of PU pairs ($C = M$) where one channel c_h^{H} from the set of available channels $\mathcal{C}$ is assigned to one PU transmitter $h \in \mathcal{M} = \{1, \ldots, M\}$. Fig. 1 shows the system model of spectrum sharing in cognitive radio network. In the system model, PU-TX, PU-RX, SU-TX, and SU-RX indicate the PU transmitter, PU receiver, SU transmitter, and SU receiver, respectively. An SU transmitter $i \in \mathcal{N} = \{1, \ldots, N\}$ senses the available channels and selects a channel c_i from the set of available channels $\mathcal{C}$ to communicate with the intended receiver. The spectrum underlay approach is assumed where a channel can be used simultaneously by a PU and SUs. Thus, the transmission of SUs appears as noise to the PU. By distributively selecting transmission channel, SUs attempt to manage intra- and inter-system interference. As the goal of dynamic spectrum sharing, the inter-system interference from SUs to PU must not exceed a certain limit.

Let p_h^{H} denote the transmit power of PU transmitter h, so that the received SINR at the PU receiver h can be expressed as

$$\gamma_h^{\mathrm{H}} = \frac{G_{hh}^{\mathrm{HH}} p_h^{\mathrm{H}}}{\sum_{i=1}^{N} G_{hi}^{\mathrm{HL}} p_i \delta_{c_h^{\mathrm{H}} c_i} + n_0}, \quad (1)$$

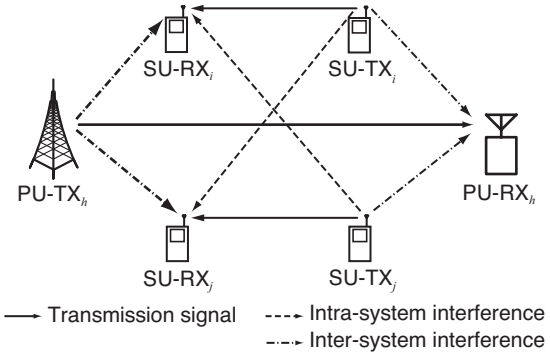


Fig. 1. System model of spectrum sharing with intra- and inter-system interference scenario.

where G_{hh}^{HH} is the link gain between the PU transmitter h and PU receiver h, G_{hi}^{HL} is the link gain between the SU transmitter i and PU receiver h, and n_0 denotes the noise power at each receiver.

Due to simultaneous transmission among users in both the higher- and the lower-priority system, the SINR measured at the SU receiver i can be expressed as

$$\gamma_i = \frac{G_{ii} p_i}{\sum_{j=1, j\neq i}^{N} G_{ij} p_j \delta_{c_i c_j} + \sum_{h=1}^{M} G_{ih}^{\mathrm{LH}} p_h^{\mathrm{H}} \delta_{c_i c_h^{\mathrm{H}}} + n_0}, \quad (2)$$

where G_{ih}^{LH} is the link gain between the PU transmitter h and SU receiver i, G_{ij} is the link gain between the SU transmitter j and SU receiver i, and p_i denotes the transmit power of the SU transmitter i. $\delta_{c_i c_j}$ is the Kronecker delta function that indicates whether or not channel c_i and channel c_j are the same:

$$\delta_{c_i c_j} = \begin{cases} 1 & c_i = c_j, \\ 0 & c_i \neq c_j. \end{cases} \quad (3)$$

Outage probability is used in measuring the performance of PUs, which is defined as the probability that the received SINR at PU receiver h falls below a certain threshold, $\gamma_h^{\mathrm{H}} < \gamma_h^{\mathrm{target}}$.

III. Game Formulation and Power Allocation Algorithm

A. Game-theoretic Framework

Since SUs in the spectrum sharing model compete for the limited spectrum resources, their interaction can be modeled as a strategic or normal form game. In particular, the considered game can be mathematically defined as $\Gamma = \left\{\mathcal{N}, \{\mathcal{S}_i\}_{i\in\mathcal{N}}, \{u_i\}_{i\in\mathcal{N}}\right\}$, where:

i) $\mathcal{N}$ is the finite set of players, i.e., SU pairs.

ii) $\mathcal{S}_i$ is the set of strategies associated with player i, i.e., channel that could be selected by player i, denoted as $s_i = c_i \in \mathcal{S}_i$. The strategy space of all players is $\mathcal{S} = \prod_{i\in\mathcal{N}} \mathcal{S}_i$.

iii) $u_i : \mathcal{S} \rightarrow \mathbb{R}$ is the set of utility functions that the players associate with their strategies. The utility of player i represents the reward received by player i as the function of the strategy chosen by player i, s_i, and the strategies selected by other players different from i, $\boldsymbol{s}_{-i}$. In other

words, $u_i = \{u_1, u_2, \ldots, u_n\}$ wishes to be maximized by each player by selfishly selecting the best strategy, considering the current strategies of the other players.

In the proposed non-cooperative spectrum sharing game, each SU only considers about her own benefit by selecting her best strategy that maximize her utility. The situation where all players are assumed to be rational and adopt selfish behavior produces an outcome that could be predicted. The most probable outcome of the non-cooperative game is termed as Nash equilibrium (NE). Players are in an equilibrium point if a change in strategies by any one of them would lead that player gain less than if the player remained with her current strategy [14]. A strategy profile for all players, denoted as $\boldsymbol{s}^\star = (s_i^\star, \boldsymbol{s}_{-i}^\star)$, is a NE if and only if satisfies the following condition

$$u_i(s_i^\star, \boldsymbol{s}_{-i}^\star) \geq u_i(s_i', \boldsymbol{s}_{-i}^\star), \quad \forall s_i' \in \mathcal{S}_i, \ \forall i \in \mathcal{N}. \tag{4}$$

NE corresponds to mutual best response strategies for all players. The best response strategy is a dynamic process of updating strategies. In this strategy update, a player attempts to choose a strategy that maximizes her respective utility, given the current strategy of other players [13]. Best response strategy of the player i to the strategy profile $\boldsymbol{s}_{-i}$ at time $t+1$, $s_i^{t+1}(\boldsymbol{s}_{-i})$, is a strategy that satisfies

$$s_i^{t+1} \in \arg\max_{s_i' \in \mathcal{S}_i} u_i(s_i', \boldsymbol{s}_{-i}^t), \tag{5}$$

where $(s_i^t, \boldsymbol{s}_{-i}^t) \in \mathcal{S}$ denotes the action profile at time t.

B. Potential Game Formulation and Utility Function

The existence of a potential function that reflects the incentives of all players due to any unilateral change in the utility of a single player is the characteristic of potential game [15]. The potential function $V : \mathcal{S} \to \mathbb{R}$ has the following property

$$V(s_i', \boldsymbol{s}_{-i}) - V(s_i, \boldsymbol{s}_{-i}) = u_i(s_i', \boldsymbol{s}_{-i}) - u_i(s_i, \boldsymbol{s}_{-i}), \tag{6}$$

where $s_i' \in \mathcal{S}_i$.

Potential game is a special class of the normal-form game that guarantees the convergence to a NE when the best response strategy is performed. In general, the best response strategy of any player does not always converge to a NE. However, if the existence of potential function can be proved in a game, the best response strategy of all players will terminate in a finite number of steps to a NE, regardless of the initial condition of the game and the order of the sequential play [13].

By extending the utility function proposed in [9], we propose a utility function that captures the selfish and cooperative behavior of each SU to minimize the intra- and inter-system interference:

$$\begin{aligned} u_i(s_i, \boldsymbol{s}_{-i}) = &- \sum_{j=1, j\neq i}^{N} G_{ji} p_i \delta_{c_j c_i} - \sum_{j=1, j\neq i}^{N} G_{ij} p_j \delta_{c_i c_j} \\ &- \sum_{h=1}^{M} G_{hi}^{\mathrm{HL}} p_i \delta_{c_h^{\mathrm{H}} c_i} - \sum_{h=1}^{M} G_{ih}^{\mathrm{LH}} p_h^{\mathrm{H}} \delta_{c_i c_h^{\mathrm{H}}}, \end{aligned} \tag{7}$$

The first and second terms in the proposed utility function represent the interference created by SU transmitter i to other SU receivers and the total interference power received by SU receiver i from other SU transmitters, respectively. While the remaining terms represent the interference created to PU receivers and the interference received by SU receiver i from PU receivers, respectively. The existence of the first and third terms in the proposed utility function will enforce the cooperation among SUs where each SU competes for the channel that is not only receives the minimum interference from other users but also that causes minimum possible interference to other users.

In order to enable SUs in measuring the potential interferers in the network, we assume that each SU has complete channel information, which can be obtained by exchanging control message, probing packets, and channel measurement through common control channel [2].

Given the proposed spectrum sharing game with utility function u, the potential function can be formulated as

$$\begin{aligned} &V(s_i, \boldsymbol{s}_{-i}) \\ &= \sum_{i=1}^{N} \left(-\frac{1}{2} \sum_{j=1, j\neq i}^{N} G_{ji} p_i \delta_{c_j c_i} - \frac{1}{2} \sum_{j=1, j\neq i}^{N} G_{ij} p_j \delta_{c_i c_j} \right) \\ &\quad + \sum_{i=1}^{N} \left(- \sum_{h=1}^{M} G_{hi}^{\mathrm{HL}} p_i \delta_{c_h^{\mathrm{H}} c_i} - \sum_{h=1}^{M} G_{ih}^{\mathrm{LH}} p_h^{\mathrm{H}} \delta_{c_i c_h^{\mathrm{H}}} \right). \end{aligned} \tag{8}$$

The potential function can be seen as the network utility in a game, where an increase in the utility of individual user contributes to the increase of the overall network utility.

C. Power Allocation Algorithm

As the the best response strategy of all SUs in the proposed potential game converges to a NE, a certain performance of the network can be observed. However, the performance of PUs depends on two factors such as number of SUs in selecting the same channel and transmit power of SU transmitter. In a spectrum sharing system, the SUs are allowed to spatially reused the spectrum band as long as they do not cause excessive interference to the PUs. Thus, each SU determines it transmit power via power control to avoid such harmful interference. The problem is equivalent to find the power vector $\boldsymbol{p}_i = \{p_1, p_2, \ldots, p_N\}$ of all SUs so as the total interference power of SUs that accessing the spectrum of a PU among all PUs is less than the maximum allowable level.

At convergence state, we consider that each channel is selected by a group of SUs such as the group transmits on the same channel. Thus, a SU causes and receives intra- and inter-system interference to and from other SUs and PUs in a group. Let $\boldsymbol{k}$ denote C-dimensional vector representing group of SUs in selecting a particular channel, represented as:

$$\boldsymbol{k} = \{k_1, k_2, \ldots, k_C\}, \tag{9}$$

where the element of $\boldsymbol{k}$ is the number of SUs that transmit on the same channel c.

In order to maintain the SINR performance of PUs exceeds an established threshold, the transmission power of SUs have to be adjusted. Let $I_{c_h^{\mathrm{H}}}^{\mathrm{target}}$ denote the interference target at PU receiver using channel c_h^{H}, which can be calculated as

$$I_{c_h^{\mathrm{H}}}^{\mathrm{target}} = \left(\frac{G_{hh}^{\mathrm{HH}} p_h^{\mathrm{H}}}{\gamma_h^{\mathrm{target}}} \right) - n_0. \quad (10)$$

In practice, the transmit power of SU transmitter i, denoted as p_i, is constrained by

$$0 \leq p_i \leq p_{\max}, \quad (11)$$

where $p_{\max}$ denotes the maximum transmit power at SU transmitter.

The total interference power of group of SUs that utilizes the same channel have to be maintained not exceed the interference target at PU receiver h that operates on channel c_h^{H}, which can be expressed as

$$\sum_{i=1}^{N} G_{hi}^{\mathrm{HL}} p_i \delta_{c_h^{\mathrm{H}} c_i} \leq I_{c_h^{\mathrm{H}}}^{\mathrm{target}}. \quad (12)$$

Under assumption that each SU in a group contributes to equal amount of interference power to PU receiver h that operates on the same channel ($c_i = c_h^{\mathrm{H}}$), the transmit power of the SU transmitter i can be calculated as

$$p_i = I_{c_h^{\mathrm{H}}}^{\mathrm{target}} / \left(k_{c_i} \cdot G_{hi}^{\mathrm{HL}} \right). \quad (13)$$

Above equation provides the power allocation of SU transmitter i to ensure that the performance of PU in the given channel is not deteriorated by severe inter-system interference.

IV. Simulation Results

In this section, we study the convergence property and performance evaluation of the proposed spectrum sharing game. The network consists of 30 pairs of SUs that are randomly distributed in a square region of dimension $400\,\mathrm{m} \times 400\,\mathrm{m}$ as shown in Fig. 2. We assume 4 pairs of PUs and 4 channels with equal bandwidth. Under this scenario, each frequency channel is assigned to each pair of PUs so that the intra-system interference among PUs is avoided.

TABLE I
Simulation Parameters

Parameter	Value
Simulation area	$400\,\mathrm{m} \times 400\,\mathrm{m}$
Number of SUs	30 pairs
Number of PUs	4 pairs
Number of available channels	4
Transmit power at PU-TX	25 mW
Max transmit power at SU-TX	5 mW
Noise power	10^{-13} W
Path loss model	Free space
Path loss exponent	2
SINR threshold	20 dB
Traffic model	Full buffer
Number of topologies	5000

At the initial assignment, random channel assignment is performed at each SU transmitter with maximum transmit

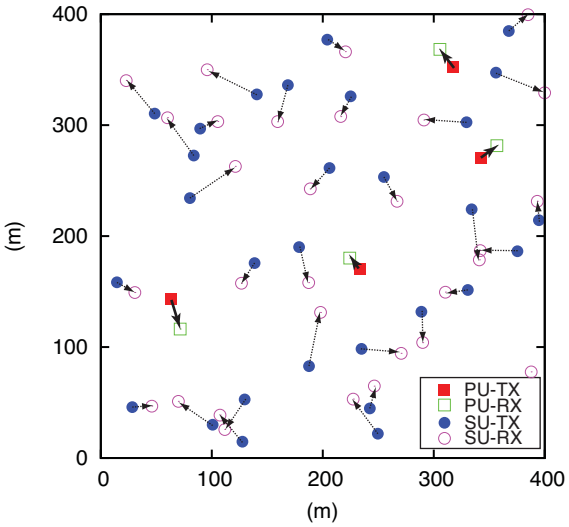


Fig. 2. Example of distribution of SUs and PUs in a network.

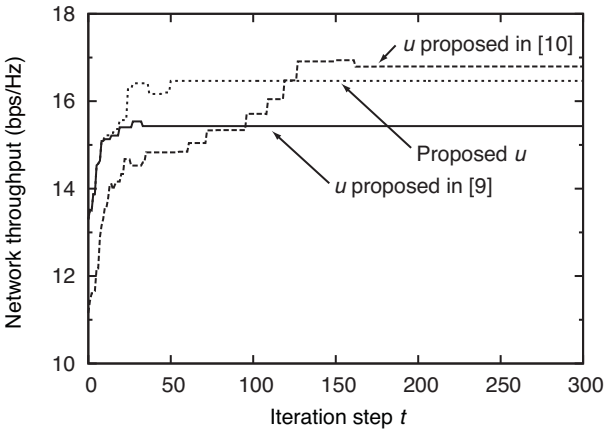


Fig. 3. Convergence of network throughput of SUs.

power $p_{\max} = 5\,\mathrm{mW}$. Free space path loss model with path loss exponent 2 is used in calculating the channel gain. For the sake of simplicity, the effect of fast fading and shadowing is not considered in the simulation. The SINR threshold at each PU receiver, $\gamma_h^{\mathrm{target}}$, is set to 20 dB. We assume that all PU receivers have identical SINR threshold. Simulation is carried out for 5000 different topologies, where each topology consists of a random distribution of PUs and SUs in the network. The relevant parameters used in the simulation is summarized in Table I.

We first study the convergence property of the proposed spectrum sharing game. The evolution of the network throughput of the SUs during iteration step t is shown in Fig. 3, where term u represents the utility function. From Fig. 3, it can be seen that the initial state occurs at $t = 0$ where each SU transmitter selects a single channel from the available channels in random fashion for transmission. At this point, the performance of network throughput is the lowest due to unmanaged intra- and inter-system interference by SUs. As the iteration step t moves, sequential play is performed

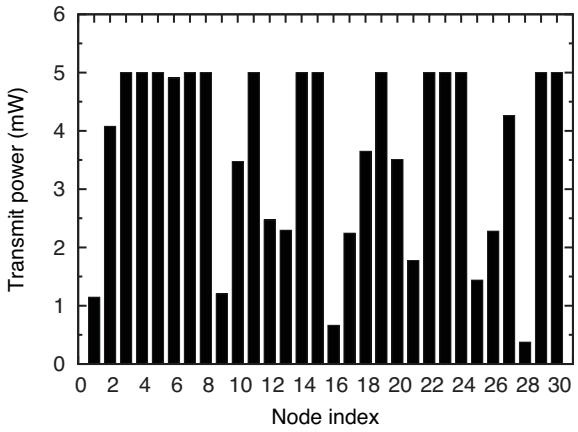


Fig. 4. Transmit power distribution of SUs at equilibrium point using power allocation algorithm.

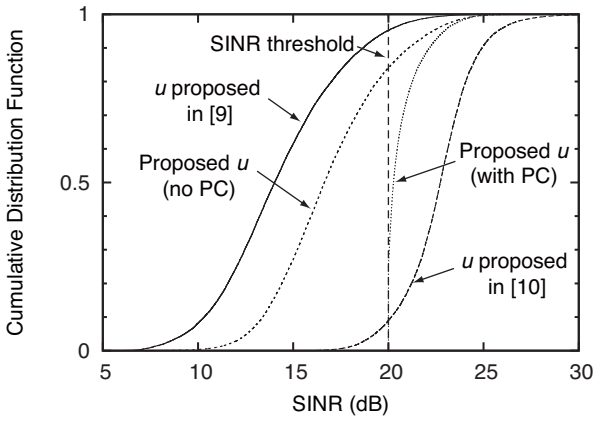


Fig. 5. Cumulative distribution function of received SINR at PUs.

TABLE II
AVERAGE TRANSMIT POWER OF SU TRANSMITTER AT EQUILIBRIUM POINT

No power control	With power control
$5\,\mathrm{mW}$	$3.23\,\mathrm{mW}$

where one SU takes action sequentially. In the sequential play, each player in turn performs the best response strategy in selecting the best channel so as to maximize her utility based on the observed opponent strategies. Thus, an increase in the throughput of individual player contributes to an increase of the overall network throughput which conforms to the definition of potential game. The players then repeatedly play the game until the convergence to a Nash equilibrium is achieved, which is represented by the steady state condition of the network throughput of SUs. At this point, neither player can improve its throughput by selecting another channel different than its best one. We also compare the convergence of network throughput of our proposed utility function with the utility functions proposed in [9] and [10].

Fig. 4 shows the transmit power distribution at SUs after the convergence when the power control (PC) is employed. From Fig. 4, we see that the transmit power distribution at SUs varies due to randomness of SU location in the network, distance between SU transmitters and PU receiver, and the selected channel at equilibrium point. Since the inter-system interference from SU to PU has to be maintained below a certain limit in the spectrum sharing scheme, the power allocation algorithm will adjust the transmit power of SU transmitter in order to avoid severe interference to the PU receiver. The average transmit power of each SU transmitter at equilibrium point with no power control and with power control is shown in Table II. It shows that each SU transmits with maximum transmit power $p_{\max} = 5\,\mathrm{mW}$ when power allocation algorithm is not performed. With power control, the average transmit power of each SU transmitter is reduced to $3.23\,\mathrm{mW}$ in order to ensure that the performance of PUs is not deteriorated. As a result, about 35.4% of energy consumption in terms of total transmit power of 30 SUs can be saved when power allocation algorithm is performed.

We further evaluate the performance of the proposed scheme at higher-priority system after convergence. We compare the performance of the proposed schemes with the scheme that proposed in [9] and in [10]. From Fig. 5, we can see that the utility function u proposed in [9] achieves the lowest performance since the it does not take into account the inter-system interference. The proposed scheme, which takes into account the intra- and inter-system interference improves the SINR performance of PUs. However, the outage probability at PUs of the proposed scheme with no power control (no PC) is still higher since the SU transmitters use maximum transmit power during transmission that may lead severe inter-system interference to the nearby PUs. The proposed scheme with power control (PC) adjusts the transmit power of SU transmitters based on the minimum required SINR at PU receiver. As a result, SINR of PUs is maintained above the threshold, and therefore, zero outage probability at PUs can be achieved. It is also clear from Fig. 5 that the proposed u with power control achieves better performance in terms of outage probability of PUs compared to our previous work [10].

V. CONCLUSION

In this paper, we have proposed the potential game approach for dynamic spectrum sharing in cognitive radio networks. In the proposed spectrum sharing scheme, the secondary users spatially reuse the spectrum and manage the intra- and inter-system interference by selecting the best channel for transmission. We then define the global objective of the game through the formulation of potential function. At the equilibrium point, power control is employed such that the inter-system interference from the lower-priority users to higher-priority users can be maintained below the maximum allowable level. The simulation results show that the proposed spectrum sharing scheme improves the performance of primary users in terms of received SINR and outage probability.

REFERENCES

[1] FCC Spectrum Policy Task Force, "Report of spectrum efficiency working group," Nov. 2002.

[2] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," IEEE Commun. Mag., vol.46, no.4, pp.40–48, Apr. 2008.

[3] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access: Signal processing, networking, and regulatory policy," IEEE Signal Proc. Mag., vol. 24, no.3, pp.79–89, May 2007.

[4] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[5] I. F. Akyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks (Elsevier)*, vol. 50, no. 13, pp. 2127–2159, Sept. 2006.

[6] A. B. MacKenzie and L. A. Da Silva, *Game Theory for Wireless Engineer*, Morgan & Claypool Publisher, 2006.

[7] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Efficient power control via pricing in wireless data networks," IEEE Trans. Commun., vol.50, no.2, pp.291–303, Feb. 2002.

[8] T. Heikkinen, "A potential game approach to distributive power control and scheduling," Elsevier Computer Networks, vol.50, pp.2295–2311, Sept. 2006.

[9] N. Nie and C. Comaniciu, "Adaptive channel allocation spectrum etiquette for cognitive radio networks," *Proc. IEEE Symp. New Frontiers in Dynamic Spectrum Access Network (DySPAN '05)*, Nov. 2005.

[10] I. W. Mustika, K. Yamamoto, H. Murata, and S. Yoshida, "Spectrum sharing with interference management for distributed cognitive radio networks: A potential game approach," *Proc. IEEE VTC Spring'10*, May 2010.

[11] Y. Xiao, X. Shan, and Y. Ren, "Game theory models for IEEE 802.11 DCF in wireless ad hoc networks," *IEEE Communications Magazine*, vol. 43, pp. S22?S26, Mar. 2005.

[12] R. S. Komali and A. B. MacKenzie, "Distributed topology control in ad-hoc networks: A game theoretic perspective," *Proc. IEEE CCNC'06*, pp. 563–568, Jan. 2006.

[13] M. Voorneveld, "Best-response potential games," *Economic Letters*, vol. 66, pp. 289–295, Mar. 2000.

[14] D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, 1991.

[15] D. Monderer and L. S. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, pp. 124–143, May 1996.

AN ANALYSIS COMPARISON OF AODV UU AND BATMAND PERFORMANCE FOR MOBILE AD-HOC NETWORK

Muflich Putera Prathama
Faculty of Electrical & Communications
Telkom Institute of Technology
Bandung, Indonesia
muflich.putera@gmail.com

Istikmal
Telkom University
Bandung, Indonesia
itm@ittelkom.ac.id

Sofia Naning Hertiana
Telkom University
Bandung, Indonesia
snh@ittelkom.ac.id

***Abstract*—Nowadays, technology tends to be developed to wireless technology. One of factors affects that development is mobile capability offered by the system to users. One of current wireless technology is mobile ad-hoc network. Important feature of mobile ad-hoc network implementation is routing protocol such as AODV and BATMAN. Implementing AODV and BATMAN in real life is by using AODV UU and BATMAN Daemon. In this paper, four nodes are implemented to evaluate both routing protocols in mobile ad-hoc network. For mobile scenario, source node moves with velocity about three km/h back and forth to destination node. Sending and receiving files in system is using Distributed Internet Traffic Generator (DITG). Analyzed parameters in this system are average packets loss percentages, average throughput, and average jitter. Implementation of no movement scenario shows that average packets loss percentage ranges from 0% to 3.358%, average throughput ranges from 393.1821813 Kbps to 409.9955152 Kbps, and average jitter ranges from 0.00052182 s to 0.00604264 s. In addition, implementation of source node movement scenario shows that average packets loss percentage ranges from 0.002% to 4.276%, average throughput ranges from 386.6820413 Kbps to 409.8114591 Kbps, average jitter ranges from 0.00202616 s to 0.00609198 s.**

Keywords—Mobile ad-hoc network; AODV UU; BATMANd; DITG

I. INTRODUCTION

At this moment, technology development tends to be wireless technology. One parameter affects this kind of development is mobility offered by system to users. Mobile ad-hoc network is current still-developed wireless technology. Application of mobile ad-hoc network ranges in many fields such as military, disaster communication, and workspace. For military, mobile ad-hoc network can be implemented to support communication in battlefield in case there is no available communication infrastructure. If mobile ad-hoc network is deployed for workspace, then employees can use it for sending/receiving document anytime anywhere or teleconference with another employees.

Routing protocols are needed to provide and decide communication route in mobile ad-hoc network. There are many types of routing protocol available for mobile ad-hoc network. In order to evaluate current mobile ad-hoc network, we use Ad-Hoc On Demand Vector (AODV) and Better Approach To Mobile Adhoc Networking (BATMAN) as routing protocols. There are several considerations why we use both routing protocols in this paper. First, both routing protocols implementation are still developed recently. Besides, both routing protocols are common-used for research evaluation. Third, both routing protocols have different method to perform routing process. So, they can be evaluated to decide which routing protocols work better. Implementation of AODV in Linux-based system is AODV Uppsala University which was last developed in 2009. Implementation of BATMAN in Linux-based system is BATMAN daemon which is still developed in 2013. This experiment is performed to evaluate performance of both routing protocols after they are implemented for mobile ad-hoc network. The performance is evaluated by analysing parameter such as average packets loss percentages, average throughput and average jitter.

In this paper, our objective is evaluating how number of hops for each routing protocol can affect network performance. Besides, we evaluate how much network performance suffers from source node mobility. In order to understand result of the implementation, network performance parameters are evaluated such as average packets loss percentages, average throughput and average jitter.

II. THEORY

Mobile ad-hoc network is group of nodes forming a network which is capable of communicating wirelessly. In order to communicate, nodes do not need fixed

communication infrastructure such as access point or base station. Nodes rely on other nodes in the network to form a communication route. Other advantage of mobile ad-hoc network is mobility. Users do not need to stay forever in one place in order to communicate. They can move freely. Some characteristics of mobile ad-hoc network are free mobility, multi-hop communication, and autonomous-terminal capability. Multi-hop communication means that communication route in mobile ad-hoc network rely on intermediate nodes to form it.

A. Ad-Hoc On Demand Vector

Ad-Hoc On Demand Vector (AODV) is reactive routing protocol for ad-hoc network [1]. Route creating is triggered by communication demand at source node. In order to find a route to destination node, flooding method is used in AODV. Creating-route message is flooded in the whole network until destination node is found. Implementation of AODV in Linux-based system is done by using AODV Uppsala University (AODV UU) which was last developed in 2009. AODV UU development is done by Erik Nordstrom Sweden. Until 2013, AODV UU has reached version 0.9.6. In this version, there are some changes such as capability of multi interface, displaying sequence number in routing table, supporting local link repair and can be deployed in kernel 2.6.x.

Function of local link repair is repairing broken link to preserve less-delay communication. If local link repair mode is not activated, broken-link-detecting node must send route-error message to source node. Then source node must update its routing table again. Of course this method consumes more time than local link repair mode. Another advantage of local link repair is node can keep sending packets to destination node but those packets will be buffered first in link-repairing node.

In AODV, there are several types of message used for communication such as RREQ, RREP, and RERR. RREQ is used for requesting route creation to destination node. RREP is used for informing source node that destination node is found. RRER is used for informing broken-link related message to nodes in precursor list.

B. Better Approach To Mobile Adhoc Networking

Better Approach To Mobile Adhoc Networking (BATMAN) is proactive routing protocol for ad-hoc network which has method in route creation by broadcasting originator message (OGM) to direct-neighbor nodes [2]. Those OGMs will be rebroadcasted with certain rule when direct-neighbor nodes receive them. Deciding best neighbor in BATMAN is using transmit quality (TQ) to predict how successful packets can be sent to destination node. Those prediction value is based on how much OGMs can be received by source node.

BATMAN daemon (BATMANd) is implementation of BATMAN routing protocol in Linux-based system. BATMANd is developed by a group of researcher in German. Until 2013, BATMANd is still developed.

To collect OGMs, sliding window is used in BATMAN. Sliding window is used for recording sequence number of every received OGM. By doing this, node knows whether those OGMs are in-window OGMs or out-of-range OGMs. In-window sequence number shows list of just-received OGMs/fresh OGMs. Out-of-range sequence number shows list of OGMs outside in-window sequence number. When node receives out-of-range OGMs, the in-window will be moved.

In BATMAN, one of parameters used for deciding whether a node can send packets successfully is link quality. Link quality is divided into three types such as receive quality (RQ), echo quality (EQ), and transmit quality (TQ) [3]. RQ shows how successful a node receives OGMs sent by other nodes. EQ shows how successful a node receives its own OGMs which are result of rebroadcasting by other nodes. TQ shows how successful a node can send OGMs to other nodes. In BATMANd, TQ is displayed in the routing table.

Every sent OGM, TQ_{OGM} value will be set to 255. Upon receiving that OGM, a node will multiply received-OGM TQ_{OGM} value with TQ_{LOCAL}. TQ_{LOCAL} is calculated based on communication with previous node. So, if more hops are passed by for communication, then the probability of sending packets successfully via that route will be lower. This will lead into higher packets loss.

III. IMPLEMENTATION CONFIGURATION

Number of nodes implemented in this research ranges from 2 nodes to 4 nodes. AODV UU and BATMANd is used in every node. In every scenario, there is one source node acts as packets sender. There is always one destination node acts as packets receiver. System implementation is done in research room, N209, Telkom Institute of Technology. Length of implementation room is only 14 metres. In order to ensure multi-hop communication established in this small room, MAC filtering is used for blocking non-direct neighbor nodes. This method is called forced multi-hop. By doing this, every node will be forced to forward the packets to its direct-neighbor node not to its non-direct neighbor nodes. Implemented scenarios in this research are divided into two general scenarios such as no-movement-at-all-nodes scenario and three-km/h-movement-for-source-node scenario. Table 1, figure 1 up to 6 provides system configuration for this research.

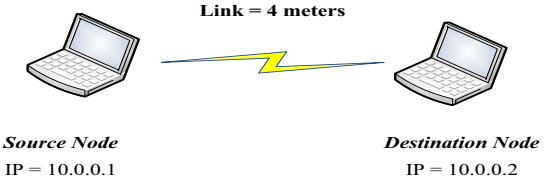


Fig. 1 One hop model for no movement scenario

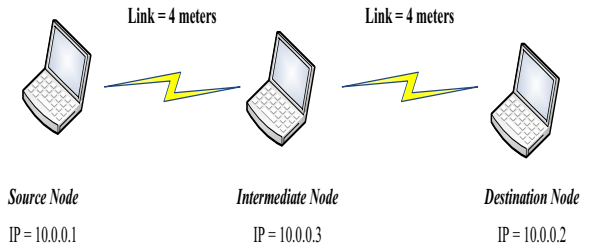


Fig.2 Two hops model for no movement scenario

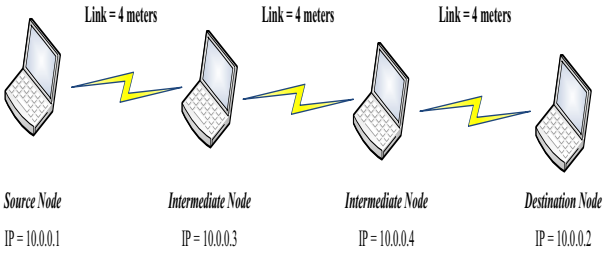


Fig. 3 Three hops model for no movement scenario

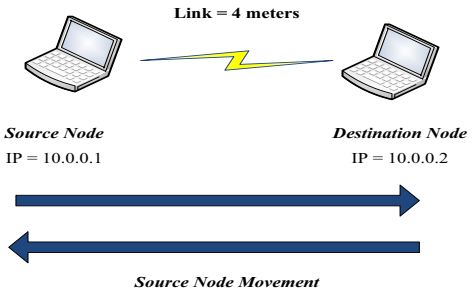


Fig. 4 One hop model for source node movement scenario

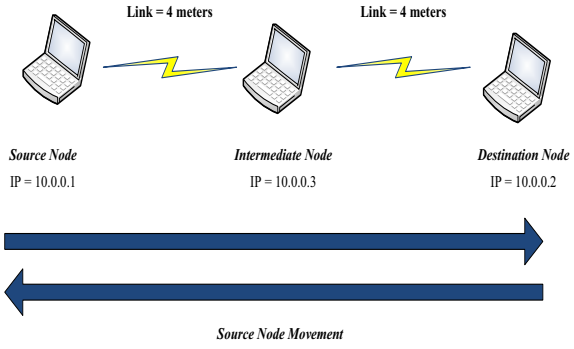


Fig. 5 Two hops model for source node movement scenario

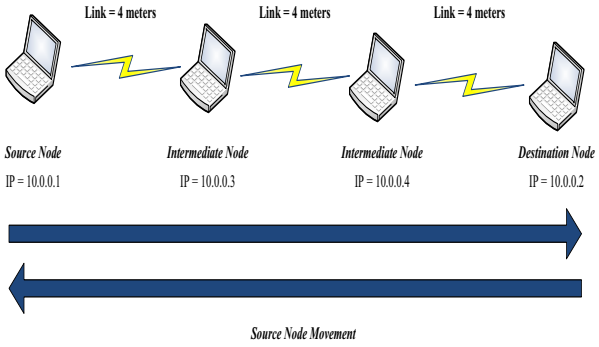


Fig. 6 Three hops model for source node movement scenario

TABLE I. SYSTEM SPECIFICATION

No	Parameter	Information
1.	Transport protocol	UDP
2.	Environment	Indoor
3.	Packets size	512 Bytes
4.	Sending duration	10 seconds
5.	Packets per sample	1000 packets
6.	Number of samples	50 samples
7.	Number of nodes	4 nodes
8.	Frequency	2.412 GHz
9.	Traffic generator	Distributed Internet Traffic Generator

As stated in Markov Chain model, higher numbers of hop passed by the packets lead to higher probability of packets loss which leads to higher packets loss [4]. For example, one-hop communication will have lower packets loss than two-hops communication. Equation (1) states Markov Chain model. N represents numbers of hop in communication route. Ln represents numbers of possible route. In this paper, number of possible route has value of 1 because implementation model is only a straight route which has no alternative route.

$$P_{B,N} = 1 - \frac{1}{N} \sum_{n=1}^{N} (1 - p_{B,N})^{Ln} \quad (1)$$

In order to make mobile ad-hoc network works, there are several softwares and operating system needed in this research such as :

1) Ubuntu Lucid Lynx

Ubuntu Lucid Lynx is used in this research because both routing protocols need kernel 2.6.x to work. Besides, AODV UU and BATMANd works in Linux-based system.

2) Aodv-uu-0.9.6

Aodv-uu-0.9.6 is used for enabling AODV as routing protocol in every node.

3) Batmand-0.3.2

Batmand-0.3.2 is used for enabling BATMAN as routing protocol in every node.

4) Distributed Internet Traffic Generator

Distributed Internet Traffic Generator (DITG) is needed for generating traffic from source node. Many parameters can be configured in DITG such as packets size, generating duration, number of packets, transport protocol, traffic distribution, etc.

5) Iptables

Iptables is needed to ensure that there is no direct communication between source node and destination node. MAC filtering is done by using this iptables so multi-hop communication can be achieved.

6) Ping

Ping is used for ensuring every node can keep communicating even though MAC filtering has been used to ensure no-direct-neighbor communication.

IV. IMPLEMENTATION RESULT & ANALYSIS

A. Analysis of Hops Number in AODV UU and BATMANd Related to Average Packets Loss Percentages at No Movement Scenario

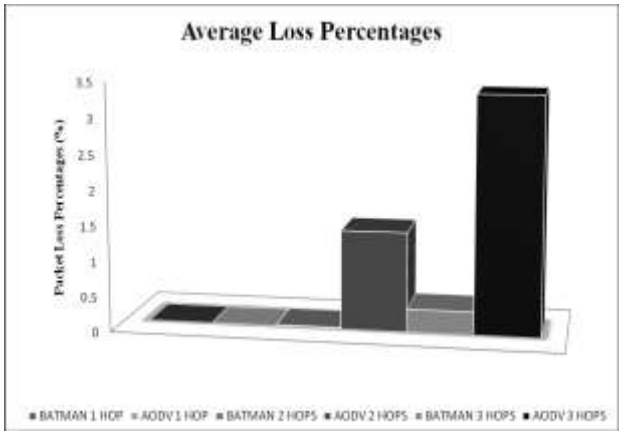


Fig. 7 Average packets loss percentages for no movement scenario

Based on Markov Chain model, it is quite clear that increased number of hops in a route causing higher packets loss probability. It will lead to higher packets loss percentages. Both AODV UU and BATMANd have higher packets loss percentages when number of hops is increased.

Figure 7 shows that in every number of hops, BATMANd has lower packets loss percentages than AODV UU. This is due to design of routing protocol algorithm. When it comes to deciding best route to destination, BATMANd collect number of OGMs, then making calculation so it can determine how good a route. This process is listed in TQ theory. This is not how it works in AODV UU. AODV UU only searches for a route to destination without determining how good quality of the link. It only searches and finds, then route is available. That is why BATMANd has lower average packets loss percentages than AODV UU.

B. Analysis of Hops Number in AODV UU and BATMANd Related to Average Throughput at No Movement Scenario

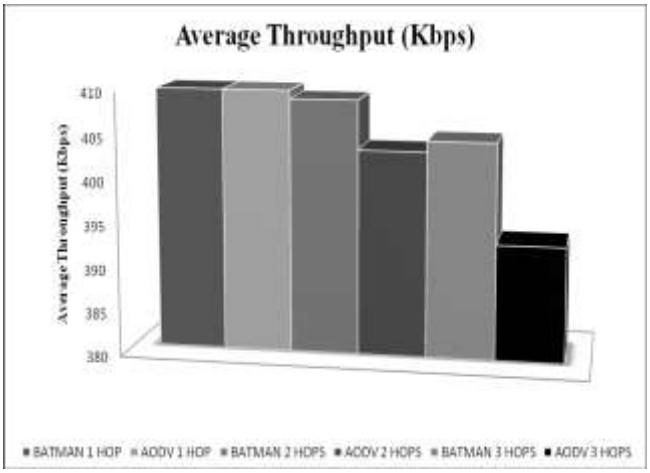


Fig. 8 Average throughput for no movement scenario

Average throughput becomes one of measured performance for this research. We choose average throughput over spontaneous throughput because average throughput gives better accuracy of throughput when there is communication. Based on figure 8, it is quite clear that higher hops cause lower achieved average throughput in destination node.

This can occur because higher hops cause higher packets loss percentages. Higher packets loss percentages means that number of successful sent packets to the destination is getting decreased which cause lower average throughput. It is clear that BATMANd throughput in every type of hops is always higher than AODV UU throughput. This happens because BATMANd has lower packets loss than AODV UU when experiment was implemented. So, packets loss affects value of system throughput.

C. Analysis of Hops Number in AODV UU and BATMANd Related to Average Jitter at No Movement Scenario

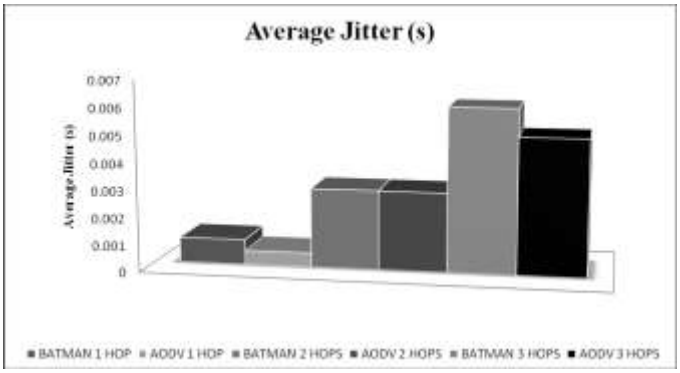


Fig. 9 Average jitter for no movement scenario

Figure 9 shows that higher number of hops tends to cause higher average jitter. This is due to longer processing delay as the number of hops increased [5]. BATMANd has higher average jitter than AODV UU in every number of hops. This could be caused by higher complexity in design of BATMANd than AODV UU such as BATMANd needs to collect number of OGMs in order to keep route updated. Then BATMANd needs to do calculation to determine TQ of the link. So, this is one of reason why BATMANd has slightly higher jitter than AODV UU.

D. *Analysis of Hops Number and Source Node Mobility in AODV UU and BATMANd Related to Average Packets Loss Percentages*

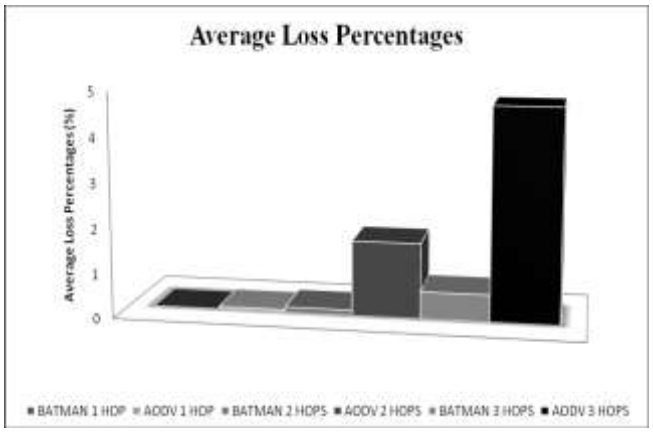


Fig. 10 Average packets loss percentages for movement scenario

Figure 10 shows that in every number of hops, BATMANd has lower average packets loss percentages than AODV UU. This is due to TQ system implemented in BATMANd. Besides, higher number of packets loss leads to higher packets loss percentages. Markov Chain model proves this theoretically. It can be noticed that movement scenario has higher packets loss percentages than no movement scenario because source node movement causes varied multipath propagation. This leads to worse performance at the implemented system.

E. *Analysis of Hops Number and Source Node Mobility in AODV UU and BATMANd Related to Average Throughput*

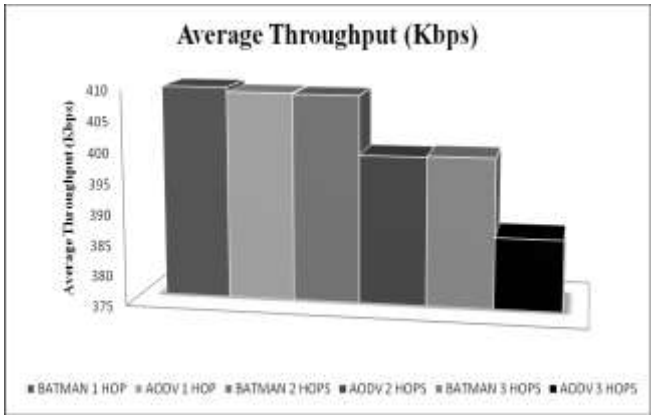


Fig. 11 Average throughput for movement scenario

Figure 11 show that movement scenario has lower average throughout than no movement scenario. This is because no movement scenario has lower average packets loss percentages which leads to lower packets loss. Lower packets loss will cause higher throughput because more packets are received at the destination node.

F. *Analysis of Hops Number and Source Node Mobility in AODV UU and BATMANd Related to Average Jitter*

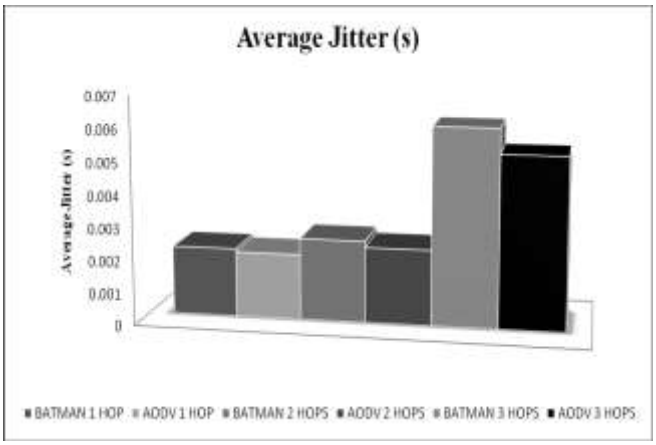


Fig. 12 Average jitter for movement scenario

Figure 12 shows that higher number of hops tends to cause higher average jitter. This is due to longer processing delay as the number of hops increased. BATMANd has higher average jitter than AODV UU in every number of hops. This could be caused by higher complexity in design of BATMANd than AODV UU such as BATMANd needs to collect number of OGMs. Then BATMANd needs to do calculation to determine TQ of the link. Besides, higher jitter in movement scenario is caused by source node keeps moving which causes longer time for sending packets to destination node. This leads to higher jitter in implemented system.

V. CONCLUSION

In this paper, we presented research about how number of hops and source node movement can affect performance of mobile ad-hoc network. Implementation result shows that higher number of hops leads to higher packets loss percentages, lower average throughput, and higher average jitter. It is proven that source node movement can worse the performance of mobile ad-hoc network too. In every measured metric performance, BATMANd outperformed AODV UU except in average jitter. This is due to route processing in BATMANd more complex than AODV UU.

REFERENCES

[1] C. Perkins, E.M. Belding-Royer, and S.R. Das, "Ad hoc on-demand distance vector (AODV) routing," Internet Engineering Task Force, 2003.

[2] A. Neumann, C. Aichele, M. Lindner, and S. Wunderlich, "Better approach to mobile ad-hoc networking (B.A.T.M.A.N.)," Internet Engineering Task Force, 2008.

[3] F. Oehlmann, "Simulation of the "better approach to mobile ddhoc networking" protocol," Technische Universitat Munchen, 2011.

[4] R. Haywood and X. Peng, "On packet loss performance under varying network conditions with path diversity," Aston University, 2012.

[5] C. Demichelis and P. Chimento, "IP packet delay variation metric for IP performance metrics (IPPM)," Internet Engineering Task Force, 2002.

An Analysis Method of Effect of Linear Polarized Electromagnetic Exposure from Mobile Phone to Human Head with Various Incident Angles

Alfredo Bayu Satriya, Eko Setijadi
Electrical Engineering Department
Institut Teknologi Sepuluh Nopember
Surabaya
alfredo12@mhs.ee.its.ac.id, ekoset@ee.its.ac.id

***Abstract*—This research is focused on the effect of various incident angles of the linear polarized electromagnetic exposure from mobile phone to the value of Specific Absorption Rate (SAR) in human head which is a safety standardization for mobile phone user. Method of Moment is chosen to calculate the electric field in the human head because it is the most suitable method. The incident electromagnetic is assumed as a plane wave with a frequency at 900 MHz and power density of antenna for 4.5 W/m^2. The integral used for calculation is Tensor Integral Equation (TIE). Human head is modelled as a homogeneous spheroid and is divided into cubical cells. The incident angles which are examined are 0°(horizontal polarization), 15°, 30°, 45°, 60°, 75°and 90°(vertical polarization). The results show that the incident angle does not affect the average SAR in the human head, but it affects the local and maximum SAR of the cells in human head. The difference of seven angles maximum SAR does not vary in the form of linear but in the form of damped sinusoidal. The angle which produces the least maximum SAR is 0°namely 0.1521 W/kg whereas the angle which produces the highest maximum SAR is produced by angle 15°namely 0.1748 W/kg. The highest SAR is obtained around the side of the head near the electromagnetic exposure source. None of the SAR produced by the variation of incident angle is above the limit of safety standard.**

Keywords—electric field; electromagnetic incident angle; homogeneous spheroid; Method of Moment; SAR

I. INTRODUCTION

Mobile phone is one of the telecommunication tools which is always close to human, especially when human makes a phone call, it is very close to the human brain. At that moment one does not know that electromagnetic wave which is produced by the phone could make a bad effect to one's brain. In Sweden a research shows that there is a headache because of mobile phone [1], another research concludes that 70% of mobile phone users also get headache [2]. The use of mobile phone also can make a sleeping disorder [3], and there are still some researches which show the dangers of mobile phone use.

What makes the danger of electromagnetic exposure is the heat which is produced by the electromagnetic exposure. The heat in ones head and brain could lead to hyperthermia. Hyperthermia is a condition when the temperature in the body increasing than its normal thermoregulatory because the body absorbs heat more than it can dissipate. Hyperthermia could lead into brain damage or even dead. The limit of increasing temperature in the human brain is 4.5°C [4]. Because of this danger of electromagnetic exposure, there are many safety standardizations made, for the example [5] and [6]. Those standardizations are made with SAR (Specific Absorption Rate) parameter. SAR is a parameter which shows the power of an electromagnetic wave which is absorbed by the human tissue.

Several works have investigated the effect of electromagnetic exposure to the human head [7-21], mostly the results show that the value of SAR is less than safety standard, but in some cases it is found that SAR is higher, namely the electromagnetic exposure of an omnidirectional antenna [9], a circumstance where the human is about one meter radius from base station [18], and human uses mobile phone in a closed elevator [19]. Based on these works we need to pay more attention on the danger of electromagnetic exposure. These works [7-21] have investigated several parameters which can affect SAR, they are mobile phone's transmit power, antenna, antenna radiation pattern, frequency, polarization, the gap between mobile phone and the human head, and the human head geometry. Each of these parameters has been computed with various method, including vertical and horizontal polarization. These two polarizations resulting different value of SAR in the human head [16].

Although there has been a paper about the effect of different polarization, but there has not a work which investigates the effect of electromagnetic incident angle to SAR in the human head. It becomes important because hand phone user usually holds hand phone in various angles while making a call and it makes electromagnetic also propagates in various incident angles. In addition, there is no standard for incident angle of hand phone antenna to produce the least SAR in the human head.

Before getting the value of SAR in the human head, it is necessary to know how big the electric field which reaches it. In order to get the electric field, the electromagnetic computational is used. Electromagnetic computation is a Maxwell Equation solution using numerical approximation. Numerical approximation is used because the solution of an actual problem is very complex and needs several assumptions

for problem simplification [22]. There are various numerical methods to solve the electromagnetic problem, namely FEM, MoM, FDTD, FVTD, DWP, DWT, TLM, Monte Carlo etc. In several previous works [7-21], MoM and FDTD are two methods mostly used in counting electric field in the human body and head. MoM is a frequency based method and it has a better accuracy than FDTD [18].

II. Methods and Models

The procedure of MOM requires subdivision of solution region, source equation, and integral equation

A. Human Head Model

Human head is modelled as a homogeneous spheroid like paper [19] and is subdivided as the procedure of MoM. The human head is modelled as a spheroid to simplify many complex details and arbitrary surface in human head, spheroid has the most same geometry with human head among other three dimensional geometry. In Table I we can see that human head consists of several tissues which have different constitutive parameters (permitivities and conductivities) but in this paper the human head structure is assumed homogeneous for simplification and because the thickness of each tissue is very thin. This model has a radius of 9 cm, permitivity ϵ_r = 45.8, conductivity σ = 0.77 and mass density ρ = 1030 kg/m^3. This model is subdivided into cubical cells/ blocks. The cubical cell size based on [23] must be $\frac{1}{4}$ wavelength or less to get a good accuracy of the computation. The maximum size of cubical cell is:

$$
\begin{aligned}
f &= 900 \text{ MHz} \\
c &= 3 \times 10^8 \text{ m/s} \\
\lambda &= \frac{c}{f} = \frac{3 \times 10^8}{9 \times 10^8} = 0.33 \text{ m} \\
s_{max} &= \frac{\lambda}{4} = \frac{0.33}{4} = 8.33 \text{ cm}
\end{aligned}
$$

With maximum cell size 9 cm, the possible modelling of the human head consists of 32 cubical cells (N = 32) of which the size is 4.5 cm, but in this paper, we select a model which consists of 160 cells 160 as presented in Fig. 1.

B. Electric Field Polarization Equation

Nowadays, there are three kinds of antenna which is often used in mobile phone namely helix antenna, planar inverted-F antenna (PIFA) and Planar Monopole Antenna (PMA) [25]. Both PIFA and PMA could produce electromagnetic wave with linear polarization which is used for mobile communication. In order not to limit the analysis for certain antenna design, this paper simulate the source electromagnetic without antenna design but by formulating the electromagnetic pattern. The incident electromagnetic wave is assumed as uniform plane wave and its electric field is assumed linear polarized. When the incident electric field propagates along z-axes towards z-positive, the electric field is defined as:

$$\tilde{E} = (\hat{x}a_x + \hat{y}a_y e^{j\delta})e^{-jkz} \quad (1)$$

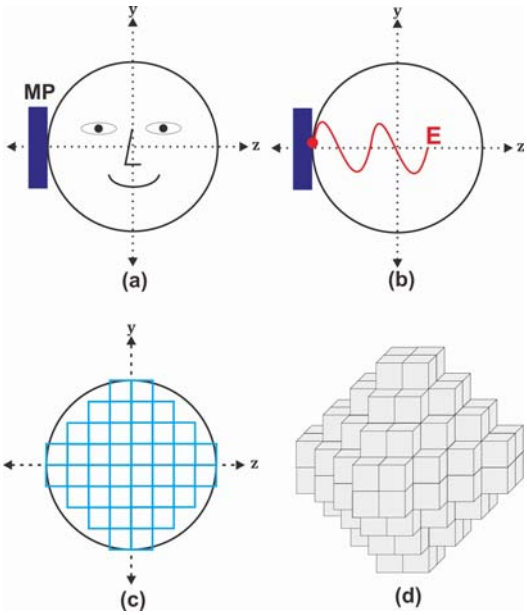


Fig. 1. Problem Modelling: (a)Human head and mobile phone,(b) Antenna feeder point, Homogeneous spheroid consists of 160 cells: (c) 2-D (d) 3-D

TABLE I. Properties of Human Head at 900 MHz

Tissues	Radius (cm)	ϵ_r	σ	ρ (kg/m^3)
Skin	9.00	39.5	0.69	1010
Fat	8.90	10.0	0.17	920
Bone	8.76	17.4	0.19	1810
Brain	8.35	44.1	0.89	1040

Suppose E has an incident angle (θ) like in Fig.2, and $a_x = a\cos\theta$ and $a_y = a\sin\theta$, equation (1) becomes :

$$
\begin{aligned}
\tilde{E} &= (\hat{x}a_x + \hat{y}a_y e^{j\delta})e^{-jkz} \\
&= (\hat{x}a\cos\theta + \hat{y}a\sin\theta)e^{-jkz} \\
\tilde{E} &= (\hat{x}\cos\theta + \hat{y}\sin\theta)ae^{-jkz} \quad (2)
\end{aligned}
$$

where δ is phase difference (δ = 0 and π for linear polarization). The value of a as an incident electric field can be calculated by :

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \quad (3)$$

$$\mathbf{H} = \frac{\mathbf{E}}{\eta} \quad (4)$$

so $a = E_0 = 41.19$ Vm^{-1}

C. Tensor Integral Equation

Tensor Integral Equation (TIE) is an equation which is proposed by Livesay and Chen to count the electric field in the human body from the exposure of the electromagnetic wave. Suppose an electromagnetic wave with incident electric field ($\mathbf{E}^i$(r)) propagates into an arbitrary biological body, there will be a part of the electric field which will propagate inside the biological body ($\mathbf{E}$(r)) and there is another part which will be scattered ($\mathbf{E}^s$(r)), the relation between these three fields is defined as :

$$\mathbf{E}(r) = \mathbf{E}^i(r) + \mathbf{E}^s(r) \quad (5)$$

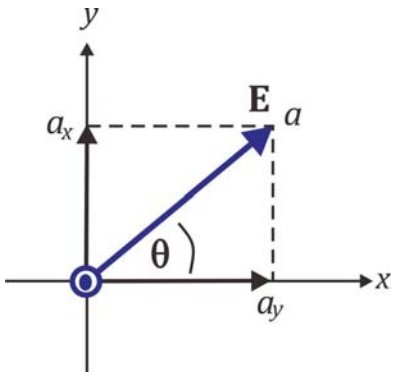


Fig. 2. Linear polarized electromagnetic wave towards z-positive with incident angle

the current density in arbitrary biological body is defined as:

$$\mathbf{J}_{eq}(r) = [\sigma(r) + j\omega(\epsilon(r) - \epsilon_0)]\mathbf{E}(r) = \tau(r)\mathbf{E}(r) \quad (6)$$

with the scattered electric field from arbitrary electric field can be expressed as:

$$\mathbf{E}^s(r) = PV \int_V \mathbf{J}_{eq}(r') \cdot \mathbf{G}(r, r')\, dV' \; - \frac{\mathbf{J}_{eq}(r)}{3j\omega\epsilon_0} \quad (7)$$

where

$$\mathbf{G}(r, r') = -j\omega\mu_0[1 + \frac{\nabla\nabla}{k_0^2}]\psi(r, r') \quad (8)$$

$$\psi(r, r') = \frac{exp(-jk_0|r - r'|)}{4\pi|r - r'|} \quad (9)$$

$$k_0 = \omega(\mu_0\epsilon_0)^{1/2} \quad (10)$$

After some derivations [23], we get the Tensor Integral Equation which is defined as :

$$\mathbf{E}^i(r) = [1 + \frac{\tau(r)}{3j\omega\epsilon_0}]\mathbf{E}(r) - PV \int_V \tau(r')\mathbf{E}(r') \cdot \mathbf{G}(r, r')\, dV' \quad (11)$$

D. Method of Moment

Method of Moment (MOM) is a numerical technique which is designed to solve an integral equation. MOM has become popular since the research made by Richmond in 1965 and Harrington in 1967. Generally, MOM has four steps, namely :
1. Derivation of the appropriate integral equation,
2. Discretization of the integral equation into matrix equation,
3. Evaluation of matrix elements,
4. Obtaining the desired parameters. [22]

The derivation, discretizaton and evaluation of the electric field in arbitrary biological body in generally has done in [23]. The discretization of (7) is :

$$\begin{bmatrix} [G_{xx}] & [G_{xy}] & [G_{xz}] \\ [G_{yx}] & [G_{yy}] & [G_{yz}] \\ [G_{zx}] & [G_{zy}] & [G_{zz}] \end{bmatrix} \begin{bmatrix} [E_x] \\ [E_y] \\ [E_z] \end{bmatrix} = - \begin{bmatrix} [E_x^i] \\ [E_y^i] \\ [E_z^i] \end{bmatrix} \quad (12)$$

where

$$E_{x_p} = \begin{bmatrix} E_{x_p}(r_1) \\ \vdots \\ E_{x_p}(r_N) \end{bmatrix} E_{x_p^i} = \begin{bmatrix} E_{x_p^i}(r_1) \\ \vdots \\ E_{x_p^i}(r_N) \end{bmatrix} \quad (13)$$

N = number of cell

Green function matrix element evaluation of (7) is defined as:

for m $\neq$ n

$$G_{x_p x_q}^{mn} = \frac{-j\omega\mu_0 k_0 \Delta v_n \tau(r_n) exp(-j\alpha_{mn})}{4\pi\alpha_{mn}^3} \times [(\alpha_{mn}^2 - 1 - j\alpha_{mn})\delta_{pq} + \cos\theta_{xp}^{mn} \cos\theta_{xq}^{mn}(3 - \alpha_{mn}^2 + 3j\alpha_{mn})] \quad (14)$$

where

$$\alpha_{mn} = k_0|r_m - r_n| \quad (15)$$

$$\cos\theta_{xp}^{mn} = \frac{x_p^m - x_p^n}{|r_m - r_n|}; \cos\theta_{xq}^{mn} = \frac{x_q^m - x_q^n}{|r_m - r_n|} \quad (16)$$

$$r_m = (x_1^m, x_2^m, x_3^m); r_n = (x_1^n, x_2^n, x_3^n) \quad (17)$$

$m, n = 1, 2, ...N$
$p, q = 1, 2, 3$

for m = n

$$G_{x_p x_q}^{mn} = \delta_{pq}\{\frac{-2j\omega\mu_0\tau(r_n)}{3k_0}[exp(-jk_0 a_n)(1 + jk_0 a_n) - 1] - [1 + \frac{\tau(r_m)}{3j\omega\epsilon_0}]\} \quad (18)$$

where

$$a_n = (\frac{3\Delta v_n}{4\pi})^{\frac{1}{3}} \quad (19)$$

E. SAR

Specific Absorption Rate (SAR) is a parameter that describes the amount of electromagnetic power absorbed by the human tissue while being exposured by elctromagnetic waves. SAR calculation is described as:

$$SAR = \frac{\sigma(r)|E|^2}{2\rho(r)} \quad (20)$$

where

$$|E| = \sqrt{E_x^2 + E_y^2 + E_z^2} \quad (21)$$

There are many organizations including FCC, IEC, CENELEC, ANSI, ICNIRP, and IEEE which have specified their safety standard for maximum SAR in the human tissue, but they all have the same limit, namely 1.6 W/kg in each gram of human tissue and 2 W/kg in each ten grams of human tissue.

III. Results

The results show that incident angle of electromagnetic does not affect the average SAR of the human head model. With the change of range 15°, from the degree of 0 till 90, the average SAR is remain 0.0216 W/kg. This value of average SAR is much less than the maximum limit of the safety standard. While the average SAR shows steady value, the maximum SAR of the human head does not. The obtained maximum SAR of simulation in the human head is never same as the incident angle varies from 0°to 90°. For getting the simultaneous effect of the incident angle, the interpolation of the maximum SAR values is presented in Fig. 3 and maximum SAR values of each angle are presented in Table II. The change

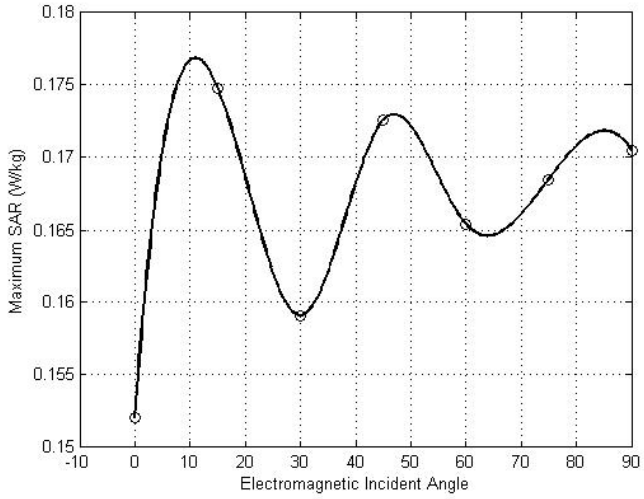


Fig. 3. Interpolation result of incident angle effect to the maximum SAR in human head

TABLE II. MAXIMUM ANGLE OF VARIOUS INCIDENT ANGLE

Incident Angle	Maximum SAR (W/kg)
0	0.1521
15	0.1748
30	0.1591
45	0.1726
60	0.1654
75	0.1685
90	0.1705

of the maximum SAR value is not linear, the interpolation of SAR produces a graphic which is like a damped sinusoidal, where the angle 0°or horizontal polarization produces the least maximum SAR that is 0.1521 W/kg and the angle 15 produces the highest maximum SAR that is 0.1748 W/kg. The different maximum SAR value between these seven angles can be affected from the local SAR of each cell. The distribution of electric field in each cell changes when the incident angle is changed. This can be seen in Fig. 4. While being exposured from electromagnetic, there are parts of the human head which get more electric field than the others, this different distribution of the electric field in each cell are squared in order to get the SAR distribution, so the differences in SAR of cells are quite significant. This significant differences can be seen from the visualization of SAR in the center of human head which are refer to Fig. 5 and presented in Fig.6 and 7 (only two angles are presented). There are parts of the head which have a relatively high SAR, about 0.17 W/kg (shown in white color) but there are also parts which are not affected by the electromagnetic exposure since there are no SAR in that part (shown in black color). However, these seven values of maximum SAR are still below the safety standard.

IV. CONCLUSION

The incident angle does not affect the average SAR in the human head, but it affects the local SAR of the cells. The incident angle affects the maximum SAR in human head since the SAR distribution also changes. The maximum SAR does not vary in the form of linear but in the form of damped

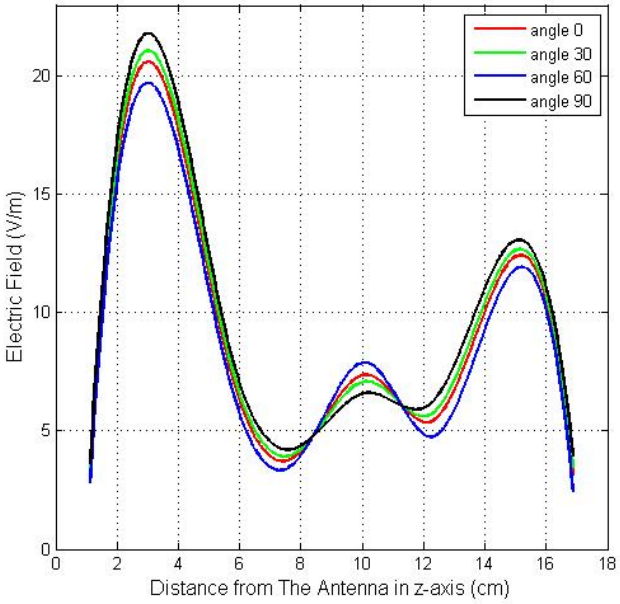


Fig. 4. Electric field in human head due to electromagnetic exposure with various incident angle

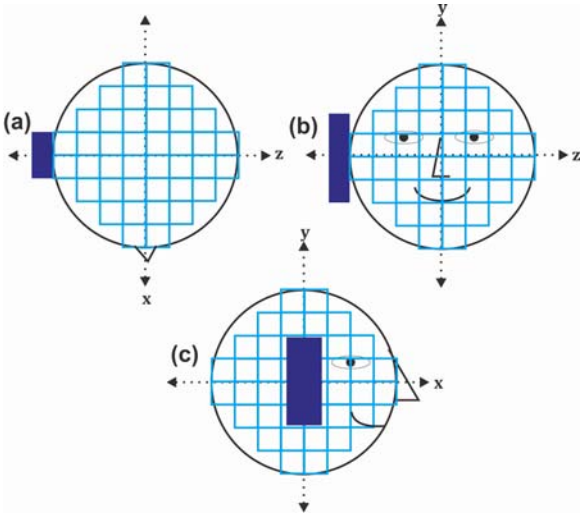


Fig. 5. Cells in the center of human head model, (a) upper view, (b) front view, (c) side view

sinusoidal. The angle which produces the least maximum SAR is 0°or the polarization whereas the angle which produces the highest maximum SAR is 15°. The distribution SAR in the human head is not prevalent, the highest SAR is obtained around the side of the head near the electromagnetic exposure. None of the SAR produced by the variation of incident angle is above the limit of safety standard.

REFERENCES

[1] G. Oftedal, J. Wilen, M. Sandstrm, and K. H. Mild, "Symptoms experienced in connection with mobile phone use," *Occupational Med.*, vol. 50, no. 1, pp. 19, 2000.

[2] Szyjkowska A., Bortkiewicz A., Szymczak W., Makowiec-Dabrowska T., "Subjective symptoms related to mobile phone use–a pilot study",

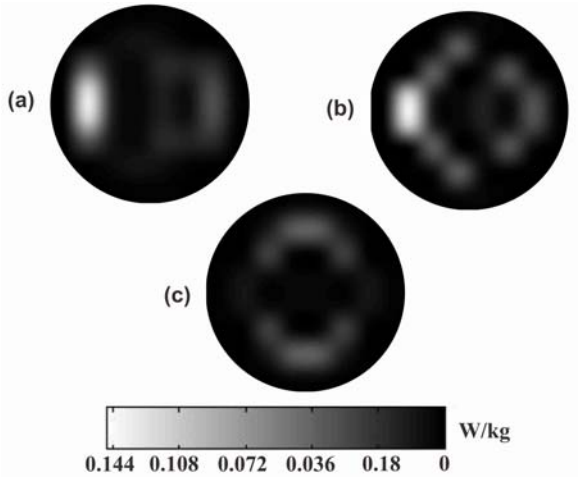


Fig. 6. SAR visualization of 0°incident angle in the center of human head from (a) upper view, (b) front view, (c) side view

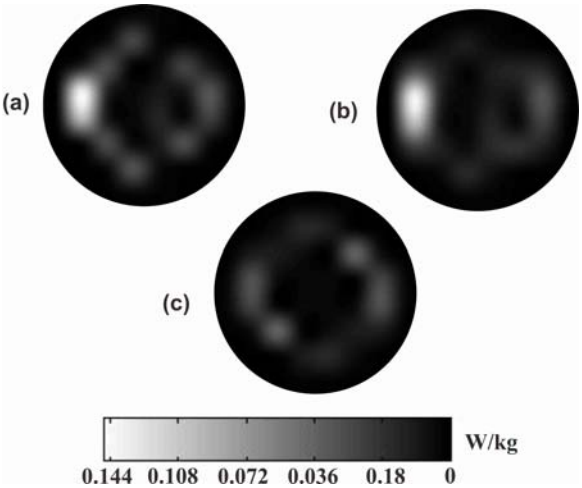


Fig. 7. SAR visualization of 90°incident angle in the center of human head from (a) upper view, (b) front view, (c) side view

PubMed.gov, US National Library of Medicine National Institutes of Health.2005.

[3] S. Thome, A. Hrenstam and M. Hagberg, "Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults - a prospective cohort study", *BMC Public Health* vol. 11, 31 January 2011.

[4] A. C. Guyton and J. E. Hall, *Textbook of Medical Physiology. Philadelphia*, PA: Saunders, 1996.

[5] "Evaluating Compliance with FCC Guidelines for Human Exposure to Radio Frequency Electromagnetic Fields," Federal Communications Commission, Washington, DC, OET Bull. 65, 1997.

[6] "Guidelines for limiting exposure to time-varying electric, magnetic and electromagnetic fields (up to 300 Ghz)," *Health Phys*., vol. 74, pp. 494522, 1998.

[7] D. M. Sullivan, D. T. Borup, O. P. Gandhi, "Use of the finite-difference time-domain method in calculating EM absorption in human tissues," *IEEE Transactions On Biomedical Engineering*, Vol. 34, No. 2, February 1987.

[8] O. P. Gandhi and J.-Y. Chen, "Electromagnetic absorption in the human head from experimental 6-GHz handheld transceivers," *IEEE Transactions On Electromagnetic Compatibility*, Vol. 37, No. 4, November 1995.

[9] K. W. Kim and Y R-Samii, "Handset antennas and humans at Ka-band: The Importance of Directional Antennas," *IEEE Transactions On Antennas And Propagation*, Vol. 46, No. 6, June 1998

[10] G. M. J. Van Leeuwen, J. J. W. Lagendijk, B. J. A. M. Van Leersum, A. P. M. Zwamborn, S. N. Hornsleth, and A. N. T. Kotte, "Calculation of change in brain temperatures due to exposure to a mobile phone", *Phys. Med. Biol.*, vol. 44, pp. 23672379, 1999.

[11] J.Wang and O. Fujiwara, "FDTD Computation of temperature rise in the human head for portable telephones," *IEEE Trans. Microwave Theory Tech.*, vol. 47, pp. 15281534, Aug. 1999.

[12] P. Bernardi, M. Cavagnaro, S. Pisa, and E. Piuzzi, "Specific Absorption Rate and temperature increases in the head of a cellular-phone user," *IEEE Trans. Microwave Theory Tech.*, vol. 48, pp. 11181126, July 2000.

[13] X. K. Kang, L. W. Li, M. S. Leong, and P. S. Kooi, "A Method of Moments study of SAR inside spheroidal human head and current distribution along handset wire antennas," *J. Of Electromagnetic Waves and Appl.*, vol. 15, No. 1, 6175, 2001.

[14] C.-M. Chen, J.-H. Tarng and J-M. Huang, "Electromagnetic absorption in human brains by using fast wavelet transform methods," *Proceedings of APMC2001*, Taipei, Taiwan, R.O.C.

[15] O. P. Gandhi, Q.-X. Li, and G. Kang, "Temperature rise for the human head for cellular telephones and for peak SARs prescribed in safety guidelines," *IEEE Trans. Microwave Theory Tech.*, vol. 49, pp. 16071613, Sept. 2001.

[16] A. Hirata, M. Morita, and T. Shiozawa, "Temperature increase in the human head due to a dipole antenna at microwave frequencies," *IEEE Trans. Electromagn. Compat.*, vol. 45, pp. 109117, Feb. 2003.

[17] A. Hirata and T. Shiozawa, "Correlation of maximum temperature increase and peak sar in the human head due to handset antennas," *IEEE Transactions On Microwave Theory And Techniques*, Vol. 51, No. 7, July 2003.

[18] Md. F. Ali, S. Mukherjee, S. Ray, "SAR analysis in human head model exposed to mobile base-station antenna for GSM-900 band," *Loughborough Antennas and Propagation Conference*. 2009.

[19] T. A. J. Mary, S. J. Priyadarshini, C. S. Ravichandran, R. George, "Analysis of SAR on human head modeling in metallic enclosures," 2010.

[20] S. Bri, S. Kassimi, M. Habibi, Ahmed Mamouni, "Specific Absorption Rate (SAR) distribution in the human head at Global System Mobil (GSM) frequencies," *European Journal of Scientific Research* ISSN 1450-216X Vol.49 No.4 (2011), pp. 590-600.

[21] L. Zhao and K.-L. Wu, "A hybrid NFM/MoM full-wave analysis of layered prolate head model exposed to handset antenna," *Progress In Electromagnetics Research*, Vol. 123, 205-225, 2012.

[22] M. N. O. Sadiku, Numerical Technique in Electromagnetics with Matlab 3rd ed. Texas, USA : CRC Press, 2009.

[23] Livesay, D. E., and Chen K.-M., "Electromagnetic field induced inside arbitrarily shaped biological bodies," *IEEE Transactions On Microwave Theory And Techniques* Vol. MTT-22, No. 12, 1974.

[24] Ulaby, F. T., Fundamentals of Applied Electromagnetics Media Ed., USA : Prentice-Hall, Inc., 2001.

[25] Rowell, C. and Lam, E. Y., "Mobile Antenna Design," *IEEE Antennas and Propagation Magazine* Vol. 54, No. 4, August 2012.

Barker Code Radar Simulation for Target Range Detection using Software Defined Radio

Jumail Soba, Achmad Munir, Andriyan Bayu Suksmono
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
jumailsoba_stelk@yahoo.com

***Abstract*— This paper present simulation of radar using Barker code signal to provide pulse compression. Barker code is the most well-known phase coding for pulse compression techniques. Pulse compression techniques has been known can provide solution for range resolution and detection range problem. Radar using Barker code simulated using software-defined radio, GNU Radio. Using GNU Radio give flexibility in operation, lower costs, faster in the realization of the design and easier to use. Radar signal processing for received signal performed on MATLAB. Radar used to detect range of the target which simulated by delaying the signal. This radar using Barker code length of 13. PRI $1.3x10^{-2}$ using for this radar, which give maximum unambiguous range 1950 km. Barker code signal of transmitter delayed 3000, 5000 and 10000 samples for simulation of target. Radar can detect the target as seen from the results of the signal processing done on matlab.**

Keywords—pulse compression; barker code; gnu radio

I. INTRODUCTION

Radar applications in human life is quite fundamental. Radar which initially appeared for the war, now has been widely used for civilian purposes. For example, for the purposes of weather forecasting and aircraft navigation. Radar is an electromagnetic system for the detection and location of objects. It operates by transmitting a particular type of waveform, a pulse-modulated sine wave for example, and detects the nature of the echo signal [1].

Barker code was first presented by R. H. Barker in 1953 for synchronization purposes in telecommunications. Beside widely used in radar technology, Barker code is also used in others telecommunications field. Barker code is used to !"#$"%&' ()*+"$,-./-0 /*1&(2$.*/ (/3()45 (6$+1 [2] and in wireless LAN IEEE 802.11 applications [3]. Barker code used in Radar application because Barker code can provide pulse compression function which can provide optimalization in both main parameter of radar, range resolution and detection range.

On the pulse compression radar, long pulse duration improves SNR and power levels emitted signal is not too high. While the output of the signal processing, the signal response is shorter so as to provide a good range resolution. By using the method of pulse compression, the two important parameters of the radar maximum detection range and range resolution can be optimized without sacrificing one of them. Where the maximum detection range of a radar is determined by the amount of energy emitted. The amount of energy rose by widening pulse. While the resolution of the distance up to shrink the width of the pulse using matching filter.

Radar with pulse compression method using Barker code signal can be implemented using software defined radio (SDR), SDR provides an advantage flexibility in operation, lower costs, faster in the realization of the design and easier to use. SDR can be combined with other software such as MATLAB to perform signal processing.

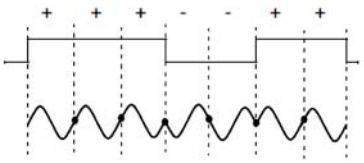
Implementations of radar using SDR has been done by many researcher, for example, implementation of FMCW radar for weather surveylance [4] and for passive radar implementation [5]. Implementation radar using SDR for pulse compression radar simulation using Barker code is still have not found in literature. In this paper, radar simulation using Barker code using the SDR is presented, and the signal processing is done with MATLAB software.

II. BARKER CODE RADAR

A. Barker Code

The most famous phase coding in pulse compression radar is Barker code. In the barker code, pulse divided into N sub-pulses. If the pulse width is !¡, then the width of subpulse !¡ ! !¡/!. Then, every sub-pulses conducted a phase shift at !! or !"#!. Phase !! (amplitude 1 Volt) can be characterized by a "1" or "+" and phase !"#! (amplitude -1 Volt) with "0" or "-". Selecton of phase !! and !"#! because both of this phase easily generated at the transmitter and the signal processor. Example of Barker code is shown in Fig. 1 [6].

This work is partially supported by the Ministry of Research and Technology, the Republic of Indonesia, under the Grant of Incentive Research Program 2013.



a.

Fig. 1. Binary phase coding with length 7

Code signal phase equation:

$$\text{!!}(\text{!}) \; \text{!} \; \text{!!}^{\text{!!}\,\text{!}\,\text{!}} \sum_{\text{!}\,\text{!}\,\text{!}}^{\text{!}} \text{!}_{\text{!}}(\text{!})\,\text{!}^{\text{!!}\,\text{!}} \quad (1)$$

Barker code of length N is denoted by $\text{!!}_{\text{!}}$. There are only 7 Barker code. List of Barker codes are shown in Table I.

In general, the autocorrelation function (output of the matched filter) of $\text{!}_{\text{!}}$ Barker code will be as wide as $\text{!}"\,\text{!}_{\text{!}}$. The width of the mainlobe $\text{!}\,\text{!}_{\text{!}}$, peak value N. There are (N-1)/2 sidelobes on each side of the mainlobe [6].

TABLE I. BARKER CODE

Code Symbol	Code Length	Code Elements	Side Lobe Reduction (dB)
$\text{!}_{\text{!}}$	2	+ - + +	6.0
$\text{!}_{\text{!}}$	3	+ + -	9.5
$\text{!}_{\text{!}}$	4	+ + - + + + + -	12.0
$\text{!}_{\text{!}}$	5	+ + + - +	14.0
$\text{!}_{\text{!}}$	7	+ + + - - + -	16.9
$\text{!}_{\text{!}\,\text{!}}$	11	+ + + - - - + - - + -	20.8
$\text{!}_{\text{!}"}$	13	+ + + + + - - + + - + - +	22.3

Maximum sidelobe reduction of Barker code of length 13 Barker code is -22.3 dB, is not enough for applications in Radar. Barker code can be combined to generate a longer code. Barker code can be combined to generate a longer code. B_M combined with B_N become B_{MN} [6] as shown in Fig. 2 .

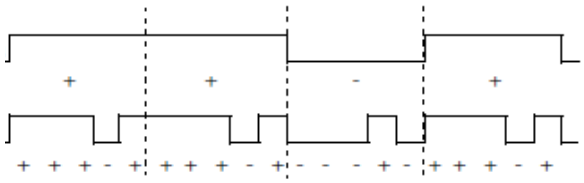


Fig. 2. Combination Barker code 5 dna 4, B_{54}

B. Software-Defined Radio

GNU Radio is a free software development toolkit that provides the signal processing runtime and processing blocks to implement software radios using readily-available, low-cost external RF hardware and commodity processors. It is widely used in hobbyist, academic and commercial environments to support wireless communications research as well as to implement real-world radio systems [7]. GNU Radio is licensed under the GNU General Public License (GPL) version 3. All of the code is copyright of the Free Software Foundation [7].

GNU Radio applications are primarily written using the Python programming language, while the supplied, performance-critical signal processing path is implemented in C++ using processor floating point extensions where available. Thus, the developer is able to implement real-time, high-throughput radio systems in a simple-to-use, rapid-application-development environment [7]. GNU Radio Companion is a GUI for GNU Radio that allows the user to design the system using schematic blocks.

III. METHOD AND DESIGN

Models of architectural simulation radar Barker code is shown in Fig. 3. Simulation of radar Barker code is done on GNU Radio. Signal of transmit and receive saved and then the signal processing is done in MATLAB.

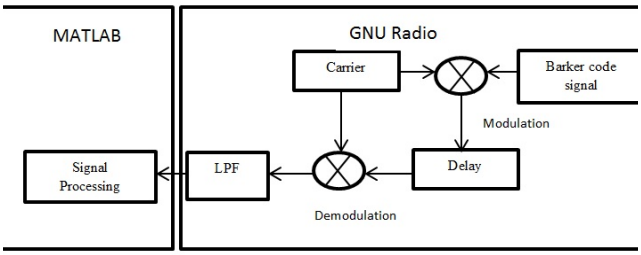


Fig. 3. The architecture of Barker code radar system based on SDR

The specification of the radar is to be simulated as shown in Table II.

TABLE II. BARKER CODE RADAR SPESIFICATION

Parameter	Value
Barker code length	13
Repeat	1000
Baseband frequency	1 kHz
Sampling Rate	1 MHz
Carrier frequency	100 kHz
Delay	3000, 5000, 10000 samples

Simulated Barker code radar on the GNU Radio Companion is done by making the scheme of the blocks that have been provided on the software GNU Radio Companion. Scheme of the blocks is as shown in Fig. 4 below :

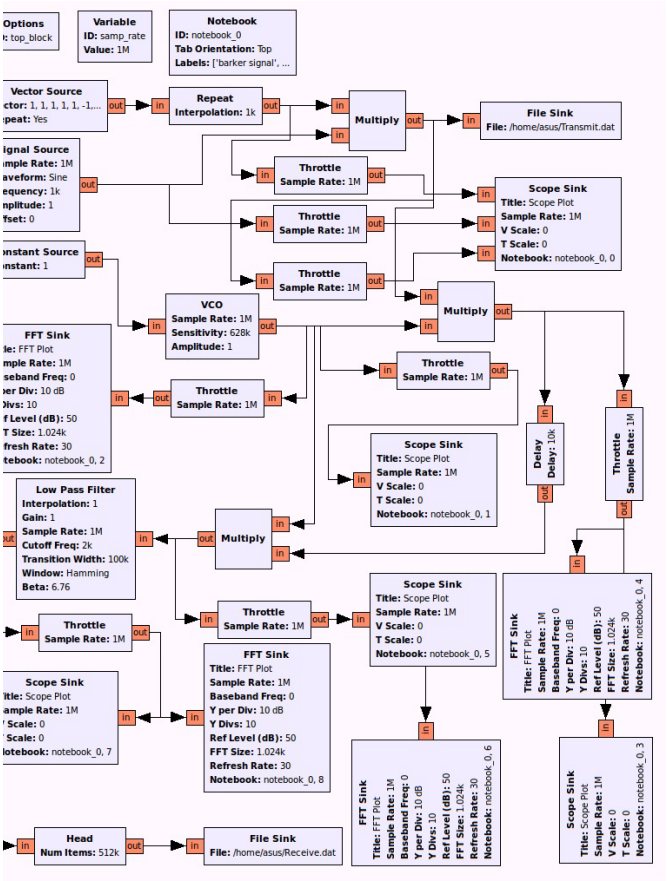


Fig. 4. Block Scheme for Barker code radar simulation on GNU Radio Companion

Values of parameters are adjusted in order to obtain the corresponding simulation results. The parameters of the blocks are given in Table III.

TABLE III. PARAMATER OF BLOCK ON GNU RADIO COMPANION

Block	Parameter
Vector Source	Output Type : Float Vector : 1, 1, 1, 1, 1, -1, -1, 1, 1, -1, 1, -1, 1 Repeat : Yes Vec Length : 1
Repeat	Type : Float Interpolation : 1000 Vector Lenth : 1
Signal Source	Output Type : Float Sample Rate : 1 MHz Waveform : Sine Frequency : 1 kHz Amplitude : 1 Offset : 0
Constant Source	Output Type : Float Constant : 1
VCO	Sample Rate : 1 MHz Sensitivity : 628000 Amplitude : 1
Multiply	IO Type : Float Number Inputs : 2 Vec Length : 1
Delay	Type : Float Delay : 3000 Vector : 1
Low Pass Filter	FIR Type : Float -> (Interpolating) Interpolation : 1 Gain : 1 Sampling Rate : 1 MHz Cutoff Frequency : 2 kHz Transition Width : 100 kHz Window : Hamming Beta : 6.76
Throttle	Type : Float Sample Rate : 1 MHz Vector Length : 1
Scope Sink	Type : Float Sample Rate : 1 MHz V Scale : 0 T Scale : 0 AC Couple : Off XY Mode : Off Num Inputs : 2
FFT Sink	Type : Float Sample Rate : 1 MHz Basebabd Frquency : 0 Y per Div : 20 dB Y Divs : 10 Ref Level (dB) : 50 FFT Size : 1024 Refresh Rate : 30 Peak Hold : Off Average : Off
Head	Type : Float Num items : 8000*64 Vec length : 1

IV. SIMULATON RESULT

A. *Simulation Result on GNU Radio*

Barker code signal generated by the Vector Source, Repeat, Signal Source and Multiply blocks as shown in the Fig. 4. Vector Repeat Souce and generate a signal box with a length of 13, according to the used Barker code length. Source Signal produces a sinusoidal signal which is then multiplied by the signal box on the Multiply block that generate Barker code signal with a length of 13, as shown in Fig. 5.

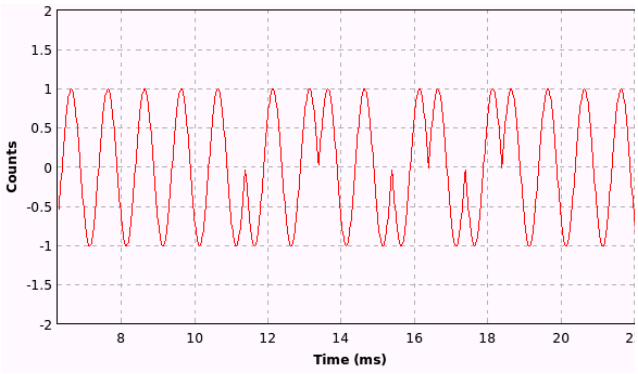
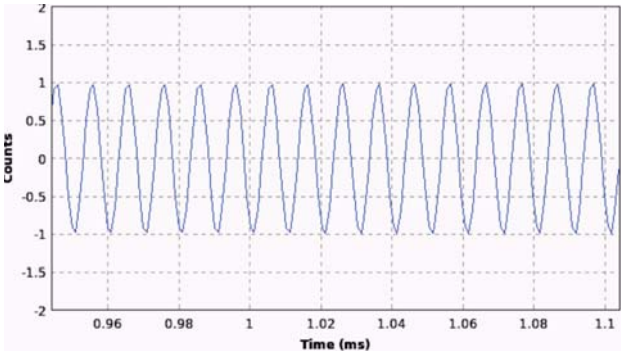


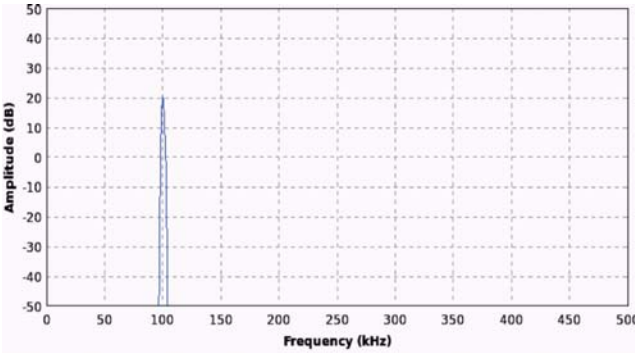
Fig. 5. Barker code signal

Carrier signal 100 kHz generated by the VCO and Constant Source blocks. Carrier signal is then modulated with the Barker code signal on Multiply block. FFT signal carrier

and the carrier signal shown in Fig. 6. Signal modulation results are shown in Fig. 7.



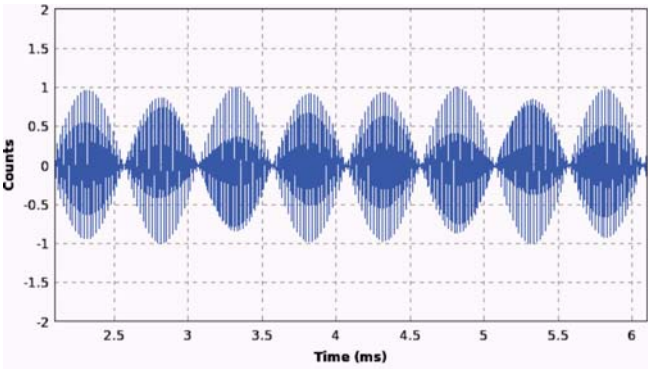
(a)



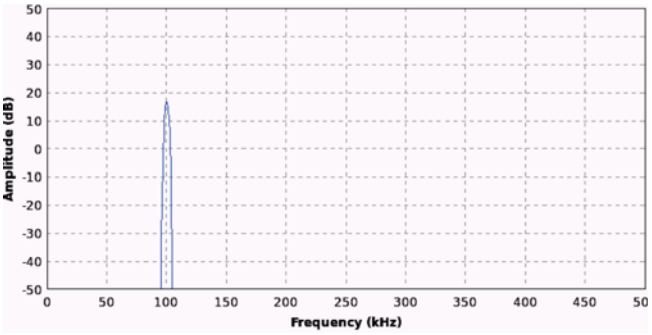
(b)

Fig. 6. Carrier signal(a) FFT of carrier signal (b)

Scope sink use to show carrier signal that have period 0.1 ms. FFT sink use to shoe the frequency at 100 kHz. This carrier signal will modulated with barker code signal on Multiply block. Modulated signal as shown in Fig. 7 (a).



(a)

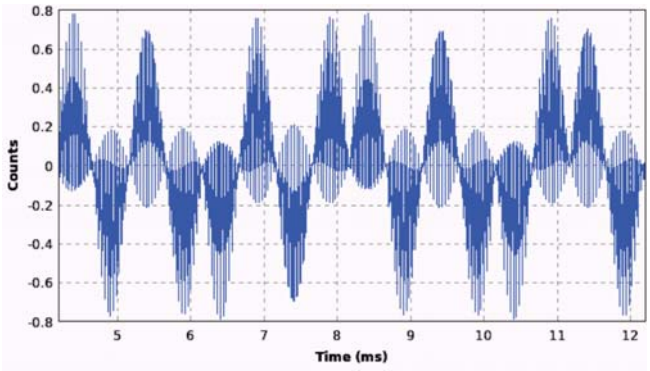


(b)

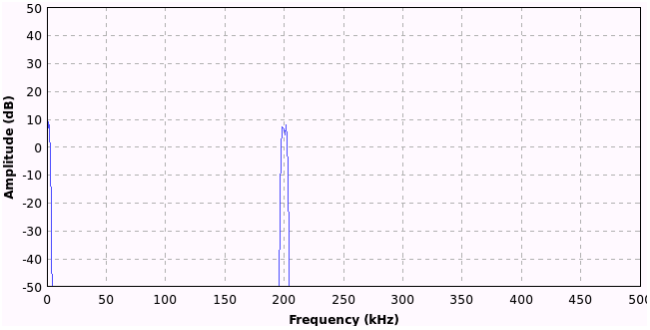
Fig. 7. Modulated signal (a) FFT of modulated signal (b)

Scope sink show that modulation type is amplitude modulation. Modulated signal between carrier signal 100 kHz and baseband frequency 1 kHz is in 99-101 kHz range.

Modulated signal then delayed alternately 3000, 5000 and 10000 samples on Delay block. Signal has delayed then demodulated by multiplying it by the same signal with the carrier signal (1 kHz). Signal of demodulation results are shown in Fig. 8.



(a)



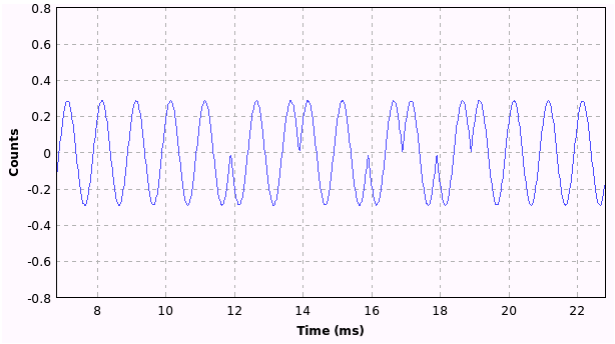
(b)

Fig. 8. Demodulated signal (a) FFT Demodulated signal (b)

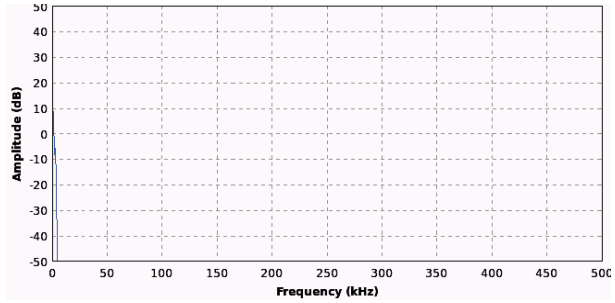
Scope sink show that after demodulated, signal have two frequency, at range 1 kHz and 199-201 kHz. Because

baseband frequency have low frequency, Low Pass Filter (LPF) used to block high frequency.

Demodulated signal results then inserted into the Low Pass Filter (LPF), the output of the LPF is stored in the File Sink and saved as Receive.dat file. Head use to limit the size of file. Output signal of LPF is shown in Fig. 9.



(a)



(b)

Fig. 9. Output signal of LPF (a) FFT of LPF output (b)

Scope show that after inserted in LPF, high frequency blocked and only low frequency passed. Output signal similar to barker code signal. Barker code signal (Fig. 5) and LPF output signal (Fig. 9 (a)) are stored in file sink for later signal processing performed on MATLAB software.

B. Simulation Result on Matlab

Send and receive signals that have been stored in a file in GNU Radio Companion and then processed in MATLAB. Basically signal processing in MATLAB done by is matched filtering between the received signal with transmit signal that has been complex conjugated using FFT as shown in Fig. 10 [1] .

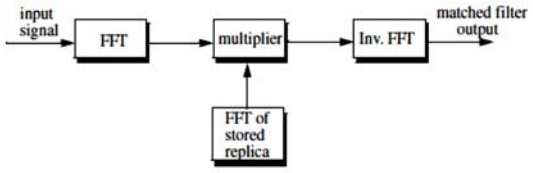
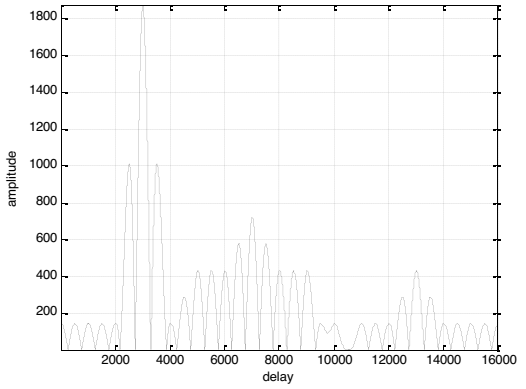
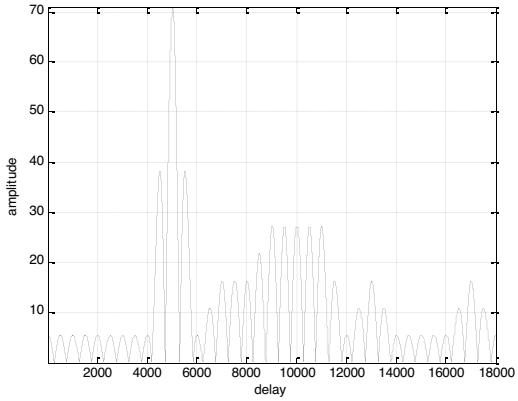


Fig. 10. Computing mached filter using FFT

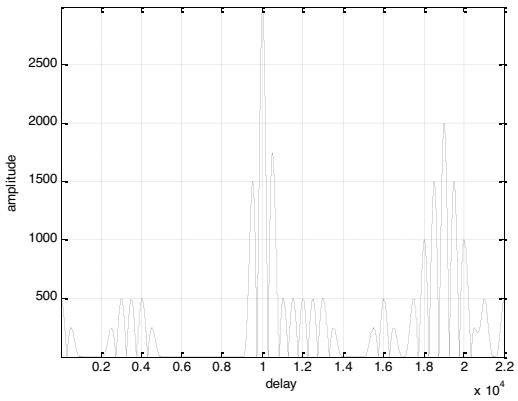
After signal processing on MATLAB, the result is shown in Fig. 11. Simulated result shown delayed signal 3000, 5000 and 10000 samples. !!



(a)



(b)



(c)

Fig. 11. Matched filtering result on MATLAB, delay 3000 samples (a), 5000 samples (b) and 10000 samples (c)

Match filtering result show that signal that delayed can be detected after processed in Matlab. High side lobe that shown

in the figure because the Barker code use sinusoidal waveform that have high sidelobe than other waveform for example chirp waveform. High sidelobe at 19000 in Fig. 11 (c) is ignored because this radar system have maximum unambiguos range 13000 in samples.

C. Analysys of Simulation Result

From radar spesification, maximum unambigous range can be calculated.

$$Ru = \frac{! \; !!! \; " \; \#}{!} \quad (2)$$

Maximum Unambiguous Range calculation :
Barker code length 13,
Sampling rate = 1 MHz,
Baseband Frequency = 1 kHz,
Repeat interpolation = 1000

dt= 1/Sampling rate = $1/10^6 = 10^{-6}$ second
PRI = 13 x 1000 x $10^{-6} = 1.3x10^{-2}$ second
Ru (sample) = $1.3x10^{-2}/10^{-6} = 13000$ samples
Ru = c x PRI/2 = $3x10^8$ x $1.3x10^{-2}$ /2 = 1950 km

From calculation we get maximum unambiguos range in samples is 13000 in samples and 1950 in km. Pulse compression function also shown in simulation result, signal after matched filter process is narrower and this is give better range resolution.

Using the formula (3) [1], range of the target can be calculated:

$$R = \frac{! \; !! \; !}{!} \quad (3)$$

For delay 3000 samples, coverted to time $3x10^{-3}$ s, the range is 450 km. For delay 5000 samples, range is 750km and for delay 10000 samples, range is 1500 km.

V. CONCLUSIONS

From these simulations it can be concluded that Radar Simulation with Barker code code length of 13 with maximum unambiguous range 1950 km using Software defined radio, GNU Radio Companion, and Signal processing using MATLAB can be implemented. This Barker code radar simulation can detect range of the target with different range as simulated with delaying the signal. Future works for this paper is this simulation can implemented with using USRP hardware and antenna. Sidelobe reduction processing can also used in signal processing in Matlab to reduce the high sidelobe

REFERENCES

[1] Skolnik, Merrill I., Introduction to Radar Systems, 1980 by McGRAW-HILL BOOK COMPANY

[2] 7$&89 8(; *< 8=8(7>?< 8=8@6#&(! "#$"%&< &".(/30 *+"$,-./-0 /*1&2$.*/(/3045 6$+1(A1*"+(B-C*.(7$?D&?(E/'&()&F>&"%&G8Proceedings of the 40th European Microwave Conference, 2010

[3] Konig, S., Schmidt, M., Hoene, C., "Precise Time of Flight measurements in IEEE 802.11 Networks by Cross-Correlating the Sampled Signal with a Continuous Barker Code", Mobile Adhoc and Sensor System (MASS), 2010 IEEE 7th International Conference, 2010

[4] Prabaswara, A., Munir, A., Suksmono, Andriyan B., "GNU Radio Based Software-Defined FMCW Radar for Weather Surveillance Application", The 6th International Conference on Telecommunication Systems, Services, and Applications 2011, 2011

[5] Heunis, S., Paichard, Y., Inggs, M., "Passive Radar using a Software-Defined Radio Platform and Opensource Software Tools", Radar Conference (RADAR), 2011 IEEE, 2011

[6] Mahafza, Bassem R., Radar Systems & Analysis and Design using Matlab, 2000 by Chapman & Hall/CRC

[7] GNU Radio – WikiStart – gnuradio.org http://gnuradio.org/redmine/projects/gnuradio/wiki, accessed on Mei, 30th 2012

Designing Cross-Coupled Bandpass Filters with Transmission Zeros in Lossy Microstrip

Mudrik Alaydrus, Dian Widiastuti and Teguh Yulianto
Department of Electrical Engineering
Universitas Mercu Buana, Jakarta, Indonesia
Email: mudrikalaydrus@yahoo.com

***Abstract*—Bandpass filters play a significant role in many wireless communication systems. The filters pass desired signals and reject unwanted one. In this paper, we design a four-pole bandpass filter using a rigorous computer simulation. The resonators used are square open-loop resonators. To get a sharper selectivity around the pass band, transmission zeros though cross-coupling between resonators are introduced. In simulation, we compare the results for relative permittivity 4.4 and 4.9, and for tangent loss 0.025 (lossy) and 0.0 (lossless). The simulation shows, the required specifications are fulfilled for lossless case, whereas for lossy case, an insertion loss of 8.3 dB observed. For verification, the bandpass filter is built in FR4. In this lossy microstrip, a measurement shows, we get an insertion loss of about 6.64 dB and a bandwidth of about 120 MHz.**

Keywords—bandpass filter, coupling coefficient, cross-coupled resonators, square open–loop, transmission zeros

I. INTRODUCTION

The recent advanced digital communications systems require efficient use of the frequency spectrum. The use of this scarce expensive resource must be controlled strictly otherwise significant interference can occur, which can deteriorate the overall system performances. Bandpass filters are designed for this purpose. Bandpass filters pass desired signals from other unwanted signals [1]. To get sharp filtering characteristics, in traditional filter approximations, such as Butterworth and Chebychev realizations, we need many resonators, which leads to big filter dimensions. The introduction of transmission zeros enhances the selectivity near the pass band [2].

Hong in [3] introduced a simple square open-loop resonator as based for building the filter. The filter designed around the frequency 2.46 GHz with a fractional bandwidth of 0.04 and implemented in low-loss material. Based on given coupling coefficients, the paper determined the separations between the resonators using some empirically found formulas. Lee et al [4], based on Hong's work [3], designed a bandpass filter for IMT-2000 around 2 GHz, they used also low-loss microstrip lines. An other resonator structure is proposed by [5], a novel bandpass filter with sharp attenuations and wide stopband is developed through the combined use of composite resonators and stepped impedance resonators (SIRs). Parallel coupled bandpass filters are developed for several frequency regions, such as around 5.75 GHz [6], and around 3.2 GHz [7]. The comparisons between simulation and measurement results, as well as with the specifications given before, are very good. In all the publications above, the microstrip used is based on expensive material like TMM10, or RT/Duroid substrate, which have a tangent delta smaller than 0.003. This small value of tangent delta belongs to materials with small loss.

In this work, we develop a bandpass filter around the frequency 2.35 GHz, which is internationally allocated to amateur radio and amateur satellite use on a secondary basis. Another potential applications would be radio location, mobile and fixed communications [8]. We use the same structure as described in [3], the square open-loop resonator. Firstly, the filter structure is simulated numerically by a commercial software package, Sonnet v13.56 [9]. In fabricating the filter, a printed circuit board (PCB) based on FR4 material is used. The relative permittivity of FR4 lies between 4.4 and 4.9 and the tangent delta of about 0.025. The thickness of the PCB used is 1.6 mm. Because FR4 has a tangent delta about 10 times higher that RT/Duroid or TMM10, in this paper we call this microstrip as lossy microstrip. In this research we observe the effects of lossy material to the coupling coefficient and to the overall filter performances. Moreover, we study the effect of the uncertain value of the relative permittivity of FR4 to the filter performances.

II. CHEBYCHEV APPROXIMATION WITH TRANSMISSION ZEROS AND CALCULATING THE COUPLING COEFFICIENTS

In this paper, we try to design a bandpass filter, whose spesifications should somewhat conform with a filter produced by AMTI, the C0423501 [10]. The filter should work at the interval 2.3 GHz and 2.4 GHz, with maximal insertion loss of 1.5 dB. The voltage standing wave ratio is 2.0:1, the ripple and flatness in the pass band are smaller than 1 dB, and the rejection at 2.2 GHz and 2.5 GHz is minimal 20 dB. We do not consider another additional specifications like temperature compatibility and the dimensions of the filter.
With the interval given above, the midband frequency becomes $f_o = \sqrt{2.3 \cdot 2.4} GHz = 2.3495 GHz$, and the fractional bandwidth $FBW = BW/f_o = 0.04256$. In designing the filter, the filter order is set to n=4 (four resonators), so that the dimension of the filter can be kept small. The VSWR of 2.0:1 leads to a reflection factor of -9.54 dB, which generally can be achieved by any filter configuration. To fulfill the requirement of the band rejection at 2.2 GHz and 2.5 GHz, we introduce a pair of transmission zeros at about 2.26 GHz and 2.44 GHz, so that we have $\Omega_a = (2.44/2.3495 - 2.3495/2.44)/0.04256 = 1.78$.

Cameron [11] gives an efficient algorithm to synthesis a generalized Chebychev filter by introducing a coupling matrix. However, in this paper we follow the procedure described in [1]. In [1] we can find component values for n =4 and RL = -20 dB. We choose the value $\Omega_a = 1.8$, so we have in page

264 in [1] the following component values $g_1 = 0.95974$, $g_2 = 1.42192$ $J_1 = -0.21083$ and $J_2 = 1.11769$.

With these values, the coupling coefficient $M_{i,i+1}$ and the external quality factors Q_e can be calculated

$$M_{1,2} = M_{3,4} = \frac{FBW}{\sqrt{g_1 g_2}} = 0.03643 \quad (1)$$

$$M_{2,3} = \frac{FBW \cdot J_2}{g_2} = 0.03346 \quad (2)$$

$$M_{1,4} = \frac{FBW \cdot J_1}{g_1} = -0.00935 \quad (3)$$

$$Q_{e,i} = Q_{e,o} = \frac{g_1}{FBW} = 22.55 \quad (4)$$

III. Realization the Coupling in Microstrip Resonators

The feeding line is defined to have the standard wave impedance of 50 Ω, leads to strip width of about 3 mm (for relative permittivity of 4.9), and for the resonator is chosen to the wave impedance of 60 Ω leading to resonator line width of about 2 mm. The line width of 2 mm gives an effective relative permittivity of $\varepsilon_{r,eff} = 3.366$, and the guided half wavelength inside the microstrip line at the mid frequency 2.35 GHz becomes about 35 mm.

We use the similar resonator as in [3], the square open-loop resonator with the total physical length of 35.5 mm, a gap length of 1.3 mm, and line width of 2 mm.

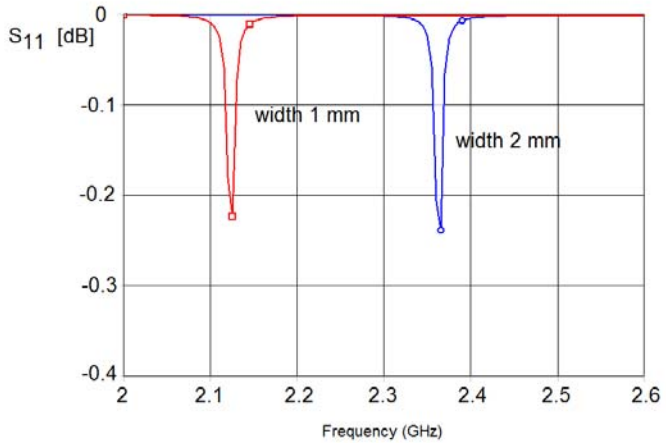


Fig. 1. resonant frequency for different line width of the resonator total physical length 35.5 mm.

Fig.1 shows comparison the position of the resonance between different line width of the resonators. In designing we base on the line width of 2 mm, the figure gives the position of the resonance at about 2.36 GHz. A smaller line width of 1 mm gives smaller effective relative permittivity. The reason is as follows. The total physical length of the resonator used here, 35.5 mm, leads to higher guided wave length. The resonance happens consequently at a lower frequency, at about 2.13 GHz.

According to the coupling coefficients obtained in the previous section, we propose to use the schematic given in fig. 2. The couplings between resonator 1 and 2 and between 3 and 4 are realized as mixed coupling (combination electric and magnetic couplings), whereas the coupling between resonator 2 and 3 is magnetically and between 1 and 4 is electrically (negative coupling value). The couplings between resonators 1 and 2, 2 and 3 and 3 and 4 are the direct couplings, whereas the coupling between resonators 1 and 4 is the cross coupling, which realizes the transmission zeros.

In the next section, we will vary the separation between resonator 1 and 2, s_1, to gain the coupling coefficient $M_{1,2} = 0.03643$, also for other coupling coefficients.

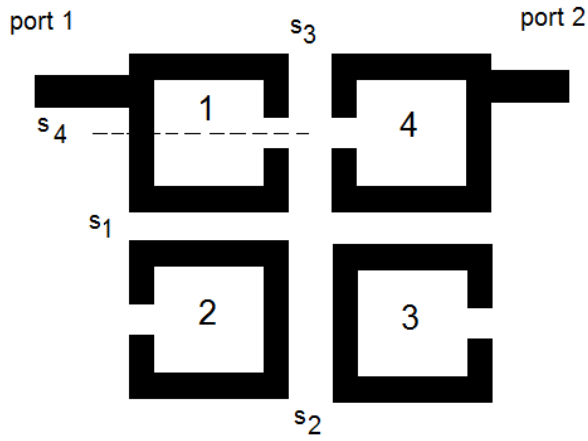


Fig. 2. Schematic for realizing the couplings

IV. Results

A. Coupling Coefficients k as Function of Separation s

From the physical consideration, that small distances will cause strong coupling and large distances lead to weak coupling, we begin the observation with very small distance (s_1=0.1 mm) and very large distance (s_1=3 mm). At resonance, the transmission factor (S_{21}) of a resonator will have an optimum value, and if two resonators are coupled mutually, there are two resonant points, the lower frequency f_1 and the higher f_2. From these resonant frequencies, the coupling coefficient can be calculated to

$$k = \frac{f_2^2 - f_1^2}{f_2^2 + f_1^2}. \quad (5)$$

Fig. 3 gives the transmission factor from port 1 to port 2 for two different distances mentioned above for lossy and lossless case. In simulating the lossy case, the tangent delta is set to 0.025, whereas for lossless case, the value is set to 0.0. For the strong coupling case, the resonant frequencies are located far away from each other, which gives big coupling coefficient value. For the lossless case we see the two peaks very well, and for the lossy case we can still recognize the maximal value of S_{21}. For weak coupling case, the resonant frequencies get closer, and for lossy case they degenerate to each other, we cannot resolve them, and there is no possibility to calculate the coupling coefficient.

Fig. 4 shows the comparison of calculated mixed coupling coefficient for lossless case (solid line) and lossy case (circle). Due to the similarity of the results, for smaller coupling we

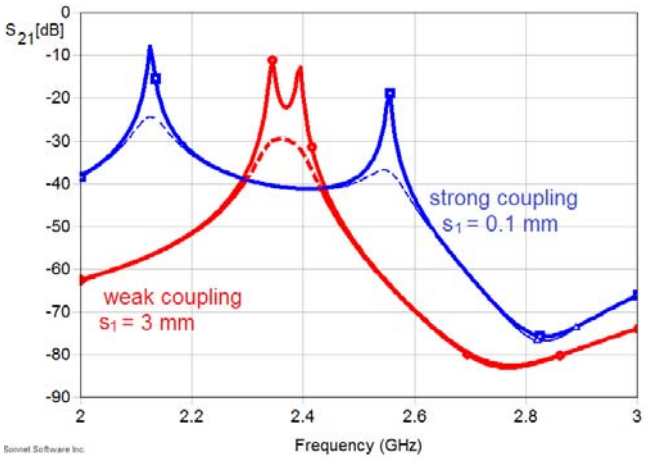


Fig. 3. Strong and weak coupling in mixed coupling between two resonators in lossless (solid lines) and lossy case (dashed lines).

prefer to use the lossless case for better resolution, which consequently leads to more accurate results in computing the coupling coefficient. In this figure, we learn to calculate the coupling coefficient, it is enough to consider the lossless version of the problem. From fig. 4, we can determine the distance s_1, which yield the coupling coefficient $M_{1,2}$=0.03643, namely $s_1 = 1.9mm$. In a similar fashion, we calculate the distances s_2=2.5 mm and s_3=2.9 mm.

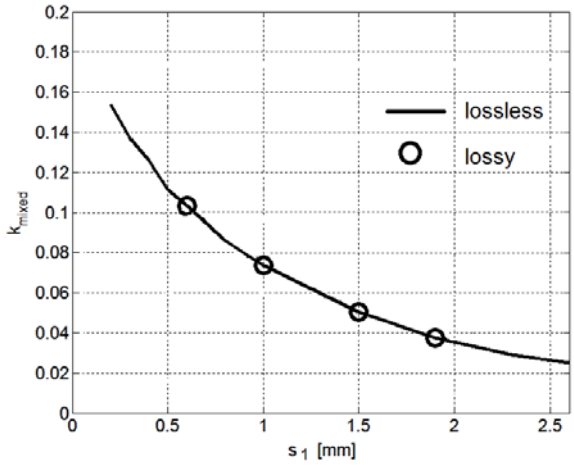


Fig. 4. Coupling coefficient for mixed coupling.

The distance from the feed point to the symmetry line of the resonator s_4 will determine the external quality factor of the resonator. For the calculation of the external quality factor we use just a resonator, and calculate again the transmission factor S_{21}. At the resonance, again the transmission factor have a maximum, so that we get the resonant frequency of the system. By getting the frequency interval of the 3-dB line, we get the bandwidth of the resonator circuit. It is clear, that a system having large bandwidth will have small quality factor, and narrow bandwidth is the effect of a system with higher quality factor. So, by dividing the resonant frequency and bandwidth to each other we get the external quality factor of the resonator fed by this microstrip line. The iterative calculation gives the value s_4 = 2.65 mm.

The complete schematic of the bandpass filter designed is depicted in fig. 5 on the left side.

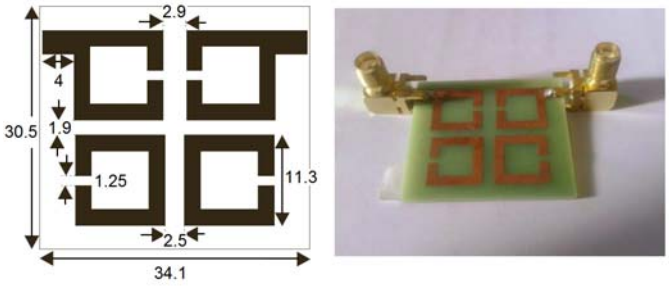


Fig. 5. left: Schematic (all dimensions in mm), right: the prototype

B. Simulation and Measurement

The schematic in fig. 5 on the left side is the structure under consideration. A full-wave computation with the software Sonnet gives a very good simulation result. Fig. 6 shows the transmission factor in the frequency interval of 2 GHz and 3 GHz. The lossless FR4 with the relative permittivity of 4.9 gives an insertion loss smaller than 1.5 dB in the intervall 2.32 GHz and 2.415 GHz. The pass band is located at around 2.36 GHz. A rejection of more than 28 dB at 2.2 GHz and about 19.7 dB at 2.5 GHz. The reflection factor for this case is smaller than -10 dB in the same frequency interval (fig. 7).

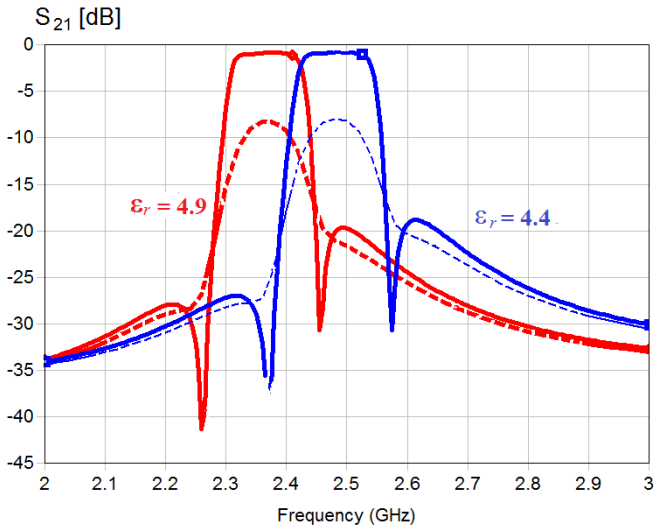


Fig. 6. Transmission factor for lossless (solid) and lossy (dashed) FR4 with different relative permittivity.

Furthermore, we see transmission zeros at about 2.27 GHz and at 2.45 GHz. All requirements given are fulfilled very well.

The condition is rather different for lossy case (dashed curves). A minimal insertion loss of about 8.3 dB is observed. However, other specifications are still fulfilled.

In other occasion, we consider, what will happen, if our designed filter is not embedded in an FR4 of relative permittivity of 4.9 but rather 4.4. In this case is certainly our filter is electrically smaller, the consequence is, the pass band

will be shifted in higher frequencies. A simple calculation gives us a prediction of a new mid frequency, 2.36 GHz $\cdot\sqrt{4.9/4.4} = 2.4905$ GHz. A computer simulation based on the software Sonnet verifies this result, as depicted in fig. 6, the pass region is now around 2.49 GHz, as predicted before.

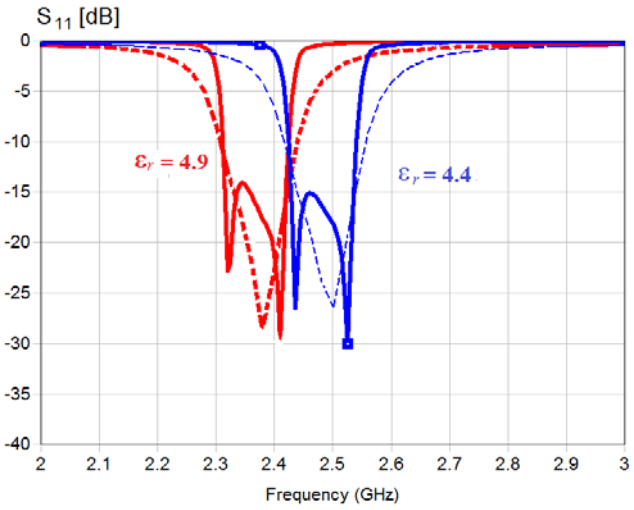


Fig. 7. Reflexion factor for lossless (solid) and lossy (dashed) FR4 with different relative permittivity.

In order to verify the simulation with measurements, a prototype is built (fig. 5 on the right side). The measurement is performed with a vector network analyzer ZVL13. The transmission factor and reflection factor is depicted in fig. 8. The mid frequency is located at about 2.45 GHz with insertion loss of about 6.64 dB. 3 dB below this value is the region between the frequency 2.41 GHz and 2.53 GHz, or the filter has the bandwidth of about 120 MHz. Transmission zeros around the pass band are not observed. The shift of the frequency response of the filter compared to the simulation result is probably due to, we simulate and measure with different values of relative permittivity. Probably, the FR4 used has a relative permittivity of about 4.4, so that the measurement results conforms with the simulation results with this relative permittivity value. A smaller insertion loss obtained by measurements could have the reason, that we in simulation used higher tangent delta, than we have in reality. From the comparison between simulation and measurement, we have the opportunity to measure the relative permittivity and the tangent delta of FR4.

V. Conclusion

Designing a bandpass filter around the frequency 2.35 GHz with a pair of transmission zeros was verified by a simulation with a computer. All specifications are fulfilled, if we use lossless FR4. The simulation of lossy FR4 shows a degradation of the insertion loss to the value of 8.3 dB. A measurement of the filter revealed, that the frequency characteristics of the filter was shifted to higher region. A possible reason is we have in reality an FR4 with the relative permittivity of about 4.4. The insertion loss obtained by measurement is about 6.64 dB. In this research, we learnt two things, firstly, to calculate the coupling coefficient is it enough to simulate the lossless microstrip, secondly, through comparison between simulation

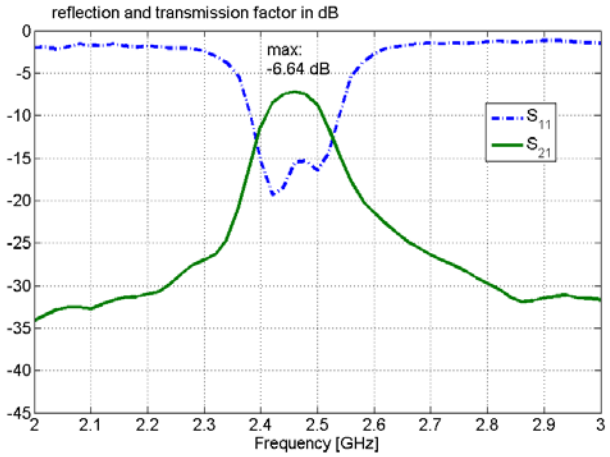


Fig. 8. Measurement Results

and measurement, we can measure the relative permittivity of the material.

References

[1] Hong, Jia-Seng, *Microstrip Filters for RF/Microwave Applications*, Second Edition. New Jersey : John Wiley & Sons, 2011.

[2] R.J.Wenzel, Understanding Transmission Zero Movement In Cross-Coupled Filters, *Microwave Symposium Digest, 2003 IEEE MTT-S International*, pp.1459 - 1462.

[3] J.S. Hong, and M.J. Lancaster, Couplings of Microstrip Square Open-Loop Resonators for Cross-Coupled Planar Microwave Filters, *IEEE Trans. Microwave Theory Tech.*, Vol. 44, No 12, Dec. 1996, pp. 2099-2109.

[4] Y.H.Lee, W.H. Kwon, B.H. Koo, C.S. Lee, A study on a RF filter of IMT-2000 terminal using coupled rectangular microstrip loop resonators, *Asia-pacific Microwave Conference 1999*, pp.295-298.

[5] H. Miki, Z. Ma, and Y. Kobayashi, A Novel bandpass filter with sharp attenuations and wide stopband developed through the combined use of composite resonators and stepped impedance resonators, *Asia-pacific Microwave Conference 2006*, pp. 1683-1686.

[6] O.A.R. Ibrahim, I.M. Selamat, M. Samingan, M. Aziz, A. Halim, 5.75 GHz microstrip bandpass filter for ISM band, *Applied Electromagnetics, 2007 Asia-Pacific Conf. on*, Dec. 2007, pp. 1-5.

[7] M. Alaydrus, Designing Microstrip Bandpass Filter at 3.2 GHz, *International Journal on Electrical Engineering and Informatics*, Volume 2, Number 2, 2010, pp.71-83.

[8] Federal Communications Commission, www.fcc.gov (verified on July 14, 2013)

[9] www.sonnetsoftware.com (verified on June 15, 2013)

[10] www.diplexer.com (verified on June 15, 2013)

[11] R. J. Cameron, General coupling matrix synthesis methods for chebyshev filtering functions, *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 4, pp. 433442, Apr. 1999.

Electrical Model of Two Element Aperture Coupled Cylindrical Dielectric Resonator Antenna Array

A. A. Baba, M. A. Zakariya*, Z. Baharudin, M. H. Md Khir, S. M. Ali and J. J. Adz.
Electrical and Electronics Engineering Department
Universiti Teknologi PETRONAS, Malaysia
*mazman_zakariya@petronas.com.my

***Abstract*—In this paper, electrical models of single element cylindrical dielectric resonator antenna (CDRA) and two element CDRA array made of CCTO material are presented. The 50 ▵ microstrip transmission line is used to excite the CDRAs through coupling slots etched on the ground plane. The electric models of single element CDRA and two element CDRA array are designed by using Advanced Design System (ADS). The electrical circuits are used to validate the CST design feasibility. The validity of RLC model is verified by comparing the return loss of the ADS model against those obtained through computer simulation technology (CST) and fabricated prototype. The results obtained through simulated (CST and ADS) designs and fabricated prototype of single element CDRA and two element CDRA array are in good agreement.**

Keywords—Cylindrical dielectric resonator (CDRA); microstrip line; electrical model.

I. INTRODUCTION

With the advent of small size modern communication devices, a small size high gain antenna becomes the requirement of communication industry. Dielectric resonator antenna (DRA) is the best alternative to microstrip patch antenna technology at millimetre wave frequencies. DRA attained the attraction of researchers and communication industry due to its many attractive features such as low conductor and surface wave losses at millimetre wave frequencies as compared to patch antennas [1], can be fabricated into different shapes [2], has high efficiency (> 95%) [3], can cover the frequency band from 0.7–35 GHz [4] and the performance of DRAs is minimally affected by the presence of nearby noisy objects (such as human bodies) [5]. DRA was first introduced by Long in 1983 [6]. Before 80's dielectric resonators (DRs) of different shapes with high relative permittivity ε_r > 20 have been used in various applications of microwave circuits [7].

The author would like to thank Universiti Teknologi PETRONAS in particularly Research and Invoation office (RIO) in providing short term internal research fund for this work.

Different shapes of dielectric resonators are found in the literature such as rectangular, annular and cylindrical. The cylindrical DRA become the attraction of researchers and communication industry due to its many attractive features. The CDRA is more directional as compared to annular and rectangular DRA [8]. Dielectric material can be easily molded into cylindrical shape by using commercially available rod [9] and it has less edges and simple field structure as compared to rectangular DR [10].

Different modes of DRA can be excited by selecting the slot position under DRA. It is mentioned in the literature that by placing slot in the middle of the DR hybrid mode can be excited and by moving away from DR center *TE* and *TM* modes are excited [7]. Various coupling techniques have been used to excite the DRA such as aperture couple microstrip transmission line [11], probe feed [12] and dielectric image line [13]. The microstrip transmission line is easy to fabricate and has a low profile as compared to probe feed and dielectric image line. From the literature it is also proved that rectangular shaped coupling slot is better in performance as compared to the circular shape aperture slot [14].

In this paper, electrical models and fabricated prototypes of single element and two elements aperture coupled CDRA array are presented for WLAN applications.

II. DESIGN METHODOLOGY

The single element CDRA and two elements CDRA array is design for IEEE 802.11a WLAN applications working at 5.0 GHz band. The proposed geometries of a single element CDRA and two element CDRA array are depicted in Fig. 1 and Fig. 2 respectively. The dimensional attributes of proposed antennas are tabulated in Table 1. The CDRAs of diameter

$2a$ = 15.5 mm, height $h = h_1 = h_2$ = 3.0 mm and permittivity ε_r = 55 are used for antenna design. The coupling slots placed at a distance of $0.9\lambda_g$ are used to excite the CDRAs through 50 ▵ transmission line. The transmission line is placed at the back of the FR4 substrate having a thickness and permittivity of 1.565 mm and 4.9 respectively.

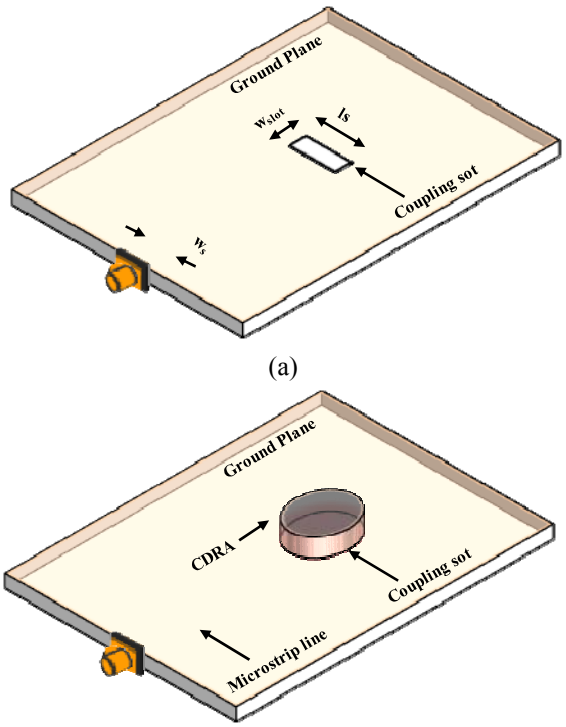


Fig. 1. Geometry of aperture coupled CDRA.

Table 1. Dimension of aperture coupled CDRA

Parameters	Length (mm)
$2a$	15.5
ε_r	55
h, h_1, h_2	3.0
w_s	2.6
w_{slot}, w_1, w_2	4.0
l_{slot}	14.0
L_1, L_2	20.0
L	6.67

The electrical models of aperture coupled single element CDRA and two element CDRA array are designed to validate the CST design as depicted in Fig. 3 and Fig 4 respectively. The *A, B* and *C* block represent the RLC blocks of microstrip line, aperture slot and DRA respectively. The Equations used to calculate the resonant resistance R_r, inductance L_r and capacitance C_r for the DR are given as [15]

$$R_r = \frac{2n^2 z_0 s_{11}}{1 - s_{11}} \quad (1)$$

$$C_r = \frac{Q_0}{\omega_0 R_r} \quad (2)$$

$$L_r = \frac{1}{C_r \omega_0^2} \quad (3)$$

where

S_{11} is the reflection coefficient.

Z_o is the characteristic impedance.

Q_o is the quality factor.

n is the coupling magnitude between excitation source and dielectric resonator.

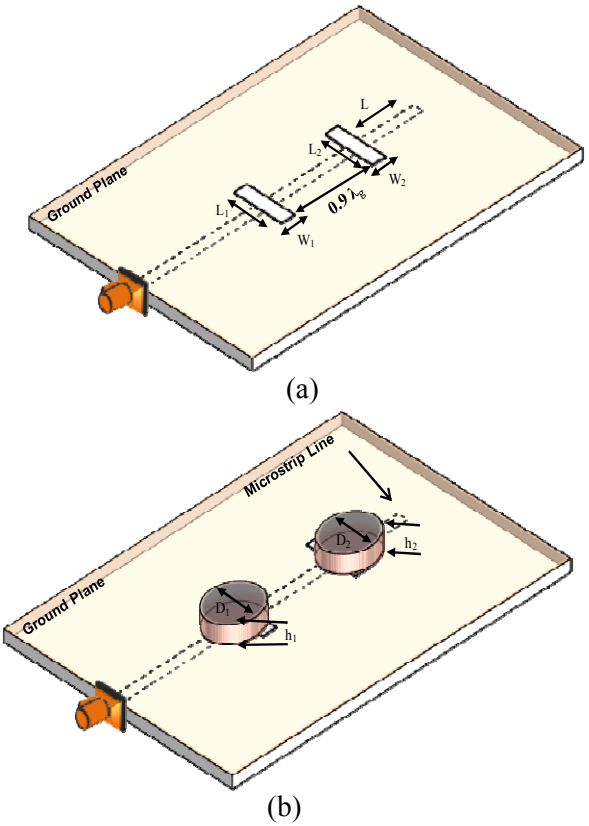


Fig. 2. Geometry of two elements aperture coupled CDRA array.

The value of R_r needs to be selected carefully because the value of L_r and C_r depends upon the R_r value. The Equation used to calculate the coupling slot impedance is given in [16]

$$Z_{slot} = Z_c \frac{2R}{1-R} + jZ_c \cot(\beta_f L_t) \quad (4)$$

where

Z_c is the characteristic impedance of the transmission line.

R is the voltage reflection coefficient.

β_f is the propagation constant.

L_t is the stub length.

The Equations used to calculate the input impedance of the transmission line is given in [17]

$$G_{rm} = \frac{160\pi^2 h^2}{Z_{cm}^2 \lambda_0^2 \epsilon_{cm}} \quad (5)$$

$$B_m = \omega C_l, \qquad C_l = \frac{l_{eq}\sqrt{\epsilon_{cm}}}{cZ_{cm}} \quad (6)$$

where

h is the substrate height.

Z_{cm} is the characteristic impedance of the microstrip.

E_{cm} is the effective dielectric constant.

l_{eq} is the equivalent extra length of microstrip line.

c is the velocity of light.

The fabricated prototypes of the proposed antennas are depicted in Fig. 5 and Fig. 6.

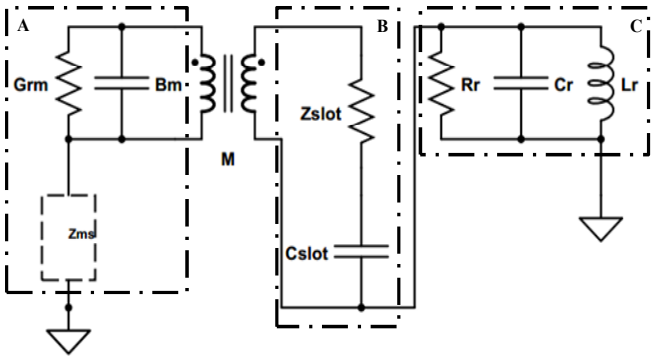


Fig. 3. Electrical model of a single element dielectric resonator antenna.

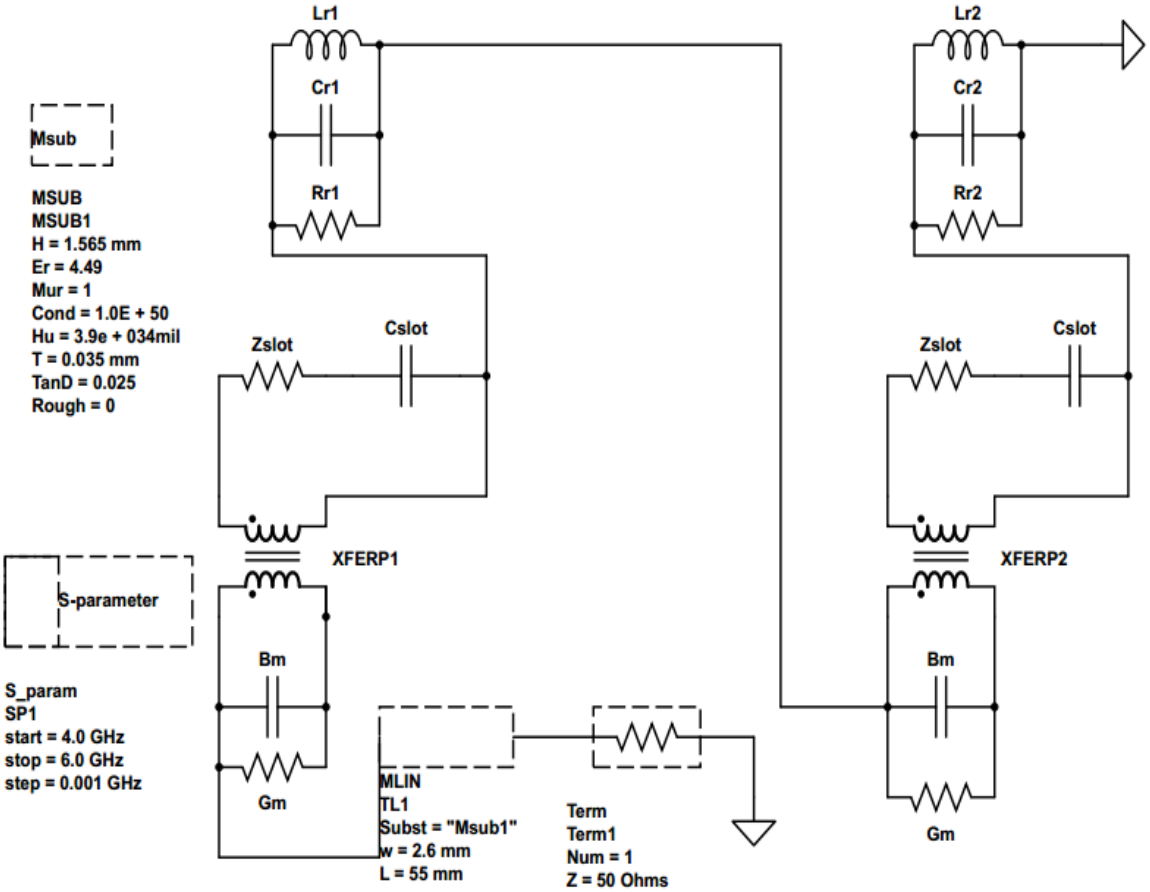


Fig. 4. Electrical model of a two element dielectric resonator antenna array..

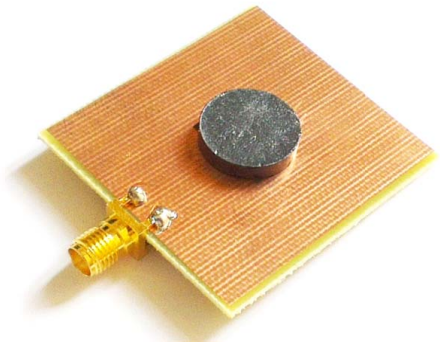


Fig. 5. Fabricated geometry of aperture coupled CDRA.

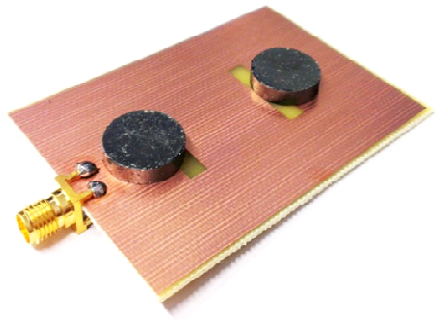


Fig. 6. Fabricated geometry of two element CDRA array.

III. Results and Discussion

The fabricated prototypes of single element and two elements aperture coupled CDRA array designed for 802.11a applications are depicted in Fig. 5 and Fig. 6 respectively. The comparison between the return loss (S_{11}) results of single element CDRA and two element CDRA array is depicted in Fig. 7. The corresponding achieved bandwidths are 0.2 GHz and 0.35 GHz respectively. The simulated (CST and ADS) and measured impedance bandwidths for singe element CDRA are 0.2 GHz (5.15-5.35 GHz), 0.2 GHz (5.15-5.35 GHz) and 0.23 GHz (5.14-5.37 GHz) respectively as depicted in Fig. 8. The simulated (CST and ADS) and measured return loss results of two element CDRA array is depicted in Fig. 9. The corresponding impedance bandwidths of two element antenna array are 0.35 GHz (5.1-5.45 GHz), 0.375 GHz (5.15-5.525 GHz) and 0.51 (4.95-5.46 GHz) respectively. Fig. 9 shows that return loss is less than -10 dB for both simulated and measured results. These values show that the antenna works at 5.15–5.35 GHz which is the requirement for IEEE 802.11a WLAN application. The shift in the simulated and measured results is due to the fabrication error (surface of the CDRs is not smooth which produce the air gap between the CDRs and ground plane). In addition to this, the main reason of shift in the results of electrical model and fabricated prototype is due the approximation taken while calculating the electrical design parameters.

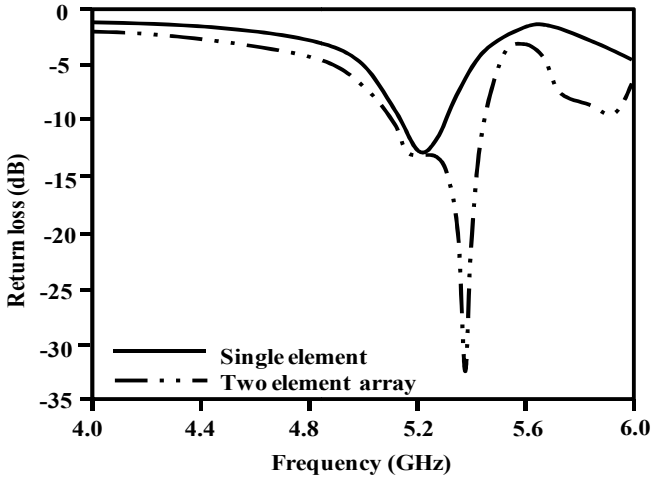


Fig. 7. Comparison between the return loss of singe element and two element CDRA array.

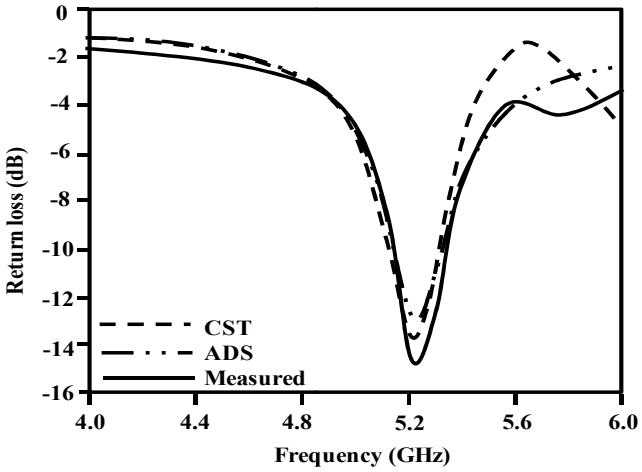


Fig. 8. Simulated (CST and ADS) and measured return loss of single element CDRA.

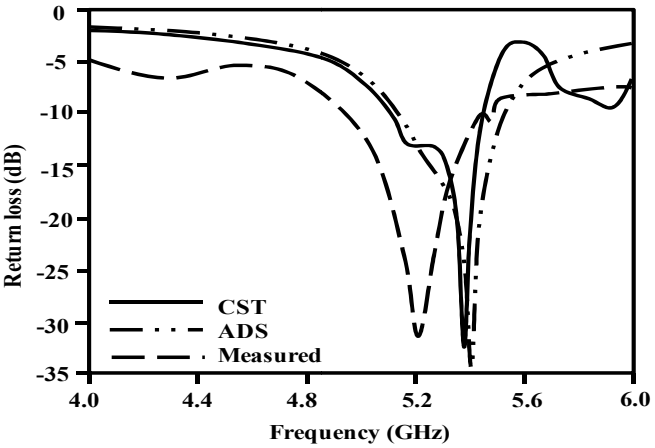
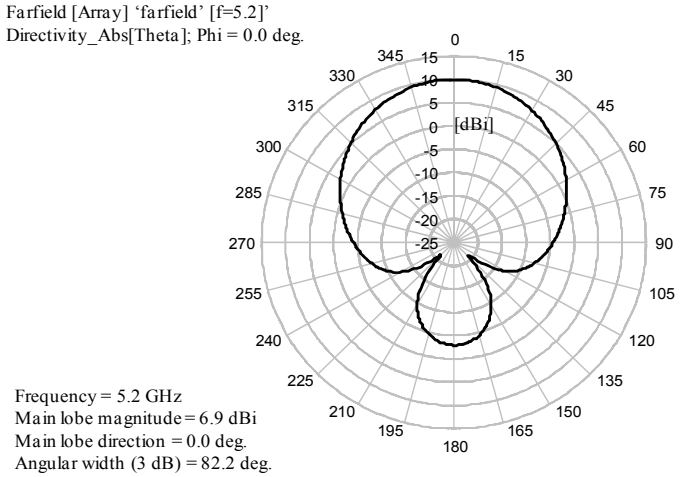


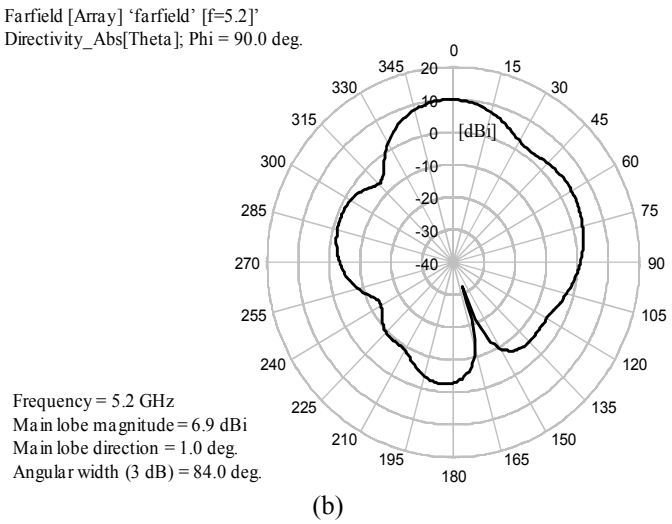
Fig. 9. Simulated (CST and ADS) and measured return loss of two element CDRA array.

The E- and H-plane radiation patterns of single element CDRA and two elements CDRA array at 5.2 GHz are depicted in Fig. 10 and Fig. 11 respectively. The corresponding magnitudes of the main beams are 6.9 dBi and 8.7 dBi respectively in the direction of 0^o. It is noticed that there is a spillover in the radiation patterns due to the small size ground plane. This kind of spillover in the radiation patterns is due to the electromagnetic scattering from the edges of the ground plane. However, simulated and measured results are in good agreement.

The directivity of the single element CDRA and two element CDRA array at 5.2 GHz is depicted in Table 2. The directivities of a single element and two element antenna array are 6.9 dBi and 8.7 dBi respectively. The gain enhancement in the two element antenna is due to array configuration.

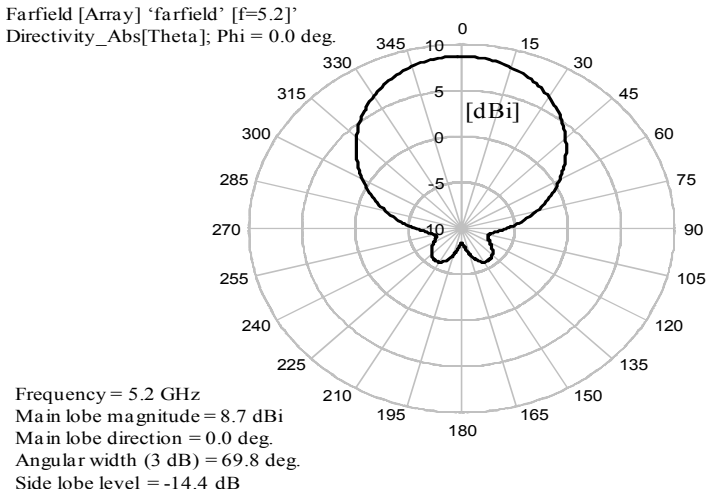


(a)

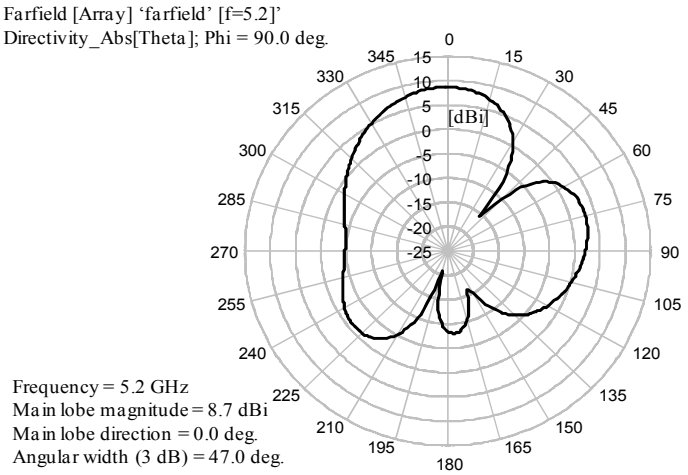


(b)

Fig. 10. The radiation pattern of single element CDRA at 5.2 GHz (a) E - plane (b) H - plane.



(a)



(b)

Fig. 11. The radiation pattern of two elements CDRA array at 5.2 GHz (a) E - plane (b) H - plane.

Table 2. Directivity of single element and two element antenna array.

Antenna Configuration	Frequency (GHz)	Directivity (dBi)
Single element	5.2	6.9
Two Element	5.2	8.7

IV. CONCLUSION

In this paper, electrical model of single element CDRA and two element CDRA array is presented. Antennas feasibility is verified by comparing the results obtained through CST, ADS against those obtained through fabricated prototypes. The two element antenna array achieved a significant enhancement in the antenna gain as compared to single element CDRA.

REFERENCES

[1] A. Petosa, “Dielectric resonator antennas Handbook”, Artech House Inc., 2007.

[2] R. Kumari, K. Parmar, et. al., "A dual band triangular shaped DRA array for WLAN/WiMAX applications", Annual IEEE India Conference (INDICON, p.p 1-4, 2011.

[3] A. Petosa, A. Ittipiboon, et. al., "Dielectric resonator antenna technology for wireless applications," IEEE Antennas and Propagation for Wireless Communications, p.p 117-120, 1998.

[4] M. S. Aras, M. K. A. Rahim, et. al., "An array of a dielectric resonator antenna for wireless application", IEEE international conference on RF and Microwave, p.p 459-463, 2008.

[5] Y. Huang and K. Boyle, “Antennas: From Theory to Practice”, John Wiley & sons, 2008.

[6] S. A. Long, M. W. McAllister and L. C. Shen, "The resonant cylindrical dielectric cavity antenna", IEEE Transactions on Antennas and Propagation, Vol. 31, p.p 406-412, 1983.

[7] K. M. Luk and K. W. Leung, “Dielectric Resonator Antennas”, Research Studies Press, Hertfordshire, U.K, 2003

[8] S. Sreekantan, Y. K. Ling, et. al., "Simulation and experimental investigators on rectangular, circular and cylindrical dielectric resonator antenna", Progress In Electromagnetics Research C, Vol. 7, p. p 151-166, 2009.

[9] L. C. Y. Chu, D. Guha, et. al., "Conformal Strip-Fed Shaped Cylindrical Dielectric Resonator: Improved Design of a Wideband Wireless Antenna", IEEE Antennas and Wireless Propagation Letters, Vol. 8, p.p 482-485, 2009.

[10] D. Yau and M. V. Shuley, "Numerical analysis of an aperture coupled rectangular dielectric resonator antenna using a surface formulation and the method of moments", IEEE proceeding on Microwaves, Antennas and Propagation, Vol. 146, p.p 105-110, 1999.

[11] M. F. Ain, Y. M. Qasaymeh, et. al., "A novel 5.8 GHz high gain array dielectric resonator antenna", Progress In Electromagnetics Research C, Vol. 15, p.p 201-210 2010.

[12] A. Derneryd, U. M. Khan, et. al., "Dual-polarized dielectric resonator antennas for base station applications", IEEE Proceedings of the 5th European Conference on Antennas and Propagation, p.p 606-610, 2011.

[13] H. Dashti, M. H. Neshati and F. Mohanna, "RDRA Array Fed by Dielectric Image Line with Improved Gain & Low Back Radiation", First International Conference on Communication Engineering, p.p 44-44, 2010.

[14] P. D. M. Pozar, "A Review of Aperture Coupled Microstrip Antennas: History, Operation, Development, and Applications," 1996.

[15] R. E. Collin, "Foundations for Microwave Engineering", 2nd Edition, Wiley - IEEE Press, 2001.

[16] M. F. Ain, Y. M. Qasaymeh, et. al., "Novel modeling and design of circularly polarized dielectric resonator antenna array", Progress In Electriomagnetic Research, Vol. 28, p.p 165-179, 2012.

[17] H. G. Akhavan and D. Mirshekar-Syahkal, "Approximate model for microstrip fed slot antennas", Electronics Letters, Vol. 30, p.p 1902-1903, 1994.

Load Distribution using Modified RED for Multipath TCP Communication

Sumet Prabhavat[1] Ruttikorn Varakulsiripunth[2] Satoshi Utsumi[3] Yasushi Kato[4]

[1]Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
[2]Faculty of Information Technology, Thai-Nichi Institute of Technology, Bangkok, Thailand
[3,4]Tsuruoka National College of Technology, Yamagata, Japan
[1]sumet@it.kmitl.ac.th, [2]kvruttik@gmail.com, [3]u.satoshi0215@gmail.com, [4]president@tsuruoka-nct.ac.jp

***Abstract*— Demand for large bandwidth network connections is growing continuously while network connections via multiple paths can exist because of high degree of connectivity. Utilization of the multiple paths is an economical solution for provisioning large network capacity to meet the users' requirements but it brings about key issues such as load imbalance (causing a large queue and buffer overflow at a gateway) and packet reordering problems that have significant impacts on TCP connections. In this paper, we propose a load distribution model for TCP transmission, called Load Distribution over Multipath with Extended Drop Slope Random Early Detection (LDM/ExRED). LDM/ExRED takes advantages of RED mechanism to maintain a small queue, avoid buffer overflow at a gateway, and mitigate packet reordering problem. In addition, it reduces peak buffer occupancy by using our modified RED (i.e., ExRED). Our work will be evaluated and compared to the other existing models by simulations under realistic traffic conditions.**

Keywords—Load Distribution; Load Balancing; Multipath TCP

I. INTRODUCTION

A wide variety of network service requirements demanded by users has been an important driving force for development of networking technologies [1]. By using today's technologies, multipath configurations can be established in various ways [2]. Their primary objectives are to improve network reliability by increasing network availability and reducing network downtime. A main path was used for data transmission while the others were backups which would be activated when the main path became unavailable. Currently, applications of multiple paths aims not only at avoiding a single point of failure but also at facilitating network provision [1–4] where its effectiveness is essential to quality of network services and Quality of Service (QoS) guarantee. Using multiple paths as a single path with aggregate bandwidth is a practical solution which is preferable rather than provisioning a large-bandwidth path because it offers a possibility to establish a very large-bandwidth connection. This improves scalability to support the future growth in bandwidth demand and affordability for network users. It also provides flexibility in bandwidth management within the communication protocol over the multipath network. Bandwidth aggregation and network-load balancing are important issues having been attractive research topics, and a large amount of load distribution approaches have been proposed. The rest of this paper is organized as follows. Section II briefly describes existing load distribution models, some studies on packet reordering problems, and an active queue management for TCP networks. Section III presents a new load distribution model with our modified active queue management mechanism. Section IV discusses performance evaluation under real traffic conditions. Conclusions are then given in Section V.

II. RELATED WORKS

A. Load Distribution Models

Load distribution can be classified into packet-based model and flow-based models. Packet-based models can be further classified into non-adaptive and adaptive models. Surplus Round Robin (SRR) [5] is a simple non-adaptive load distribution model. SRR is an enhanced version of round robin family model which is widely used schemes in the Internet because it can achieve starvation-free and acceptable load balancing efficiency. Least-loaded first (LLF) [6–8] is an adaptive load distribution model. LLF is a well-known load-sharing approach introduced to manage task loads with heavy-tailed distribution, where a task is assigned to the least-loaded server. In multipath load balancing with this scheme, a path having the smallest load or the shortest queue will be selected for an arrived packet. The most important drawback is that they do not consider task sequence, which can result in the packet reordering problem [9].

On the other hand, Direct Hashing (DH) [10], Table-based Hashing (TH) [11], and Fast Switching (FS) [12] are flow-based models which can prevent packet reordering but they can cause load imbalance problem due to inability to cope with variance of the flow size distribution. This problem can cause a large buffer occupancy leading to a high risk of buffer overflow [2]. Adaptive load distribution models, e.g., Load Distribution over Multipath (LDM) and Load Balancing for Parallel Forwarding (LBPF) [13–15] were introduced to mitigate the load imbalance problem. However, there is trade-off between balancing load and maintaining a low risk of packet reordering. This depends on the respective algorithms as well as their set parameters.

B. Packet Reordering Problem

Several works have studied on packet reordering problem [16–22]. Packet reordering can occur with high probabilities of splitting a flow and switching a path [18], [19]. On the other hand, the possibility of packet reordering will decrease if the interarrival time of two consecutive packets belonging to the same flow is longer than the longest time required to deliver a packet via the parallel paths [23]. The packet interarrival time must be greater than the difference between the longest and the shortest delays among parallel paths to ensure packet order preservation. If the load is not properly distributed among the parallel paths, some paths may experience congestion (leading to large queues) while the others are idle; consequently, the delay difference can be very large.

C. Active Queue Management for TCP Gateway

Random Early Detection (RED) [24] is a well-known active queue management mechanism introduced to improve performance of TCP communication. RED itself consists of two main parts, i.e., the estimation of an average queue size and the decision of whether or not to drop an incoming packet. An average queue size is calculated by a current queue size using an exponentially weighted moving average (EWMA). Based on the average queue size, RED decides whether or not to drop an incoming packet. Packet arrival rate of a flow plays an important role in determining packet drop probability. The higher the packet arrival rate, the higher the packet drop rate. This feature of RED is to prevent two major problems that have a significant impact on TCP connections: buffer overflow and synchronization of TCP flows [25] due to burst traffic. In LDM/RED [26], we uses RED for sensing congestion and marking a packet arriving at the time congestion occurs. The marked packet will be switched to a less congested path.

III. Load Distribution (over Multipath Networks) with Extended Drop Slope Random Early Detection

Figure 1 describes our approach called Load Distribution over Multipath with Extended Drop Slope Random Early Detection (LDM/ExRED), which is developed from LDM/RED [26] and our modified RED algorithm (i.e., ExRED) [27]. Assume that multiple and parallel connections are established between a source and a destination before data transmission is started. For each path p, bandwidth is μ_p. The other parameters, e.g., hop counts and path delay, are assumed reflected by a measured queue length q_p. The input traffic is split into flows, each of which independently takes a path determined by a path selection component. LDM/ExRED is a variant version of FS [12] and LDM [13], a table-based model which selects paths based on the flow-path mapping table. A packet belonging to an existing flow is sent via the same path as its previous ones. When a new flow appears, a packet belonging to the new flow will be sent via a new path having the shortest queue.

In LDM/ExRED, a mapping table is modified and the path for an existing flow will be changed. The existing flow is switched (to a new path having the shortest queue) with probability π_s which is determined by ExRED mechanism, as illustrated in Fig. 2. Based on an average queue length $\hat{q}_p$,

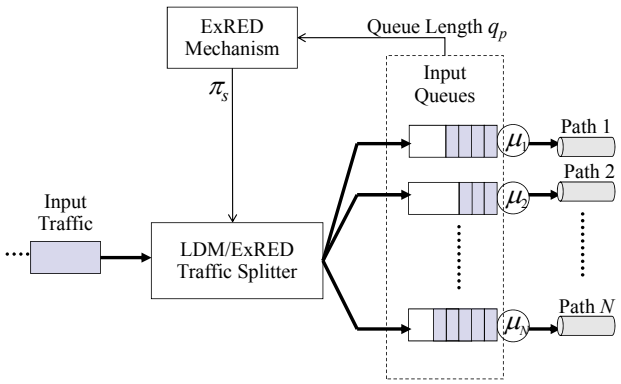


Figure 1. Description of the proposed model, LDM/ExRED

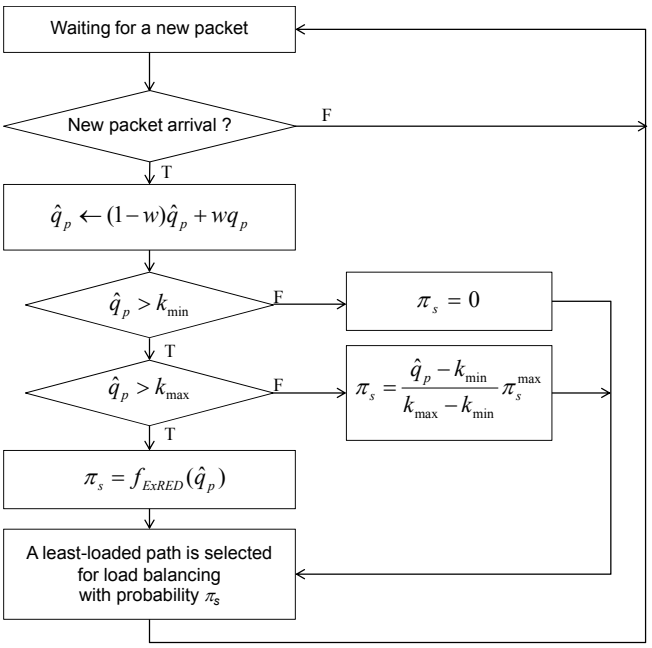


Figure 2. Flowchart to calculate π_s in traffic splitting algorithm

LDM/ExRED makes a decision of whether or not to split the flow. There are two queue thresholds: minimum and maximum thresholds, i.e., $k_{\min}$ and $k_{\max}$. When $\hat{q}_p > k_{\min}$, the flow is split with probability $\pi_s > 0$. The larger the average queue size $\hat{q}_p$, the higher the probability π_s. When $\hat{q}_p > k_{\max}$, π_s increases according to polynomial function in (1) until $\pi_s = \pi_s^{\max}$ (which is a desired maximum value of splitting probability). When $\pi_s^{\max} = 1$ and $\pi_s = \pi_s^{\max}$, all flows are always moved to a new path.

$$f_{ExRED}(\hat{q}) = a_2\hat{q}^2 + a_1\hat{q} + a_0 \quad (1)$$

where coefficients of the positive increasing function are

$$a_2 = \frac{(k_{\max} - k_{\min}) - (K - k_{\min})\pi_s^{\max}}{(k_{\max} - k_{\min})(K - k_{\max})^2}$$

$$a_1 = \frac{(k_{\max}^2 + K^2 - 2k_{\max}k_{\min})\pi_s^{\max} - 2k_{\max}^2 + 2k_{\max}k_{\min}}{(k_{\max} - k_{\min})(K - k_{\max})^2}$$

$$a_0 = \frac{k_{\max}^3 - k_{\max}^2 k_{\min} - (k_{\max}^2 + 2k_{\max}k_{\min} + k_{\min}K)K\pi_s^{\max}}{(k_{\max} - k_{\min})(K - k_{\max})^2}$$

Equation (1) shows that ExRED provides flexible increasing of path-switching probability. It can prevent a consecutive path-switching in case of the average queue size is over maximum threshold and not exceeds limit of buffer size (K). We expect that this will allow us to utilize a particular path as much as possible while not causing overload on the path.

IV. Performance Evaluation

In this section, comparative performance under different conditions of real traffics is demonstrated.

A. Simulation model and parameters

Three simulation scenarios are conducted to compare the performance of load distribution models, by using 1-hour long real traffic traces [28], i.e., DSet1, DSet2, and DSet3, which are captured from primary Internet access point between Digital Equipment Corporation and the Internet. Characteristics of the traces are listed in Table I. Three paths exist between a pair of source and destination. Bandwidth capacities (or mean service rates) of path 1, path 2, and path 3 are 1, 4, and 7 Mbps, respectively; the total bandwidth capacity is 12 Mbps. As compared to the total bandwidth capacity, traffics generated from trace DSet1 and DSet2 incur heavy load and some load-spikes. Moreover, we use trace DSet3 to generate extremely heavy traffic which peak load is much higher than the total bandwidth capacity, thus incurring network overload.

With the set-up simulation environment, LDM/ExRED, LDM, LLF, and SRR are evaluated. In LDM/RED and LDM/ExRED, the parameters are chosen: EWMA parameter w=0.02, $k_{\min}$=10, $k_{\max}$=30, and $\pi_s^{\max}$=0.1. These parameters setting are recommended in [24], [27], and [29]. In SRR, weights assigned for path 1, path 2, and path 3 are 1, 4, and 7, respectively, which are used for specifying desired forwarding rates on each path [5]. In LBPF, the size of the table is 20, W=1000, and P=20, where W is window size for detecting aggressive flows and P is period for splitting the aggressive flows. LBPF's parameters can be further read at [14].

B. Load Balancing Efficiency

Tables II shows that LDM/ExRED can achieve load balancing, as compared to the others since probabilities of buffer occupancy of all paths are almost equal. This can be implied that all paths are fairly utilized. Moreover, Table III shows that LDM/RED and LDM/ExRED can prevent buffer overflow; maximum queue sizes incurred by both of them are significantly smaller than those incurred by SRR and LDM. This is because burst traffic (which can cause a large queue) is detected by RED and then distributed over multiple paths.

TABLE I. Profile of Traffic Traces

Trace ID	# Packets x10^6	Traffic Rate (Mbps.)			# Different Flows	Flow Size (Packets)		Flow Rate (Flows/Second)		
		Mean	Min.	Max.		Mean	CV	Mean	Min.	Max.
DSet 1	2.66	5.91	2.07	13.65	5865	453.87	7.52	137.89	77	204
DSet 2	2.87	6.38	0.46	12.24	12903	222.71	5.98	175.32	44	247
DSet 3	3.86	8.58	1.86	15.45	12710	303.88	7.11	184.50	90	269

TABLE II. Probability of Buffer Occupancy

Load Distribution Model	Dset 1			Dset 2			Dset 3		
	Path 1	Path 2	Path 3	Path 1	Path 2	Path 3	Path 1	Path 2	Path 3
LDM/ExRED	0.71	0.66	0.65	0.73	0.68	0.65	0.86	0.82	0.80
LDM/RED	0.70	0.66	0.65	0.73	0.67	0.65	0.87	0.82	0.80
LDM	0.86	0.71	0.59	0.89	0.73	0.58	0.95	0.85	0.77
LLF	0.77	0.53	0.47	0.80	0.55	0.48	0.91	0.74	0.70
SRR	0.35	0.66	0.66	0.37	0.68	0.68	0.61	0.81	0.82

TABLE III. Maximum Queue Size (Packets)

Load Distribution Model	Dset 1			Dset 2			Dset 3		
	Path 1	Path 2	Path 3	Path 1	Path 2	Path 3	Path 1	Path 2	Path 3
LDM/ExRED	83	81	82	85	83	82	319	318	317
LDM/RED	100	98	100	104	97	102	376	375	375
LDM	1685	422	180	3845	378	76	5976	2780	522
LLF	92	92	93	83	83	83	293	293	293
SRR	43	86	185	33	88	165	110	350	557

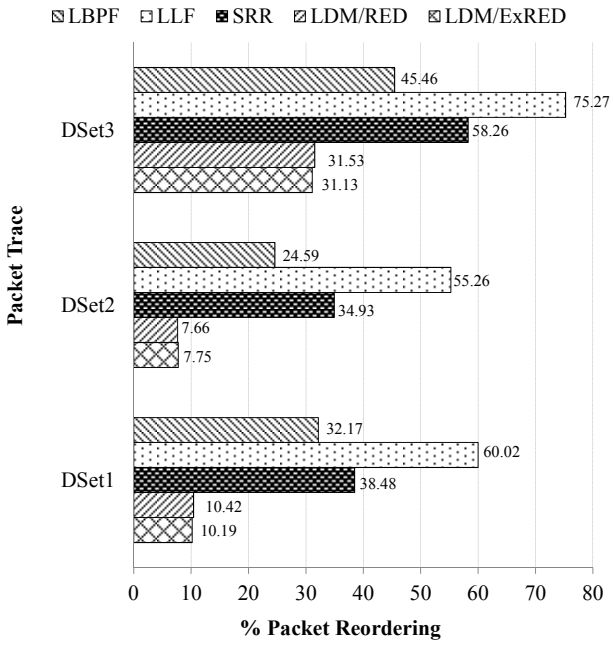


Figure 3. Risk of Packet Reordering

However, when buffer is highly utilized, RED causes consecutive marked packets while ExRED can prevent such problem. Therefore, LDM/ExRED can reduce peak buffer occupancy. A gateway device requires only a small buffer.

C. Packet Order Preservation

Fig. 3 illustrates that LDM/ExRED and LDM/RED can efficiently alleviate packet reordering inherently existing in packet-based models such as SRR and LLF. LDM is a flow-based model which can perfectly prevent packet reordering. SRR sends packets in a round robin manner. It can cause a risk of packet reordering. LLF choosing only the path with the shortest queue size can cause a high risk of packet reordering. LBPF splits a group of largest flows, thus leading to the risk of packet reordering. The discussion on this issue can be found in [2], [23], and [30]. As compared to LDM/RED, LDM/ExRED has a potential to maintain low risk of packet reordering because of a smaller path-switching probability. However, difference between delay on high congested path and that on a less congested path is also a factor of the risk of packet reordering [2], [23]. This is an open issue for further studies.

V. Concluding Remarks

This paper proposes a load distribution model for TCP transmission. LDM/ExRED takes advantages from RED mechanism in traffic splitting and distribute load subject to necessary condition to prevent packet reordering. With a modified RED, i.e., ExRED, it can prevent buffer overflow (due to burstiness of TCP flows) at a gateway and maintain low risk of packet reordering. The superiority of LDM/ExRED is presented by comparative performance among LDM/ExRED and the current existing load distribution models by simulations under variety of real traffic conditions.

References

[1] L. Golubchik, J. Lui, T. Tung, A. Chow, W. Lee, G. Franceschinis, and C. Anglano, "Multi-path continuous media streaming: What are the benefits?," *Performance Evaluation*, vol. 49, pp. 429–449, Sep. 2002.

[2] S. Prabhavat, H. Nishiyama, N. Ansari, and N. Kato, "On Load Distribution over Multipath Networks," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 3, pp. 662–680, 2012.

[3] M. Menth, R. Martin, A. Koster, and S. Orlowski, "Overview of resilience mechanisms based on multipath structures," in *Proc. the 6th International Workshop on Design and Reliable Communication Networks (DRCN)*, La Rochelle, France, Oct. 2007.

[4] N. Wang, K. Ho, G. Pavlou, and M. Howarth, "An overview of routing optimization for Internet traffic engineering," *IEEE Communications Surveys and Tutorials*, vol. 10, no. 1, pp. 36–56, 2008.

[5] H. Adiseshu, G. Parulkar, and G. Varghese, "A reliable and scalable striping protocol," *ACM SIGCOMM Computer Communication Review*, vol. 26, no. 4, pp. 131–141, Oct. 1996.

[6] K. G. Shin and C. J. Hou, "Design and Evaluation of Effective Load Sharing in Distributed Real-Time Systems," *IEEE Trans. Parallel and Distributed Systems*, vol. 5, no. 7, pp. 704–719, Jul. 1994.

[7] O. Kremien and J. Kramer, "Methodical analysis of adaptive load sharing algorithms," *IEEE Trans. Parallel Distribution Systems*, vol. 3, no. 6, pp. 747–760, Nov .1992.

[8] C. C. Hui and S. T. Chanson, "Hydrodynamic load balancing," *IEEE Trans. Parallel and Distributed Systems*, vol. 10, no. 11, pp. 1118–1137, Nov. 1999.

[9] Z. Tari, J. Broberg, A. Zomaya, and R. Baldoni, "A least flow-time first load sharing approach for distributed server farm," *Journal of Parallel and Distributed Computing*, vol. 65, no. 7, pp. 832–842, Jul. 2005.

[10] C. Villamizar, "OSPF optimized multipath (OSPF-OMP)," Internet draft draft-ietf-ospf-omp-02.txt, Feb. 1999.

[11] D. Thaler and C. Hopps, "Multipath issues in unicast and multicast next-hop selection," RFC 2991, Nov. 2000.

[12] A. Zinin, Cisco IP Routing, packet forwarding and intra-domain routing protocols. Reading, MA: Addison Wesley, 2002.

[13] J. Song, S. Kim, M. Lee, H. Lee, and T. Suda, "Adaptive load distribution over multipath in MPLS networks," in *Proc. IEEE ICC*, Anchorage, Alaska, May 2003, pp. 233–237.

[14] W. Shi, M. H. MacGregor, and P. Gburzynski, "Load balancing for parallel forwarding," *IEEE/ACM Trans. Networking*, vol. 13, no. 4, pp. 790–801, Aug. 2005.

[15] S. Kandula, D. Katabi, S. Sinha, and A. Berger, "Dynamic load balancing without packet reordering," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 2, pp. 53–62, Apr. 2007.

[16] V. Paxson, G. Almes, J. Mahdavi, and M. Mathis, "Framework for IP performance metrics," RFC 2330, May. 1998.

[17] C. Demichelis and P. Chimento, "IP packet delay variation metric for IP performance metrics (IPPM)," RFC 3393, Nov. 2002.

[18] N. M. Piratla, A. P. Jayasumana, A. A. Bare, and T. Banka, "Reorder buffer-occupancy density and its application for measurement and evaluation of packet reordering," *Computer Communications*, vol. 30, no.9, pp.1980–1993, Jun. 2007.

[19] N. M. Piratla and A. P. Jayasumana, "Reordering of packets due to multipath forwarding – An analysis," in *Proc. IEEE ICC*, Istanbul, Turkey, Jun. 2006, pp. 829–834.

[20] A. Morton, L. Ciavattone, G. Ramachandran, S. Shalunov, and J. Perser, "Packet reordering metrics," RFC 4737, Nov. 2006.

[21] A. Jayasumana, N. Piratla, T. Banka, and R. Whitner, "Improved packet reordering metrics," RFC 5236, Jun. 2008.

[22] S. Jaiswal, G. Iannaccone, C. Diot, J. Kurose, and D. Towsley, "Measurement and classification of out-of-sequence packets in a tier-1 IP backbone," *IEEE/ACM Trans. Networking*, vol. 15, no. 1, pp. 54–66, Feb. 2007.

[23] S. Prabhavat, H. Nishiyama, N. Ansari, and N. Kato, "On the Performance Analysis of Traffic Splitting on Load Imbalancing and Packet Reordering of Bursty Traffic," in *Proc. IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC 2009)*, Beijing, China, Nov. 2009.

[24] S. Floyd and V. Jacobsan, "Random Early Detection Gateways for Congestion Avoidance," *IEEE/ACM Trans. Networking*, vol. 1, no. 4. pp. 397-413, Aug. 1993.

[25] S. Floyd and V. Jacobson, "On traffic phase effects in packet-switched gateways," *Internetworking: Research & Experience*, vol. 3, no. 3, Sep. 1992, pp.115-56, UK.

[26] S. Prabhavat, N. Kato, and R. Varakulsiripunth, "Load Distribution Mechanism for Multipath TCP Communication," in *Proc. International Symposium on Technology for Sustainability (ISTS 2011)*, Bangkok, Thailand, Jan. 2012, pp. 321–324.

[27] S. Prabhavat and R. Varakulsiripunth, "Performance Improvement on RED Based Gateway in TCP Communication Network," in *Proc. International Conference on Control, Automation and Systems (ICCAS 2004)*, Bangkok, Thailand, Aug. 2004, pp. 782–787.

[28] (Online Sources) P. Danzig, J. Mogul, V. Paxson, and M. Schwartz. (1995, Mar.). *The Internet Traffic Archive*. Available: http://ita.ee.lbl.gov/index.html.

[29] (Online Sources) S. Floyd (1967, Nov.), "RED: Discussions of Setting Parameters," Available: http://www.icir.org/floyd/REDparameters.txt.

[30] S. Prabhavat, H. Nishiyama, N. Ansari, and N. Kato, "Effective Delay-Controlled Load Distribution over Multipath Networks," *IEEE Trans. Parallel and Distributed Systems*, vol. 22, no. 10, pp. 1730–1741, Oct. 2011.

Microwave Bandpass Filter Using QMSIW

M. Z. Ur Rehman, Z. Baharudin*, M. A. Zakariya, M. H. M. Khir, M. T. Khan
Communication Cluster, Electrical and Electronics Engineering Department
Universiti Teknologi PETRONAS
*zuhairb@petronas.com.my

***Abstract*- A planar bandpass filter based on a technique that utilizes quarter mode of substrate integrated waveguide (SIW) is presented. This concept reduces the circuit foot print of SIW to a quarter of its size, along with miniaturization good performance and high quality factor is maintained by the structure. The design concept for single-pole resonator structure is presented; and by coupling the resonators a two-pole substrate integrated waveguide bandpass filter is achieved for the frequency of 5.9 GHz. The simulation results show that the filter insertion losses are better than 2 dB and return losses are less than 15 dB.**

Keywords- Substrate Integrated Waveguide, QMSIW, High quality factor, bandpass filters

I. INTRODUCTION

Filters have received a particular attention with the advent of various wireless systems, this interest has dramatically increased with the introduction and development of new millimeter waves applications over the past years. A variety of applications have been recently proposed including wireless local area networks [1], automotive radars [2], intelligent transportation systems [3] and imaging sensors [4]. Efficient filters demand has also increased with the development of chip-sets operating at 60 GHz or even higher frequencies by a number of semiconductor industries [5].

A promising candidate for filters is the Substrate Integrated Waveguide (SIW) structure which is achieved through incorporating the rectangular waveguide structure into the microstrip substrate [6]. SIWs are dielectric filled and are formed from the substrate material using two rows of conducting vias or slots connecting top and bottom metal plates. These vias are embedded in dielectric filled substrate; hence providing simple integration with other planar devices and size reduction. Although size reduction involving dielectric filled substrate instead of air-filled reduces the quality factor (Q), but the entire circuitry including structure and microstrip transitions can be fabricated by using standard printed circuit board (PCB) technology or other techniques, like LCP [9] and LTCC [8].

Technical literature review reveals that most of the waveguide components have been implemented in SIW structure. However, most of these SIWs have been operative in the microwave frequency range, with a few exceptions at higher frequency. However these filters are relatively large in contrast to microstrip planar filters and this is more evident in the lower microwave frequency band. In order to reduce the size of the SIW devices, some novel structures based on SIW technology have emerged; these include HMSIW, FSIW, FHMSIW and QMSIW structures.

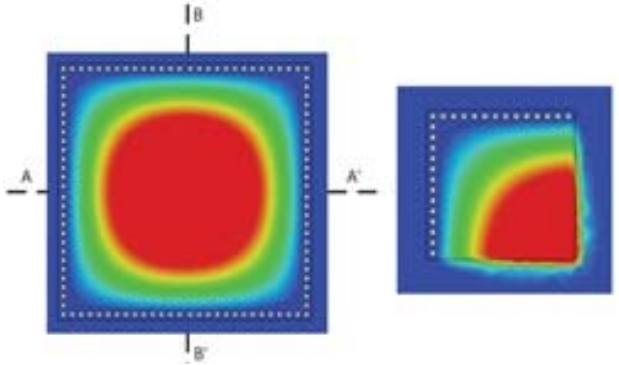


Fig. 1 Electric field distribution of SIW and Quarter mode SIW

Wei Hong et al in 2006 proposed the half-mode substrate integrated waveguide (HMSIW) concept which can reduce the size of SIW components by half [11]. An HMSIW is obtained through placement of an open circuit along the symmetric plane in the transmission direction; i.e. the remaining half keeps the remaining distribution fields intact if the cutting plane is a magnetic wall. The open side aperture of the HMSIW is nearly equivalent to a perfect MW due the high ratio of width to height, and it has been shown that HMSIW retains low loss characteristics obtained through SIW structures. Performances at microwave and millimeter wave bands show that the attenuation of it is less than that of planar microstrip and SIW. Thus, further compression can be brought in the size of a microwave or millimeter wave integrated circuit based on these wave structures.

The quarter-mode SIW (QMSIW) concept was proposed to further reduce the size of SIW components, its size is only a quarter of SIW resonator cavity. Fig. 1 represents the electric field of a QMSIW's resonant mode and its comparison with SIW electric field distribution. QMSIW based bandpass filters have been reported in [12].

In this paper a QMSIW resonator based bandpass filter is presented. This resonator structure is a miniaturized form of SIW cavity, and exhibits similar low-loss and high quality characteristics as found in its other counterparts. The proposed bandpass filter configuration is suitable for integration with planar devices and its small footprint area allows other devices to be easily integrated on a single board.

II. SUBSTRATE INTEGRATED WAVEGUIDE

Design and Implementation of SIW filters are performed through defined Practical methods so far. The most common

Table 1. Design Specifications of the Bandpass filter

Key Parameters	Values
Passband Centre Frequency	6 GHz
Passband Return Loss, S_{11}	< -10dB
Passband Insertions Loss, S_{21}	> -3db
Passband Bandwidth at -10dB	> 900 MHz
Stopband Rejection	> 30dB

technique is to form the SIW cavity through metallic sidewalls [10] as shown in Fig 2. A dielectric substrate with width of *h* forms the resonator/cavity and the resonator has a width of *W* and length of *L*. The top and bottom of the cavity are constructed through placing metallic plates and conducting posts/vias going through the substrate connects the top and bottom plates; hence forming the sidewalls of the cavity. The vias are of diameter *d* and the separation between two neighboring vias is given as *s*. The choice of diameter and separation between the two vias forms the basis of the SIW filters, therefore these should be selected in a manner that minimum radiation loss is exhibited.

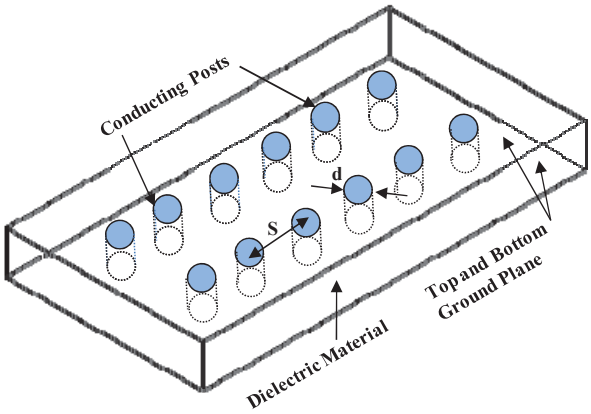


Fig. 2 SIW structure Geometry

The Dealandes and W_u [7] study reveals two primary design rules for SIW structures as given in Eq. 1; these rules are followed in order to ensure same design and modeling methodology adopted for rectangular waveguides. These rules pertain to the diameter *d* of the via posts and the via post spacing *s*:

$$d < \frac{\lambda_g}{5} \quad (1)$$

$$s \leq 4d \quad (2)$$

In our design *d* and s are chosen to be 0.5 mm and 1 mm respectively, these values ensure less radiation losses and the SIW cavity acts closely to a rectangular waveguide. For the dominant TE_{101} mode, the dimensions of the SIW resonator structure are determined by using the relation in eq. 3 [7].

$$f_{TE_{101}} = \frac{c}{2\sqrt{\mu_r \epsilon_r}} \sqrt{\left(\frac{1}{W_{eff}}\right)^2 + \left(\frac{1}{L_{eff}}\right)^2} \quad (3)$$

W_{eff} and L_{eff} are the effective width and length of the SIW cavity, respectively, and are given by:

$$W_{eff} = W - \frac{d^2}{0.95s}, L_{eff} = L - \frac{d^2}{0.95s} \quad (4)$$

Where; *W* and *L* are the real width and length of the SIW cavity, *c* is the velocity of light in free space. In this design of QMSIW only a quarter of the footprint of the SIW square resonant cavity is used. Therefore the width and length of the single-pole resonator structure is chosen to be half of the computed width and length using equation 3 and 4 as shown in fig. 3. Utilizing this method, the requirements in Table 1 provide the SIW bandpass filter.

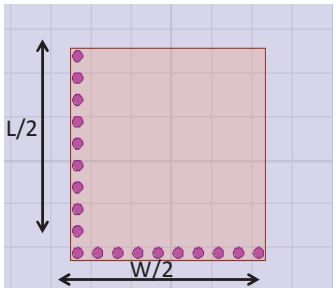


Fig. 3 Dimension of the single pole filter QMSIW resonator

QMSIW bandpass filter is designed using Roger RT/Duriod 5880 material substrate with a dielectric constant of 2.2 and substrate thickness of 787µm. Theoretically, the resonance frequency does not depend on the thickness of the substrate. However, it has been observed in literature that it does play a role on the loss (mainly on radiation loss). The thicker the substrate the lower is the loss or higher Q. It has been shown that slight increase or decrease in the substrate thickness changes the unloaded quality factor.

To accomplish two pole bandpass SIW filter design, once a single pole QMSIW cavity resonator is created for a specific resonant mode, the design methodology closely resembles conventional simulation-based microstrip filter design [7]. Two single cavity resonators are coupled together through capacitive coupling along the RF input and output microstrip-to-SIW transition areas and a capacitive coupling located in-between two SIW cavities as can be seen in figure 4.

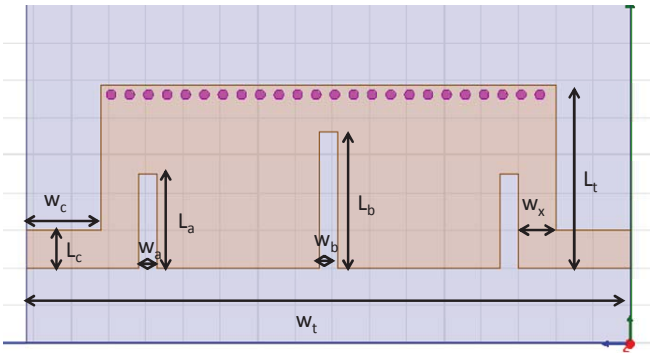


Fig. 4 Two pole bandpass filter dimensions

Table 2. Dimensions of the two pole filter resonator structure

Notation	Value (mm)	Notation	Value (mm)
L_a	5	W_a	1
L_b	7.2	W_b	1
L_c, W_x	2	W_c	4
L_t	9.7	W_t	32.4

Coupling coefficients and external quality factors are calculated from derived expressions based on lowpass prototype parameters [10].

$$M_{1,2} = \frac{FBW}{\sqrt{g_1 g_2}}, \qquad Q_{e1} = \frac{g_0 g_1}{FBW}, \qquad Q_{e2} = \frac{g_2 g_3}{FBW} \tag{4}$$

These values are then compared to the simulated extracted coupling coefficients and external quality factors for a particular cavity geometry.

$$Q_{ext} = \frac{f_0}{\Delta f_{-3dB}} \tag{5}$$

Iterations and adjustments to the dimensions of the coupling areas of the filter are performed until the calculated values match the extracted values from full-wave simulation, providing the desired filter.

III. Results and Discussions

The desired fixed filter described in the specifications in Table 1 and its corresponding designs structure shown in Fig. 3 and Fig. 4 are realized with the responses shown in Fig. 5 and Fig. 6 respectively. The simulations are carried out using HFSS full-wave simulation.

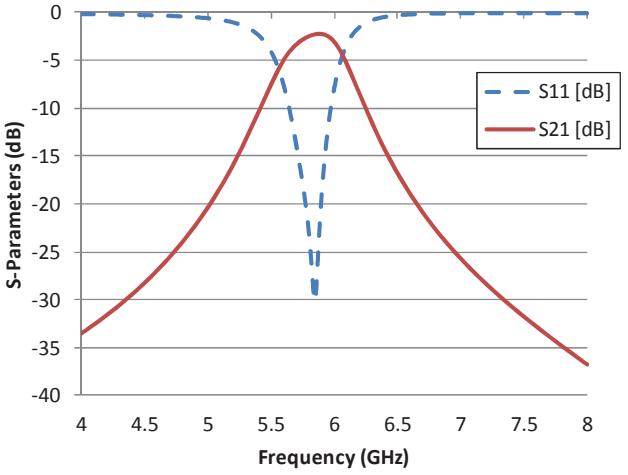


Fig. 5 Response of the Single Pole filter

Figure 5 shows the response of the Single Pole filter designed for 6 GHz, the response of the filter is obtained using the resonator structure shown in Fig. 3, the resonator designed based on the requirements using the equations presented in Section 2.The simulated S21 and S11 response of the QMSIW bandpass filter for the single pole reveals that the S11 value at the center frequency of 5.85 GHz is less than -25 dB, whereas the S21 response is greater than -3 dB and the passband bandwidth at -10 dB is greater than 800 MHz.

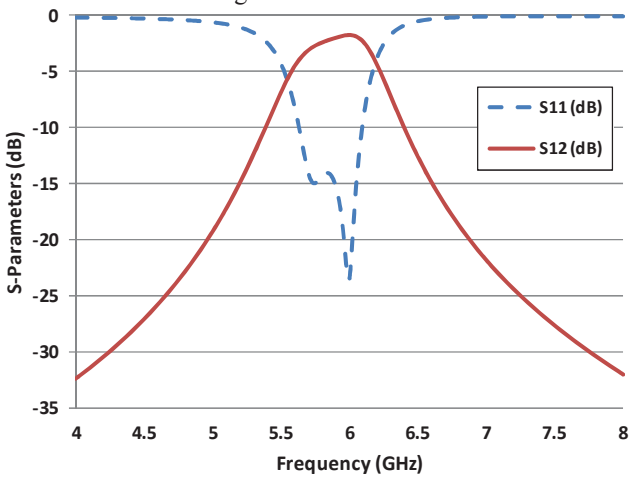


Fig. 6 Response of the two pole filter

The simulated S21 and S11 response of the two pole bandpass filter is shown in fig. 6. The S21 response at the centre frequency of 5.9 GHz is better than -2 dB and its corresponding S11 response at the centre frequency is less than -15 dB. The passband bandwidth at -10 dB is greater than 900 MHz, whereas the lower and upper stopband rejections are better than -25 dB. These responses are summarized in Table 3. The performance in the insertion loss might be due to the structure being a miniaturized form of the SIW structure.

Table 3. Summary of the Filter performance

Key Parameters	Values
Passband Centre Frequency (GHz)	5.9
Passband Return Loss, S_{11} (dB)	< -15
Passband Insertions Loss, S_{21} (dB)	> -2
Lower and upper stopband rejection (dB)	> -25
Passband Bandwidth at -10dB level (GHz)	0.9

Conclusion

A 6 GHz bandpass filter based on QMSIW structure is proposed in this paper, the filter has presented good performance and a miniaturized version of the SIW structure is exploited in the design process. The filter has also presented broad bandwidth at the -10 dB level. This filter presents good performance in its small size, and it can be easily integrated with planar circuits and devices.

References

[1] James, J.; Pengbo Shen; Nkansah, A.; Xing Liang; Gomes, N.J.; , "Millimeter-Wave Wireless Local Area Network Over Multimode Fiber System Demonstration," *Photonics Technology Letters, IEEE* , vol.22, no.9, pp.601-603, May1, 2010

[2] Li Wang; Glisic, S.; Borngraeber, J.; Winkler, W.; Scheytt, J.C.; , "A Single-Ended Fully Integrated SiGe 77/79 GHz Receiver for Automotive Radar," *Solid-State Circuits, IEEE Journal of* , vol.43, no.9, pp.1897-1908, Sept. 2008

[3] Yan, Xinping; Zhang, Hui; Wu, Chaozhong; , "Research and Development of Intelligent Transportation Systems," *Distributed Computing and Applications to Business, Engineering & Science*

(DCABES), 2012 11th International Symposium on , vol., no., pp.321-327, 19-22 Oct. 2012

Millimeter, and Terahertz Waves (IRMMW-THz), 2012 37th International Conference on , vol., no., pp.1, 23-28 Sept. 2012.

[5] Niknejad, A.M., Hashemi, H.: 'Millimetre-wave silicon technology: 60 GHz and beyond' (Springer, 2008)

[6] Hirokawa, J., Ando, M.: 'Single-layer feed waveguide consisting of posts for plane TEM wave excitation in parallel plates', IEEE Trans. Antennas Propag., 1998, 46, (5), pp. 625–630.

[7] Deslandes, D., Wu, K.: 'Accurate modeling, wave mechanisms, and design considerations of a substrate integrated waveguide', IEEE Trans. Microw. Theory Tech., 2006, 54, (6), pp. 2516–2526.

[8] Junfeng Xu; Zhi Ning Chen; Xianming Qing; Wei Hong; , "140-GHz Planar Broadband LTCC SIW Slot Antenna Array," *Antennas and Propagation, IEEE Transactions on* , vol.60, no.6, pp.3025-3028, June 2012.

[9] Ki Seok Yang; Pinel, S.; Il Kwon Kim; Laskar, J.; , "Low-Loss Integrated-Waveguide Passive Circuits Using Liquid-Crystal Polymer System-on-Package (SOP) Technology for Millimeter-Wave Applications," *Microwave Theory and Techniques, IEEE Transactions on* , vol.54, no.12, pp.4572-4579, Dec. 2006.

[4] Wilson, J. P.; Schuetz, C. A.; Martin, R.; Dillon, T. E.; Yao, P.; Prather, D. W.; , "Polarization sensitive millimeter-wave imaging sensor based on optical up-conversion scaled to a distributed aperture,"*Infrared,*

[10] Tang, H.J., Hong, W., Hao, Z.C., Chen, J.X., Wu, K.: 'Optimal design of compact millimetre-wave SIW circular cavity filters', Electron. Lett., 2005, 41, (19), pp. 1068–1069

[11] W. Hong, B. Liu, Y.Q. Wang et al., "Half Mode Substrate Integrated Waveguide: A New Guided Wave Structure for Microwave and Millimeter Wave Application," *2006 IRMMW-THz Conference Digest*, pp. 219.

[12] Zhenyu Zhang; Ning Yang; Ke Wu; , "5-GHz bandpass filter demonstration using quarter-mode substrate integrated waveguide cavity for wireless systems," *Radio and Wireless Symposium, 2009. RWS '09. IEEE* , vol., no., pp.95-98, 18-22 Jan. 2009

Performance Comparison of IEEE 1609.4/802.11p and 802.11e with EDCA Implementation in MAC Sublayer

Doan Perdana
Department of Electrical Engineering
Faculty of Engineering, University of Indonesia
Kampus Baru UI, Depok 16424, Indonesia
doan.perdana@ui.ac.id

Riri Fitri Sari
Department of Electrical Engineering
Faculty of Engineering, University of Indonesia
Kampus Baru UI, Depok 16424, Indonesia
riri@ui.ac.id

***Abstract*-One of the most challenging issues in IEEE 1609.4 is the assurance of Quality of Service (QoS), i.e. to improve throughput and reduce delay in the sublayer Medium Access Control (MAC) IEEE 1609.4. The prioritization of each service package, using Enhanced Distributed Channel Access (EDCA) at the MAC sublayer is designed based on the IEEE 802.11e with some modifications to the transmission parameters.**

In this paper, we evaluate the throughput and delay performance with EDCA implementation on the IEEE MAC sub layer 1609.4/802.11p based on simulation. Our work also evaluate the performance of the service differentiation capability of the IEEE1609.4/802.11p, in terms of throughput and queuing delay, compared to that IEEE 802.11e. On this work, we use ns 2.34 simulator to evaluate the performance of the MAC sub layer IEEE 1609.4.

From the simulation and based on channel performance, it was found that the delay performance based on IEEE1609.4/802.11p standard has been degraded 13.87%, and the throughput is 57.49% better compared with IEEE 802.11e standard. Otherwise, based on the prioritization of each service package on IEEE1609.4/802.11p standard, we found that the queuing delay of AC1 and SCH3 (CW*min* = 3 and CW*max* = 14) in IEEE1609.4/802.11p is much larger than the rest of the AC queues.

Keywords : ***IEEE 1069.4/802.11p,IEEE 802.11e, Enhanced Distributed Channel AccessFunction (EDCA), MAC Sublayer***

I. INTRODUCTION

Vehicular Ad Hoc Network (VANET) is a sub-group of MANET which can support both Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications [1,2]. The main characteristic of both MANET and VANET are self-organization and movement [1,2]. Recently, IEEE 802.11 working group developed an amendment to the 802.11 standard in order to support vehicular ad hoc network [1].

The Multichannel Operations IEEE 1609.4 is an emerging standard which support the co-existence of safety (CCH Channel) and non-safety (SCH Channel) message application in vehicular networks [3,4].

This strictly concerns on the transmission slot synchronization of CCH/SCH based on Wireless Access in Vehicular Environments (WAVE) to achieve QoS requirements (i.e. to improve packet delivery ratio and reduce delay) [3,4].

The IEEE 1609.4 standard [2,4] for multichannel operation in VANETs defines a number of channels, each channel for different application, with different characteristics as shown in Fig.1. In addition to different frequency being used, different maximum transmit powers are allocated for different channels [3,4]. IEEE 1609.4 for multichannel operation, from the control channel (CCH) to the six service channels (SCH), which are utilized in time multiplexed fashion, with the control channel being served every other timeslot [3,4]. There is of the time slots are used by the different service channels, depending on the actual application requirements [3,4].

The IEEE 802.11p uses an Enhanced Distributed Channel Access (EDCA) MAC sublayer protocol designed based on that of the IEEE 802.11e with some modifications to the transmission parameters [5]. The physical layer of the IEEE 802.11p is similar to that of the IEEE 802.11a standard [5]. The IEEE 802.11p supports the transmission rates ranging from 3 to 27 Mb/s (payload) over a bandwidth of 10 MHz, which is half of the bandwidth in 802.11a [5]. The IEEE 802.11p aims to provide both V2V and V2I communications in ranges up to 1000 m in a variety of environments (e.g., urban, suburban, rural, and motorway) with relative vehicle velocities of up to 30 m/s [5]. Considering the fast movement and frequent trajectory changes in vehicular ad-hoc networks (VANETs), on the MAC sublayer, frequent handshakes and authorization are expected to be limited to reduce the high rate of link disconnections [5].

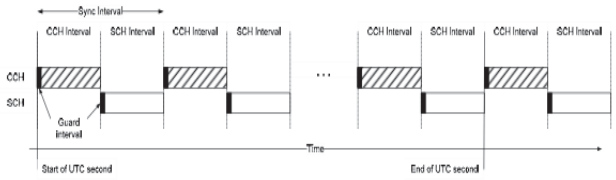


Fig. 1. Syn Interval, guard interval, CCH interval, and SCH interval [1,2]

The problem that is being addressed in this paper is to evaluate the channel performance of the service differentiation capability of the IEEE1609.4/802.11p using Enhanced Distributed Channel Access (EDCA) at the MAC sublayer is designed based on the IEEE 802.11e with some modifications to the transmission parameters. The EDCA proposed in IEEE 802.11e [13] is designed for contention-based prioritized QoS support. The EDCA mechanism defines four ACs that provide support for data traffic with four priorities [5,13]. Each AC queue works as an independent DCF station (STA) with enhanced distributed channel access function (EDCAF) to contend for Transmission Opportunities (TXOP) using its own EDCA parameters [5,13].

The contribution of this paper are to evaluate the performance of the service differentiation capability of the IEEE1609.4/802.11p, in terms of throughput and queuing delay and also to evaluate the channel performance in IEEE 802.11p/1609.4 standard using Enhanced Distributed Channel Access (EDCA) at the MAC sublayer is designed based on the IEEE 802.11e with some modifications to the transmission parameters.

This paper is organized as follows. In section 2, we provide arelated work and motivation. In Section 3,we provide a scenario and simulation for the performance of MAC Sublayer IEEE 1609.4/80211p. We also evaluate the performance of the service differentiation capability of the IEEE1609.4/802.11p. Finally, we conclude the paper and suggest the future work in Section 4.

II. RELATED WOK

Perdana, D. et al. [3] perform study the performance of the IEEE 1069.4 based on multi-hop disseminationin vehicular network to get SCH and CCH channel performanceusing Enhanced Distributed Channel Access Function (EDCA). We use [3] to simulate and evaluate the performance of Multi-channel operations of the IEEE 1609.4 based on channel perfomance using EDCA implementation.

Ali J.Ghandour, et al. [4] perform modeling and simulation of the protocol IEEE 1096.4 by contributing to improvethe packet delivery ratio and delay to the safety application and solve problems early in the transmission slot synchronization CCH/SCH. We use [4] to simulate and evaluate the performance of Multi-channel operations of the IEEE 1609.4.

Han, Chong et al. [5] proposed an analytical model for the performance of analysis of the IEEE 802.11p MAC, which helps researchers to be confident about the accuracy of simulators and proceed to do further research on IEEE 802.11p. In this paper [5] has indicated that 802.11p provides effective service differentiation mechanism that is suitable for mission-critical ITS applications.We use [5] to evaluate the performance of the service differentiation capability of the MAC Sublayer IEEE1609.4/802.11p, in terms of throughput and queuing delay, compared to that IEEE 802.11e.

Han, Chong et al. [5] proposed an analytical model for the throughput of the Enhanced Distributed Channel Access (EDCA) mechanism in IEEE 802.11p MAC Sublayer.The analytical model is suitable for both basic access and the Request-To-Send/Clear-To-Send (RTS/CTS) access mode.In this paper [6], the proposed analytical model is validated against simulation results to demonstrate its accurancy.

Grafling S. et al. [7] evaluate the performance of the IEEE 1609.4 and IEEE 802.11p based on the traffic prioritization schemes selected for the standards, and simulate in presence of multi-channel operation implemented by the IEEE 1069.4 the delay control messages and defined the QoS priority requirement standard in multi-channel operation.

One of most challenging issues in IEEE 1609.4/802.11p is the performance analysis in the sublayer Medium Access Control (MAC) IEEE 1609.4. Most of research on IEEE 1609.4/802.11p MAC Sublayer focused on methods in achieving QoS requirements (i.e. to improve packet delivery ratio and reduce delay) and increasing saturated throughput. On the other hand, most of existing work above does not include evaluate and study the performance of the service differentiation capability based on the channel performance, i.e SCH channel performance of the IEEE1609.4/802.11p.

In this paper, we evaluate the throughput and delay performance with EDCA implementation on the IEEE MAC sublayer 1609.4/802.11p based on simulation. Our work also evaluate the performance of the service differentiation capability of the IEEE1609.4/802.11p, in terms of throughput and queuing delay, compared to that IEEE 802.11e. On this work, we use ns 2.34 simulator to evaluate the performance of the MAC Sublayer IEEE 1609.4.

III. SIMULATION

A. Scenario

By using ns 2.34 simulator [8], we evaluate the performance of MAC Sublayer IEEE 1609.4/80211p based on channel performance. We also evaluate the performance of the service differentiation capability of the IEEE1609.4/802.11p. We simulate the scenario with the payload size of 512 bytes, the bitrate 6 Mbps, and the number of cars is 60. The channel configuration using default values for control and service channel intervals are 50ms, and guard intervals value is 4 ms. Meanwhile, Enhanced Distributed Channel Access Function (EDCA) values using EDCA parameter standard for IEEE 802.11p.

Table I. presents all parameters used in our simulation.While some parameters stay fixed, others are varied in order for us to observe the changing behavior of the network.

TABLEI.SIMULATION PARAMETERS

Parameters	Values
Numberofvehicles(max)	60
Numberofchannels(K)	3
Channel data rate (R)	6 Mb/s
Payload size	512 bytes
CCH slot duration	50ms
SCH slot duration	50ms
Guard Interval(GI)duration	4ms
Simulation time	20 seconds
AC0 Cw_{min}	3 time slots
AC0 Cw_{max}	15 time slots
AC1 Cw_{min}	3 time slots
AC1 Cw_{max}	14 time slots
AC2 Cw_{min}	15 time slots
AC2 Cw_{max}	1023 time slots
AC3 Cw_{min}	15 time slots
AC3 Cw_{max}	1023 time slots

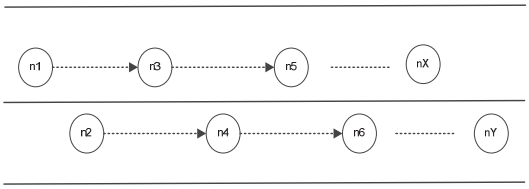


Fig.2. Path of forwading data from source to destination on Single-hop Dissemination [3]

Fig. 2 shows the single-hop dissemination. The path of forwarding data from source to destination has a direct route to destination based on Multi-channel operations IEEE 1609.4 [3].

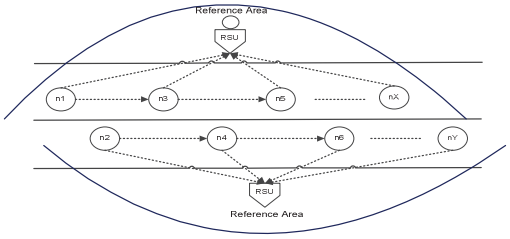


Fig. 3. The simulation scenario [5]

In our work, we use implementation of 802.11e EDCA [10] in NS-2 from TKN group in Technical University of Berlin. Some original simulation codes have been changed according to draft standard IEEE 802.11p [5,11] in our work, such as the AIFS and CW parameters. To analyze the performance of MAC sub-layer in 802.11p/IEEE 1609.4 standard, we configure the priority queue based on service differentiation. This paper using single-hop dissemination and some reference area to evaluate the performance of MAC Sublayer 802.11e and 802.11p/IEEE 1609.4, i.e the average delay, nmubers of packet delivered, and system throughput performance. We also analyze based on channel performance, i.e CCH and SCH channel performance in IEEE 1609.4 standard. Based on Han, Chong et al. [5], the reference area is a circular area around the RSU, with a radius which equals to the transmission range. Vehicles enter the reference area of the RSU in equal intervals, which means the distance between two adjacent vehicles are fixed in each scenario [5]. The number of vehicles in the reference area is controlled by changing the distance among vehicles [5]. When a vehicle enters communication range of the RSU, it initiates a Constant Bit Rate (CBR) data transmission to the RSU [5].

B. Performance Evaluation

Base on the scenario, in Figure 4 until Figure 8 the simulation was performed to obtain data according to four aspects to be measured. They are the average delay, packet delivery ratio, and throughput.

1) Performance comparison of EDCA 802.11e and IEEE 1609.4

a) Average Delay

Fig. 4 shows the comparison of the average delay of 802.11p/IEEE 1609.4 compared with 802.11e standard, by varying the number of nodes based on EDCA implementation.

We focus on the average access delay which calculate on MAC layer. Delay and access will be used interchangeably on this work by varying the number of nodes. This can be seen Fig. 4.

The following is the equation for the average delay $E[\delta]$ derived as [4] :

$$E[\delta] = E[\chi] + E[\theta] \quad (1)$$

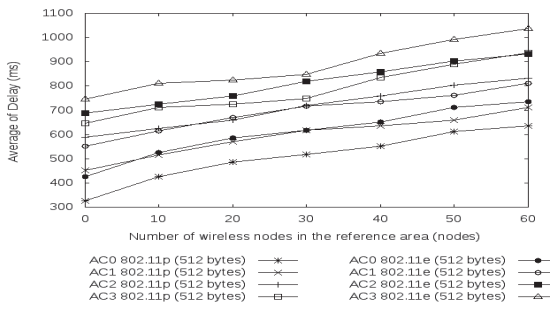


Fig. 4. Comparison of the average delay EDCA 802.11p/IEEE 1609.4 and 802.11e standard

From Fig. 4, we found that the average delay of AC3 (CW*min* = 15 and CW*max* = 1023) based on 802.11e standardis the highest average delay compared with 802.11p/IEEE 1609.4 standard. We can analyze the average delay performance of the service differentiation capability of the IEEE1609.4/802.11p using Enhanced Distributed Channel Access (EDCA) at the MAC sublayer is designed based on the IEEE 802.11e with some modifications to the transmission parameters. In addition, we fix the bugs of the RTS/CTS mode to make successful transmission using the RTS/CTS mechanism [5]. We use this model to analyze the performance of the MAC sublayer in IEEE 802.11p standard and also analyze the channel performance IEEE 802.11p standard. Based on that some modifications and fix the bugs RTS/CTS mode to make successful transmission, we get the average delay in802.11p/IEEE 1609.4 standard is better than IEEE 802.11e standard.

b) Throughput

Fig. 5 shows throughput of 802.11e compared with 802.11p/IEEE 1609.4 standard, by varying the number of nodes based on EDCA implementation. Throughput T*i* (t) is the rate of successful packet delivery through a network connection per unit time. We focus on the throughput which calculate at MAC layer, then T*i* (t) derived as [3,14] :

$$Throughput\ T_i(t) = x^*(1-p)^*d^*data\ rate \quad (2)$$

Where $d = DATA/(DIFS+PACKET+SIFS+ACK)$
xis the number of nodes
T*i* (t) is the throughput
a is the distance of nodes
p is the collision probability for a transmission

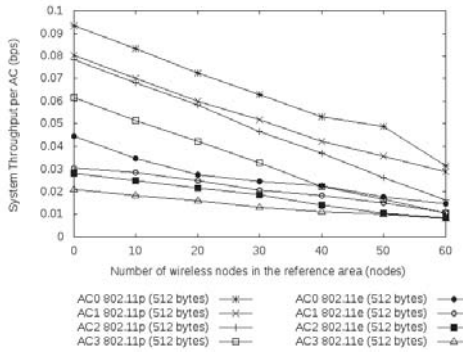


Fig. 5. . Comparison of the system throughput EDCA 802.11p/IEEE 1609.4 and 802.11e standard

From Fig. 5, we found that the throughput of AC0 (CW*min* = 3 and CW*max* = 15) based on 802.11p/IEEE 1609.4 standard is the highest performance compared with 802.11e standard. We evaluate the throughput performance of the service differentiation capability of the IEEE1609.4/802.11p using Enhanced Distributed Channel Access (EDCA) at the MAC sublayer is designed based on the IEEE 802.11e with some modifications to the transmission parameters. Some parts of the original simulation codes have been changed according to the draft standard IEEE 802.11p [11] in our work, such as the AIFS and CW parameters. In addition, we fix the bugs of the RTS/CTS mode to make successful transmission using the RTS/CTS mechanism [5]. Based on that, we get the throughput performance in 802.11p/IEEE 1609.4 standard is better than IEEE 802.11e standard.

c) Numbers of Packet Delivered

Fig. 6 shows the numbers of delivered packets of the 802.11p/IEEE 1609.4 compared with 802.11e standard, by varying thenumber of using EDCA implementation. The numbers of delivered packets is defined by varying number of nodes as parameters change. We focus on the numbers of delivered packets which calculate on MAC layer.

Let τ be the probability to transmit in a given slot. If we assume a uniform probability distribution to select a slot within the current *Contention Window* (CW), then τ can be derived as [4,9]:

$$T = 1/E[CW]+1=2/CWmax+1 \quad (3)$$

Where *CWmax* is the maximum size of *CW* for broadcast messages. Let p_{idle} be the probability that a channel is idle in a given slot, and p_{busy} its converse [4]. Similarly, let $p_{success}$ be the probability that a slot is occupied by a successful transmission, and p_{coll} is the probability that a collision occurs during a slot.

If we assume a scenario with M nodes with the above mentioned assumptions, it is easy to verify that p_{idle}, p_{busy}, $p_{success}$, and p_{coll} are as followed [4]

$$P\ idle = (1-\tau)\ M \quad (4)$$
$$P\ busy = 1-P\ idle \quad (5)$$
$$P\ success = M\cdot\tau\cdot(1-\tau)M-1 \quad (6)$$

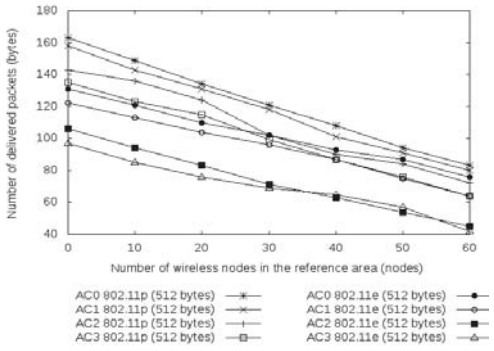


Fig.6. Comparison of the numbers of delivered packetsEDCA 802.11p/IEEE 1609.4 versus 802.11e standard

From Fig.6, we found numbers of delivered packets of the 802.11p/IEEE 1069.4 standard with EDCA implementation have AC0 (CW*min* = 3 and CW*max* = 15) with the highest performance compared with 802.11e standard. We evaluate the delivered packets performance of the service differentiation capability of the IEEE1609.4/802.11p using Enhanced Distributed Channel Access (EDCA) at the MAC sublayer is designed based on the IEEE 802.11e with some modifications to the transmission parameters. In addition, we fix the bugs of the RTS/CTS mode to make successful transmission using the RTS/CTS mechanism [5]. We use this model to analyze the performance of the MAC sublayer in IEEE 802.11p standard and also analyze the channel performance IEEE 802.11p standard. Based on that, we get the numbers of delivered packets performance IEEE1609.4/802.11p using Enhanced Distributed Channel Access (EDCA) better than IEEE 802.11e standard.

2) Performance comparison of channel performance based on EDCA 802.11p/IEEE 1609.4 standard

a) Average Delay

Fig.7 shows the average delay of 802.11p/1609.4 standard with EDCA implementation based on SCH channel performance.

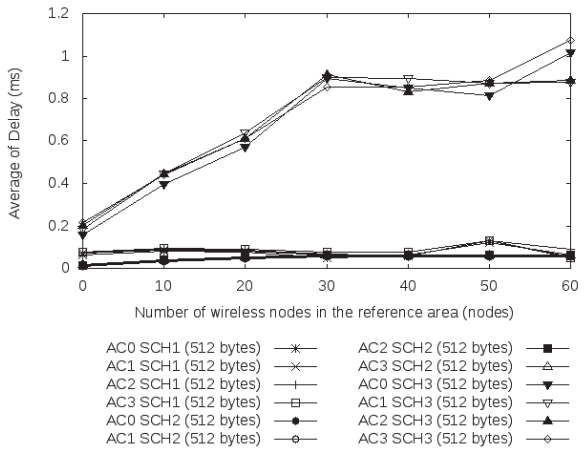


Fig.7. Comparison of the average of delay based on the SCH channel performance

From Fig.7, we found that the average delay of the 802.11p/IEEE 1069.4 standard with EDCA implementation have AC1 in SCH3 channel (CW*min* = 3 and CW*max* = 14) with the highest performance based on SCH channel performance. We evaluate the average delay based on SCH Channel performance of the service differentiation capability of the IEEE1609.4/802.11p using Enhanced Distributed Channel Access (EDCA) at the MAC sublayer with some parts of the original simulation codes have been changed according to the draft standard IEEE 802.11p [11] in our work, such as the AIFS and CW parameters.

b*) Numbers of Packet Delivered*

Fig.8 shows the numbers of delivered packets 802.11p/IEEE 1609.4 compared with 802.11e standard based on SCH channel performance, by varying the number of node using EDCA implementation. We focus on the numbers of delivered packets which calculate on MAC layer.

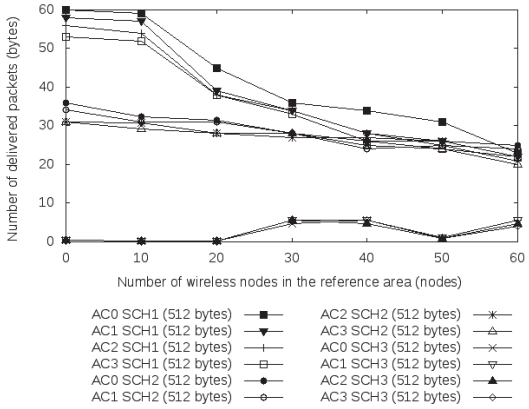


Fig.8. Comparison of the numbers of delivered packets based on SCH channel performance

From Fig.8, we found that numbers of delivered packets of the 802.11p/IEEE 1069.4 standard with EDCA implementation have AC0 in SCH1 channel (CW*min* = 3 and CW*max* = 15) with the highest performance based on SCH channel performance. We evaluate numbers of delivered packets based on the SCH Channel performance of the service differentiation capability of the IEEE1609.4/802.11p using Enhanced Distributed Channel Access (EDCA) at the MAC sublayer with some parts of the original simulation codes have been changed according to the draft standard IEEE 802.11p [16] in our work, such as the AIFS and CW parameters.

c) Throughput

Fig.9 shows the throughput of 802.11p/IEEE 1609.4 based onthe SCH channel performance, by varying the number of nodesbased on EDCA implementation.

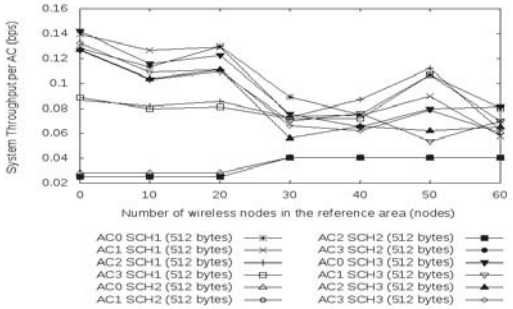


Fig.9 Comparison of the system throughput based on SCH channel performance

From Fig.9, we found that throughput of the 802.11p/IEEE1069.4 standard with EDCA implementation

have AC1 in SCH1 channel (CW*min* = 3 and CW*max* = 14) with the highest performance based on SCH channel performance.We evaluate the throughput based on SCH Channel performance of the service differentiation capability of the IEEE1609.4/802.11p using Enhanced Distributed Channel Access (EDCA) at the MAC sublayer with some parts of the original simulation codes have been changed according to the draft standard IEEE 802.11p [11] in our work, such as the AIFS and CW parameters. In addition, we fix the bugs of the RTS/CTS mode to make successful transmission using the RTS/CTS mechanism [5].

IV. CONCLUSION

We have evaluated the performance of the EDCA 802.11p/1609.4 compared with EDCA 802.11e based onthe service differentiation capability.We conclude that the performance of EDCA 802.11p/1609.4 based on the service differentiation capability the average delay has been degraded 13.87% compared with EDCA 802.11e, and the throughput is 57.49% better than the EDCA 802.11e standard.Based on SCH channel performance 802.11p/IEEE 1609.4,we have AC1 in SCH3 channel (CW*min* = 3 and CW*max* = 14) with the lowest average delay performance, also we have AC0 in SCH1(CW*min* = 3 and CW*max* = 15) with the highest number of delivered packets performance. Otherwise, the throughput performance based on SCH channel performance, we have AC1 in SCH1 channel (CWmin = 3 and CWmax = 14) with the highest performance.

REFERENCE

[1] Jafari, A., Performance Evaluation of IEEE 802.11p for Vehicular Communication Networks. PhD Dissertation, Faculty of Arts, Computing, Engineering, and Sciences, Postgraduate Program, Sheffield Hallam University, South Yorkshire, England, UK, 2011.

[2] IEEE Std 1609.4-2010, IEEE Standard for Wireless Access in Vehicular Environments (WAVE) Multichannel Operation, IEEE, 2011.

[3] Perdana, D. Sari, RF (2013), "Performance Evaluation of Multi-Channel Operation IEEE 1609.4 Based On Multi-hop Dissemination", International Journal of Computer and Network Security (IJCSNS), VOL.13 No.3, March 2013, pp. 32-41.

[4] Ali J. Ghandour, Marco Di Felice, Hassan Artail and Luciano Bononi, "Modeling and Simulation of WAVE 1609.4-based Multi-channel Vehicular Ad Hoc Networks", 5th ACM International Conference on Simulations Tools and Techniques (SIMUTools 2012), March 19-23, 2012, Sirmione-Desenzano, Italy.

[5] Chong, H., Mehrdad D., Rahim T., Ralf K., Xuemin S., Analytical Study of the IEEE 802.11p MAC Sublayer in Vehicular Networks. Intelligent Transportation Systems, IEEE Transactions on, 2012. 13(2): p. 873-886.

[6] Chong, H., , Mehrdad D., Rahim T., Ralf K., Xuemin S., Throughput Analysis of the IEEE 802.11p Enhanced Distributed Channel Access Function in Vehicular Environment. in Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd. 2010.

[7] Sebastian Grafling, Petri Mahonen, and Janne Riihijarvi. Performance Evaluation of IEEE 1609 WAVE and IEEE 802.11p for Vehicular Communications. In *International Conference on Ubiquitous and Future Networks*, pages 344-348, 2010.

[8] The networksimulator-2,http://www.isi.edu/nsnam/ns/

[9] F.Cali', M.Contiand E.Gregori. Dynamic Tuning of the IEEE 802.11Protocol to Achieve a Theoretical Throughput Limit. IEEE/ACM Transactions on Networking, 8(6),pp. 785-799,2000.

[10] S. Wietholter, M. Emmelmann, C. Hoene, and A. Wolisz, "TKN EDCA Model for ns-2. "Telecommun. Netw. Group, Technische Univ. Berlin, Berlin, Germany, Tech. Rep. TKN-06-003, Jun. 2006.

[11] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amandement 7 : Wireless Access in Vehicular Enviroments*, IEEE Unapproved Draft Std. P802.11p/D8.0, Jul. 2009.

[12] Qi Chen, Daniel Jiang, and Luca Delgrossi,IEEE 1609.4 DSRC Multi-Channel Operations and Its Implications on Vehicle Safety Communications.Vehicular Networking Conference (VNC), 2009 IEEE, 28-30 Oct. 2009, pp.1-8.

[13] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements, IEEE Std. 802.11e, Nov. 2005.

[14] Jonathan C.L. Liu, Ph.D. Paper#4 : Overload Issues in 802.11b Ad-Hoc Network. Lecture Slides. Department of Computer, Information Science and Engineering (CISE), University of Florida

[15] Inan, I., F. Keceli, and E. Ayanoglu, Analysis of the 802.11e enhanced distributed channel access function. Communications, IEEE Transactions on, 2009. 57(6): p. 1753-1764.

Performance Evaluation of ZigBee-based Wireless Sensor Network for Monitoring Patients' Pulse Status

Muhammad Niswar, Amil Ahmad Ilham, Elyas Palantei, Rhiza S. Sadjad, Andani Ahmad, Ansar Suyuti, Indrabayu, Zaenab Muslimin, Tadjuddin Waris, Puput Dani Prasetyo Adi

Department of Electrical Engineering,
Faculty of Engineering, Hasanuddin University
Makassar, Indonesia

Abstract— **Recently, the wireless sensor network has been widely deployed for medical care purpose. We have developed a wireless sensor network that can monitor the patients' pulse status for triage purpose, so that a medical team can monitor remotely the health condition of patient and they can treat the patient based on severity of patients' health condition. We developed an electronic triage operating as a sensor node (SN) tagged in patient's arm. The SN consists of microcontroller ATmega328P, ZigBee and pulse sensor to detect patient's pulse. Operating as an electronic triage, the pulse rate from sensor is classified into three categories of severe conditions, i.e., major, minor, and normal status by the microcontroller in SN and sent to the coordinator node (CN) through ZigBee interface. Our system can be deployed in emergency room, triage room, pre/post-op surgery in hospital as well as in disaster area. This paper aims to evaluate the performance of the ZigBee-based wireless sensor network that we developed. We evaluate the effective distance between CN and SN to deliver patients' pulse rate via ZigBee as well as the effective number of SNs that can be accommodated by single CN. The experimental results shown that the effective distance between CN and SN to deliver the pulse rate data is less than 30 meters and the maximum number of SNs can be accommodated by a single CN is 3 (three) nodes.**

Keywords— ZigBee, Wireless Sensor Network, Microcontroller, Pulse Sensor

I. INTRODUCTION

Application of wireless sensor network technology for medical purpose has been widely deployed especially in medical emergency response. In emergency situation, medical team usually applies a START [1] triage protocol to determine the priority of patients' treatments based on the severity of their condition. The START protocol triage the patient based on 3 (three) factors, i.e., respiration, pulse and mental status. According to the condition of these factors, patient will be classified into 4 (four) priorities of patients' treatment, i.e., immediate, urgent, delayed and deceased. In this study, we have developed an electronic triage to detect the pulse rate of a patient. This electronic triage is tagged in patient arm and plays as a sensor node (SN) attached in ZigBee-based wireless sensor network. Since our electronic triage can only detect the pulse status of the patient, therefore, this paper is focused on the triage of patients based on their pulse status only. In the near future, we will add a respiration sensor in our electronic triage to fulfill the requirement of START triage protocol. Furthermore, our electronic triage should be used after medical team has sort the patients or victims based on four priority categories (primary triage) and then tagged to the patient categorized as delayed and urgent, i.e., the patient that need a continuous monitoring and support from medical team. Therefore, our electronic triage is suitable for secondary and tertiary triage and can be deployed in emergency room, triage room, pre/post-op surgery in hospital and in disaster area.

This study intends to evaluate the performance of our ZigBee-based wireless sensor network for monitoring the pulse status of patients so that medical team can monitor continuously the patients' health condition of patient and they can treat the patient based on severity of his/her condition. This paper examines the performance of wireless sensor network in terms of the effective distance for delivering patients' pulses from electronic triage operating as an SN to the coordinator node (CN) as well as the effective number of SN that can accommodated by single CN.

This research was supported by the InSINas Grant from Ministry of Research and Technology, Indonesia .

II. RELATED WORKS

Many researchers have been proposed a network architecture for remote health monitoring using wireless sensor network (WSN). Reference [2] proposed a three-layer architecture: Body Sensor Layer, Personal Network Layer and Global Network Layer. Reference [3] discussed the implementation issues and described the overall system architecture for a bluetooth sensor network for patient monitoring and corresponding heart activity sensors. Reference [4] proposed a new low-cost, low-power wireless sensor platform implemented using the IEEE 802.15.4 wireless standard, and describe the design of compact wearable sensors for long-term measurement of electrodermal activity, temperature, motor activity and blood volume pulse. Reference [5] proposed a wireless infrastructure for emergency medical care system using low-power, wireless vital sign sensors, PDAs, and PC-class systems (Code Blue project). However, all proposed system described the overall system and there is no detail description about the performance evaluation of wireless sensor network. In this study, we developed an electronic triage operating as an SN to detect patients' pulse rate and we evaluated network performance between SN and CN in delivering pulse rate.

III. SYSTEM DEVELOPMENT

We developed an electronic triage that can monitor patients' pulse status remotely through ZigBee-based wireless sensor network. The system consists of a number of SNs, a CN collecting the pulse rate from SNs through ZigBee wireless interface, and a web interface for displaying the graph of patients' pulse rate. Our electronic triage, which is basically an SN, consists of pulse sensor [6], microcontroller Atmega 328P and ZigBee wireless interface as shown in figure 1. The electronic triage is placed in patients' wrist (see Figure 2) and the pulse sensor bind the patients' finger to detect the patients' pulse. The microcontroller reads the patients' pulse from pulse sensor, then classifies into 3 (three) categories of severity levels, i.e., major, minor, and normal status. The patients' pulse status is displayed in unit of beat per minute (BPM) on 16x2 LCD. Three LED colors indicate the severity level of patient's status (major: red, minor: yellow and normal: green). If pulse status is between 60-100 BPM, then green LED indicator will turn ON indicating that patient is in good health condition and classified as in the normal status. If the pulse status is between 44 and 60 BPM or between 100 and 116 BPM, then yellow LED indicator will turn ON to indicate that the patient is in close to critical health condition and classified as in the minor status. Otherwise, red LED indicators will turn ON to indicate that the patient in a critical health condition and classified as in the major status. The pseudo-code 1 shows the algorithm to classify the patients' severity level embedded in microcontroller.

After classifying the patients' pulse status, the electronic triage will send the pulse status to CN through the ZigBee wireless interface. We use XBee Series 2 [7] wireless interface (ZigBee-based protocol) operating in 2.4 GHz frequency band. The XBee series 2 supports point-to-point, point-to-multipoint, and mesh communication .It is specified to handle a data transmission rate up to 250 Kbps using 2 mW transmit power that can achieve a distance range up to 120 meters. The CN sends the information of patients' pulse status to web server. The information is displayed in web interface for allowing medical team to monitor and assess the patients' condition remotely from anywhere using internet browser. We can restrict the access to the web interface by allowing only the medical team to view this web page. Figure 3 shows the our overall developed system

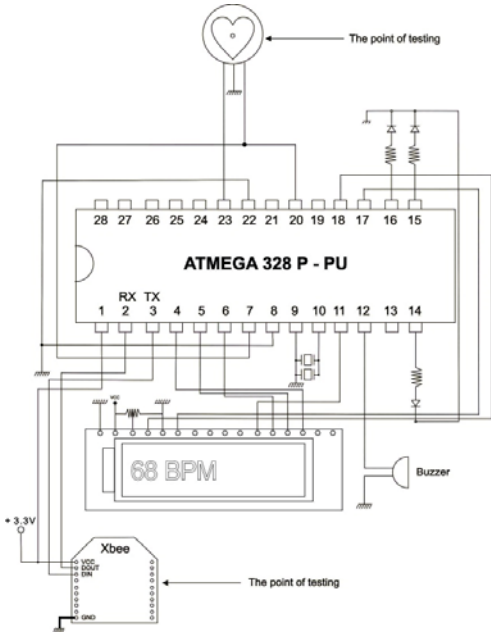


Figure 1. Sensor Node Schematic Diagram

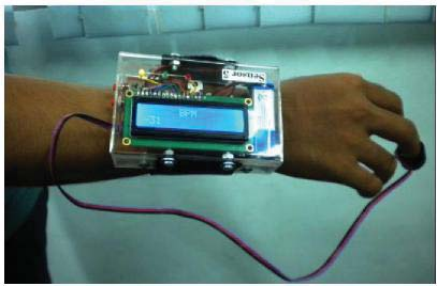


Figure 2. Pulse Sensor Node

```
Begin
inData f- Read(Port)
While(inData=”\n”)
{
        if(inData=’P1/P2/P3/P4 or PS’) then
                BPMf-int(inData)
        endIf
        if(inData=’B’) then
                HRVf-int(inData);
HeartBeat(On)
        endIf
}
InData (BPM)
Show(BPM)
If(BPM>59 and BPM<101) then
        Red_Indicator(Off)
        Yellow_Indicator(Off)
        Green_Indicator(On)
Else
If(BPM>44 and BPM<60) or (BPM>100 and
BPM <116) then
        Red_Indicator(Off)
        Yellow_Indicator(On)
        Green_Indicator(Off)
Else
        Red_Indicator(Off)
        Yellow_Indicator(On)
        Green_Indicator(Off)
EndIf
End
```

Pseudo-code 1. Algorithm for classification of patients’ pulse status

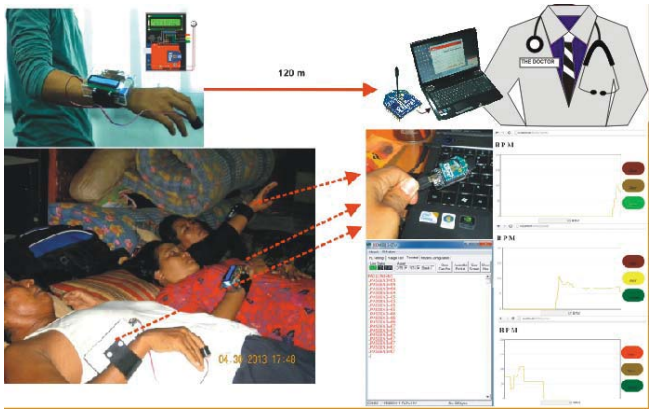


Figure 3. Communication between SN and CN

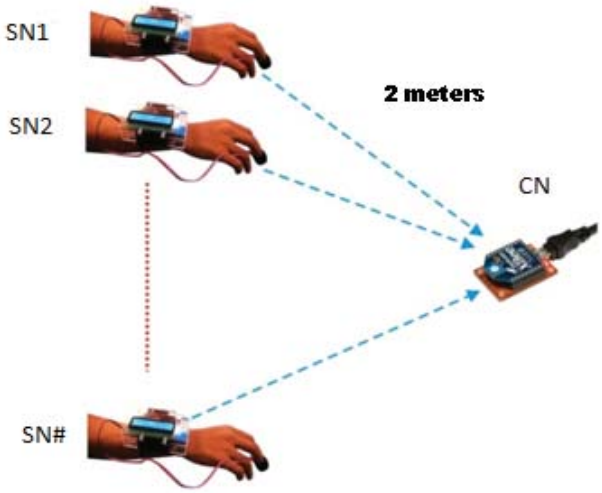


Figure 4. Experiment Scenario

IV. PERFORMANCE EVALUATION

We conducted experiments to evaluate the performance of our ZigBee-based wireless sensor network. In this study, we only evaluate point-to-point and multipoint communication modes. There are two scenarios in these experiments. First scenario aims to determine the effective distance between the CN and the SN to deliver pulse rate data with no or less packet loss. Second scenario aims to determine the maximum number of SNs that can be accommodated by a single CN. We use packet loss rate in unit of percentage as our performance indicator. The packet loss rate is calculated as the number of packet received by the CN devided by the number of packet transmitted by SN. The wireless sensor network performance is better if less packet loss rate occurs in CN.

In the first scenario, we located a SN starting from 10 to 120 meters (increasing by ten meters) away in the free space to find the effective distance between the CN and the SN. In this experiment, SN sends the 7 bytes of data (pulse rate) every 10 seconds to CN in point-to-point communication mode. We assumed that the tolerated packet loss should be less than 5% packet loss. The experiment result in figure 5 shows that the packet loss increases as the distance between the CN and the SN is increasing. From figure 5, we can see that packet loss reaches more than 5% at distance of 30 meters. Therefore, we can conclude that the effective distance between the CN and the SN to deliver the pulse rate data is less than 30 meters.

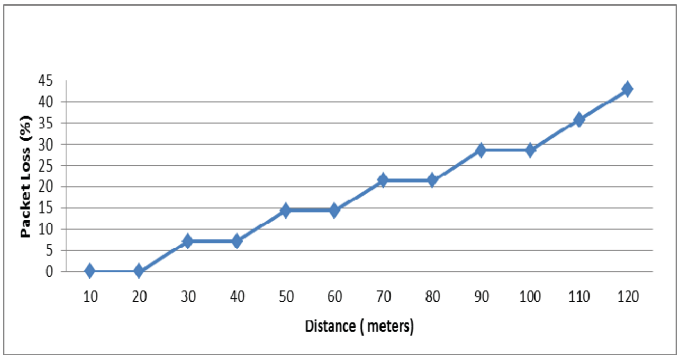


Figure 5. Packet loss vs. Distance between CN and SN

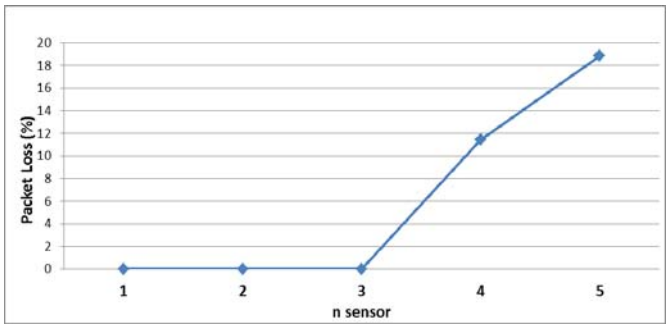


Figure 6. Packet loss vs. Number of SNs in the single CN

In the second scenario, we located an SN closed to the CN (2 meters) and let the SN send the pulse rate data to CN. We then increased the number of SN up to 5 nodes while they are sending the data to CN as shown in figure 4 (point-to-multipoint communication mode). From figure 6, we can see that the packet loss occurs after the number of SN reaches 4 (four) nodes. Therefore, we can conclude that the maximum number of SNs that can be accommodated by single CN is 3 (three) nodes.

V. CONCLUSION AND FUTURE DIRECTION

We developed a ZigBee-based wireless sensor network that can monitor patients' pulse status remotely. This paper presents the system development and performance evaluation of our ZigBee based wireless sensor network for monitoring patients' pulse rate. We conducted experiments to determine the effective distance between the CN and the SN that can deliver patients' pulse rate through ZigBee interface. We also evaluated the maximum number of SNs that can be accommodated by a single CN. Experimental results have shown that effective distance between the CN and the SN to deliver the patients' pulse rate is less than 30 meters and the maximum number of SNs that can be accommodated by a single CN is 3 (three) nodes. Since most data sent by SNs are classified as normal status, meaning that most patients are in good condition, we can reduce the number of pulse rate data by allowing SNs to send only pulse rate data that classified into major and minor status, i.e., when patients' pulse status are in critical and closed to critical condition. By reducing the number of the pulse rate data, the CN can accommodate more SNs.

In the near future, we will evaluate the performance of our ZigBee-based wireless sensor network in mesh communication mode. In mesh communication, every SN plays a role as router node, so, it can pass the information to another SN to reach the CN. Hence, an SN can be located even if it is out of the effective distance range from the CN.

.

REFERENCES

[1] Super G, et al. "START: Simple Triage and Rapid Treatment Plan," Newport Beach, CA: Hoag Memorial Presbyterian Hospital, 1994.

[2] Kuryloski, P, et al, "DexterNet: An Open Platform for Heterogeneous Body Sensor Networks and its Applications", Body Sensor Networks 2009 (BSN 2009 Berkeley, CA June 3-5, 2009)

[3] Y. Zhang and H. Xiao, Bluetooth-based sensor networks for remotely monitoring the physiological signals of a patient, WPAN Transactions on Information Technology in Biomedicine, vol. 13, no. 6, pp. 10401048, 2009

[4] Fletcher, R.R.; Ming-Zher Poh; Eydgahi, H.; , "Wearable sensors: Opportunities and challenges for low-cost health care," Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the WPAN , vol., no., pp.1763-1766, Aug. 31 2010-Sept. 4 2010

[5] D. Malan, T. Fulford-Jones, M. Welsh, and S. Moulton, CodeBlue: An ad hoc sensor network infrastructure for emergency medical care, Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks, *pp.* 12-14, June 6, 2004.

[6] _____: Pulse Sensor, Online available at http://pulsesensor.com/

[7] _____: Xbee Series 2 OM RF Modules, Online available at ftp://ftp1.digi.com/support/documentation/90000866_A.pdf

Study and Design of the Video for Resource Limited Mobile Communication

Andik Setyono
Department of Informatics Engineering
Faculty of Computer Science, UDINUS
Semarang, Indonesia
setyonoandik@gmail.com

Sendi Novianto
Department of Informatics Engineering
Faculty of Computer Science, UDINUS
Semarang, Indonesia
to_sendi@yahoo.com

***Abstract*— Multimedia files such as image, audio, and video are in general very large in size. The limitations of the current mobile communications technology with respect to media storage and bandwidth make transmission of large-size multimedia files very difficult. In this paper, we explore the use of compression technique along with video coding for designing a video transmission framework which can be applied in mobile communication networks such as GPRS or UMTS which have limited bandwidth. We develop applications for video transmission on a mobile client-server system using streaming technique. We present the simulation and experiment results of the video streaming process implemented in a peer-to-peer environment using video coding with different protocols such as HTTP, RTP and RSTP. The proposed framework is then applied to develop a mobile telemedicine system.**

Keywords*— *video coding; compression; video streaming; transmission method

I. INTRODUCTION

Nowadays, Multimedia communication such as listening music, watching television, browsing internet, playing movie and calling telephone has become part of our daily life. Multimedia is a combination of several media such as text, graphics, audio, video and animation which is used to produce stunning presentations [1]-[2]. Multimedia communication is highly interactive. For computer users, multimedia can be defined as computer information that can be presented via audio, video, text, graphics and animation [3]. Multimedia helps the users to understand the information more easily. Multimedia information is more effective because it is easy for people to remember by seeing, hearing and performing than by reading the text only. Currently, multimedia devices are growing very rapidly. This is because information using text cannot provide enough information compared to multimedia. One form of multimedia file is video which can display audio and visual together. The audio visual media can make abstract concepts more concrete [4].

At one time, we need to sit down at one place to watch a movie but now we can do it in mobile environment using mobile devices. In line with the rapid growth of information and communication technology, many tools have emerged in the market that can display multimedia using mobile device for mobile communications. Using mobile devices (e.g. mobile phones, PDAs and notebooks), we can perform many activities such as watching television news, movies and others in the mobile environment. Currently, many features that are provided in the mobile phone to exchange multimedia information enable us to access various Internet applications and streaming audio and video files [5].

Mobile technology can be used to develop mobile Internet applications such as mobile messaging, mobile telephone, mobile Internet and mobile network. It is possible to use mobile technology to transmit the video data for mobile communications [6]-[7]. There exist several techniques to upload the video data from the client to the server. We can use MMS or email client on mobile phone or desktop computer that is connected with Internet network as multimedia message. *File transfer protocol* (FTP) and *Hypertext transfer protocol* (HTTP) can also be used to transfer multimedia file from the client (mobile and desktop) to the server as input media. For retrieving video from server techniques such as downloading, progressive downloading, multicasting or streaming can be used.

In this paper, we focus on the study and design of the video transmission framework for resource-limited mobile communication, which specifically addresses the transmission methods in developing mobile client-server system application. The motivations for developing such a framework are: (i) there is a need to use multimedia data for mobile communications and (ii) there is a need to propose appropriate transmission techniques for mobile communications. This topic assumes importance due to the fact that video files are generally too large to be able to be transmitted using wireless, wired or mobile network, thus requiring the use of compression and decompression (video coding) techniques to reduce the file size without degrading the data quality. We propose in these paper suitable techniques to transmit large multimedia (video) files in a mobile environment with limited bandwidth.

The goals of this research are: (i) to explore video technology for mobile communications, (ii) to propose an adaptive video transmission framework for resource limited mobile communications, (iii) to propose mobile client-server system application and (iv) to evaluate and analysis the

proposed mobile video transmission framework. The remaining part of the paper is organized as follows: Section 2 discusses literature review. Proposed video transmission framework is presented in Section 3, simulation of multimedia streaming is presented in Section 4, Section 5 discusses performance and evaluation and conclusion and future work are presented in Section 6.

II. LITERATURE REVIEW

Digital video data tends to take up a large amount of storage or transmission capacity. Video coding is essential for any application in which storage capacity or transmission bandwidth is constrained [8]. Video coding is the process for compressing and decompressing a digital video signal. Multimedia files in general require large storage capacity. For example, a bitmap image without compression (BMP) with size 600 x 800 pixels will generate a file of 1.37 Mbytes. If it is compressed with a compression ratio of 16:1, the file size will be about 86 Kbytes [9]. For a high-definition television (HDTV) signal of size of 720 x 1280 pixels per frame with a scanning rate of 60 frames per second, the bit rate will be around 1.3 GB per second [9]. The HDTV has a channel bit rate of 20 Mbps, and hence a compression ratio of more than 70:1. Table 1 shows some examples of video applications and the required compression ratio. Maintaining the Integrity of the Specifications.

TABLE I. VIDEO APPLICATION AND COMPRESSION RATIO

Video Format	Pixels/Frame	Fps	Bits/ pixels	Raw bit rate	Channel bit rate	Required Compression Ratio
HDTV	720 x 1280	60	24	1.3 Gb/s	20 Mb/s (HDTV)	~ 70
SDTV	480 x 720	30	24	250 Mb/s	5 Mb/s (DVD)	~ 50
CIF	288 x 352	30	24	73 Mb/s	384 Kb/s (3G)	~ 190
QCIF	144 x 176	15	24	9 Mb/s	64 Kb/s (3G)	~ 140

We can see in Table 1, the transmission of the common intermediate format (CIF) video by using a 3G mobile network needs compression up to 190:1. This paper deals with different transmission methods that are suitable for large multimedia files.

A. Video Compression

Compression technique is the act or process of compacting data into smaller number of bits [10]. Video compression is the process of converting digital video into a format suitable for transmission or storage. It means that the compression is necessary for practical storage and transmission of digital video. Compression involves a complementary pair of systems, a compressor (encoder) and a de-compressor (decoder). The encoder converts the source data into a compressed form occupying a reduced number of bits, prior to transmission or storage, and the decoder converts the compressed form back to the original video data.

Data compression is achieved by removing the redundant components that are not necessary for faithful reproduction of the data. Many types of data contain statistical redundancy and can be effectively compressed using lossless compression, so that the reconstructed data at the output of the decoder is a perfect copy of the original data. Unfortunately, lossless compression of image and video information gives only a moderate amount of compression [11]. The best that can be achieved with lossless image compression standards such as JPEG-LS is a compression ratio of around 3-4 times. Lossy compression is necessary to achieve higher compression. In a lossy compression system, the decompressed data is not identical to the source data and much higher compression ratios can be achieved at the expense of a loss of visual quality. Lossy video compression systems are based on the principle of removing subjective redundancy which means removal of elements of the image or video sequence that can be removed without significantly affecting the viewer's perception of visual quality.

B. Video Transmission Methods

The video source is encoded first before it is transmitted to the destination receiver. The video that has been encoded can then be broadcast, or stored in a server or in a portable storage. Video transmission for broadcasts (such as digital television broadcasting TV programs) is encoded prior to transmission over a limited-bandwidth terrestrial, satellite or cable channels. For video that is stored on a server can be transmitted using a streaming method using the Internet where the coded video is transmitted (streamed) over the Internet, decoded and displayed on a client. The other method is to distribute the coded video by storing it in a portable storage such as DVD or other storage medium. A DVD player reads the disk and decodes video for display. We can also use video calling in which each participant includes an encoder and a decoder. Video from a camera is encoded and transmitted across a network, decoded and displayed. This occurs in two directions simultaneously. Fig. 1 shows the video coding distribution scenarios.

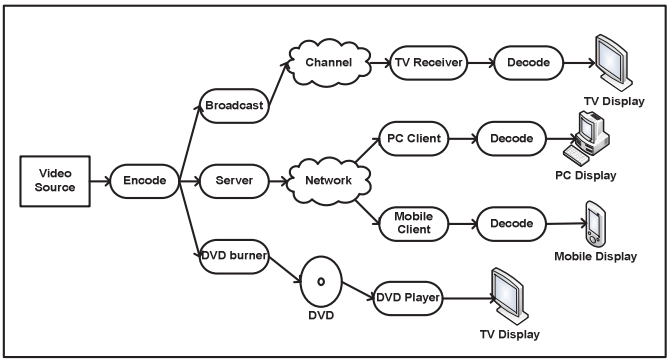


Fig. 1. The video coding distribution scenarios

In this paper, we focus on the discussion of video transmission by using servers and networks which can be accessed in static and mobile environments. We use the streaming method for transmission from server to client. Besides using streaming method for video retrieval, we can also download the coded video that has been uploaded to the server or we can also use the messaging system using the MMS and email. In this paper, we use streaming method for transmitting large video files.

Streaming media is an activity to distribute or disseminate a resource/multimedia material constantly online and in real-time

through a communications network from server to client using a streaming protocol. Online activities includes accessing a resource in a network where the data is always available which can be accessed at any time. Real-time operation indicates that access to the data source is done directly using a certain time delay that allows any new data that exist on the server to be directly received by the client. There exist some communication protocols that were developed along with streaming media needs such as *Real-time Streaming Protocol (RTSP), Real-time Transport Protocol (RTP), Microsoft Media Services (MMS), Hypertext Transfer Protocol (HTTP), and Real-time Messaging Protocol (RTMP)*. We describe the streaming concept that is implemented on the desktop computer or on the mobile computer. The streaming can use live, simulated live, multicast and store and forward methods. Fig. 2 shows the application of streaming applications in the communication network.

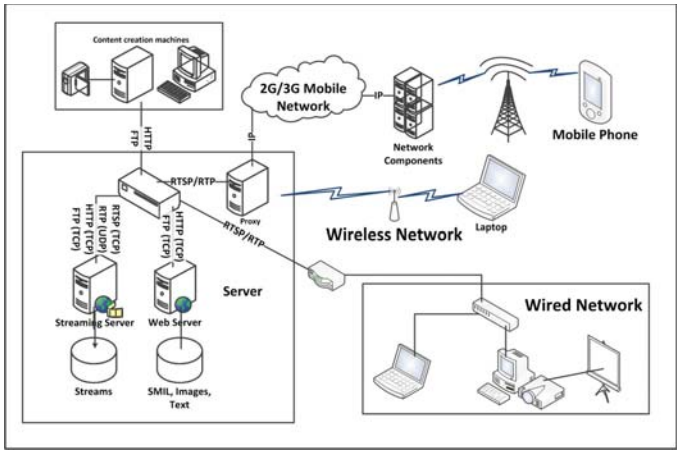


Fig. 2. Streaming application on the network

III. Proposed Video Transmission Framework

We propose a video transmission framework based on the existing streaming methods. There are four different methods of streaming: live, simulated live, interactive and store and forward. On the client side, a media player can be installed on desktop computers or mobile phones to run the streaming process.

A. Proposed Design of a Mobile Client-Server System

In this paper, we propose streaming method to implement mobile telemedicine system, especially for tele-consultation. On the server, there are many video files that can be retrieved by users on the client side using download and streaming methods. The video contents are about telemedicine information and some guidance to be followed for teleconsultation using this system. Users can get information about any disease and the curing methods via video that is available on the server using personal computer or mobile phone device. Streaming method is used to transmit video file for tele-education of patient. The streamable video file is already stored on the server. Video streaming can also be used by doctor to open patient's data that has been sent to the server. Doctor can also perform video streaming in mobile environment by using mobile devices such as mobile phone, PDA or notebook. Fig. 3 shows the proposed streaming framework for mobile telemedicine system.

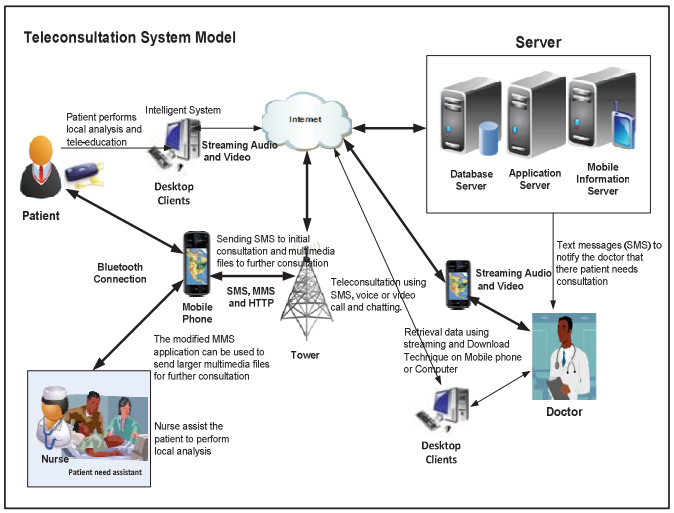


Fig. 3. The proposed streaming framework for mobile telemedicine system

We assume that on mobile phone and computer an expert or intelligent system is installed to detect the abnormality conditions from bio signals such as ECG (*local analysis*). If any abnormality is detected, the result is sent to the server using SMS and then the server sends it to the doctor. The doctor makes the diagnosis and the patient can have consultation using chatting, SMS or phone calling. If the doctor needs more data for further consultation, the patient can send the data (multimedia file) using MMS (modified MMS technology) or HTTP protocol to server. The patient's data can be retrieved by the doctor by using downloading and streaming.

IV. Simulation of Multimedia Streaming

On the server side, the first step is installing software such as Xampp software that can change a computer to act like a server (called localhost), then run it and activate the 'Apache service' in the Xampp software. Video file that will be streamed is stored in the 'htdocs' sub-folder residing on Xampp folder (for example, the video file 'phantom.mpg' is used for simulation). Whether the file can be accessed or not can be tested by typing the URL address http://localhost/phantom.mpg on the server. If the connection is successful, then the file will be played.

Mobile video streaming application is developed on the (receiver) client using J2ME and NetBeans emulator. Simulation of the mobile video streaming uses RTSP and HTTP protocols. For example, video file of phantom.mpg is streamed using HTTP protocol. The URL address http://localhost/phantom.mpg is typed on the client phone. But, if the server is on a different computer, IP address of the server is used, (e.g. 192.168.0.1) and the URL address will be http://192.168.0.1/phantom.mpg. Similar procedure is followed when using a separate mobile phone device. If the video file on the server is identified, then this application will play the video as shown in Fig. 4.

Fig. 4. Simulation mobile video streaming on the mobile phone

V. PERFORMANCE EVALUATION

In this research, we perform video streaming on peer-to-peer network by using streaming protocols such as RTP, RTSP, HTTP and MMS. We observe the streaming process, quality of video and buffering time for each protocol. The streaming process will be smooth after the connection with server is established. To simulate video streaming on mobile phone using mobile communications network, we must connect with the existing streaming server and then open using media player on mobile phone. The experimental results show that the duration of video files does not have any effects on the packet loss. The second experiment is conducted by streaming video from the DS to the mobile receiver using HTTP and RTSP protocols. Fig. 5 Comparison of buffering times for playing the video for RTSP and HTTP.

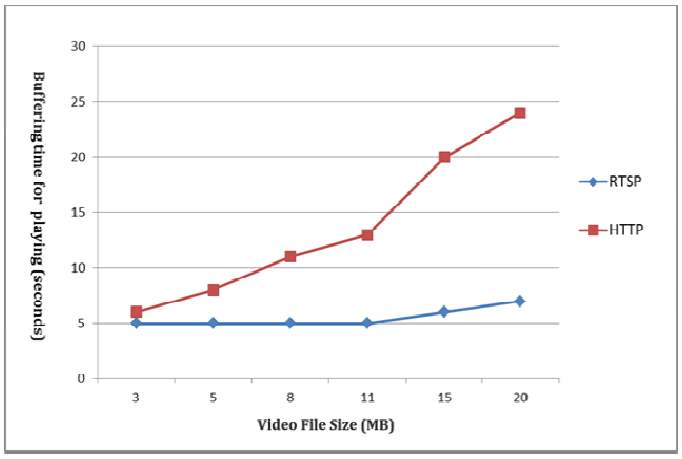


Fig. 5. Comparison of buffering times for playing the video

Based on the experiments conducted, it can be concluded that for the process of video streaming, RTSP is the best if the streamed video files are very large, while HTTP is the best for small sized files. This is because RTSP works by running piece by piece of large file, therefore it does not take longer time to play streaming video files. In addition, the use of RTSP protocol gives added benefit for mobile phones, because it does not need large storage media to play large video files. We note that the performance of RTSP is better than HTTP protocol. This is because that HTTP is not originally developed for streaming but only for web-based data transmissions. The RTSP is a protocol that is developed specifically for streaming.

A new framework for transmitting large multimedia data between two mobile clients using a developed server (DS) has been proposed. This framework has focused on multimedia streaming from the developed server to mobile client side as a client-server system. The multimedia streaming protocols including MMS protocol have been investigated to carry out the multimedia streaming. The proposed frameworks can be used to transmit large multimedia data files in limited bandwidth and mobile environments without any degradation in the data quality. The proposed framework can provide adaptive and cost-effective solution for mobile multimedia communications. The proposed framework is appropriate not only for MMS protocol but also for web-based service and streaming protocols. It provides several alternatives to transmit large multimedia data by adapting to different protocols, bandwidth available and media storage. The proposed framework can be used with resource- limited mobile phones. It is suitable for developing mobile telemedicine system that is focused for remote, rural, sea and island areas with limited bandwidth and mobile environments.

VI. CONCLUSION AND FUTURE WORK

Video coding technology continues to evolve with the aim of achieving improved results, especially for mobile communications. The need of information exchange in a mobile environment makes mobile technology develop at rapid rate providing a variety of increasingly sophisticated features. Currently, we can access the information available on the internet in the form of multimedia using mobile devices. The limitation of memory and bandwidth makes it difficult to retrieve data using mobile devices. Currently H.264/MPEG4-video coding is found to be most suitable for mobile communications.

In this paper, we have proposed a design method for mobile client-server system that can be used for developing a mobile telemedicine system, which includes video streaming. Peer-to-peer video streaming is used to simulate streaming process using some popular protocols. The experimental results show that the proposed streaming method is suitable for transmission of large size video files. The streaming method is more effective and efficient compared to downloading method especially for mobile communications. For future work, we will attempt to optimize H.264 video coding for mobile communications.

REFERENCES

[1] Nigel, C. and Jenny, C. (2009). Digital multimedia, John Wiley & Sons, Third Edition, England.

[2] Yun, Q. S. and Huifang, S. (2000). Image and Video Compression for Multimedia Engineering, Fundamentals, Algorithms, and Standards, CRC Press, USA.

[3] Vaughan T. (2007). Multimedia: Making It Work, McGraw-Hill Osborne, Seventh Edition.

[4] Hartsell, T. and Yuen, S. (2006) Video Streaming in Online Learning, AACE Journal, volume 14(1), pp. 31-43.

[5] Mabel, V. B., and Pierre, V. (2007)An Adaptable Architecture for Mobile Streaming Applications, IJCSNS International Journal of Computer Science and Network Security, VOL.7 No. 9.

[6] Iain, E. R. (2010). The H.264 Advanced Video Compression Standard, A John Wiley and Sons, Second Edition.

[7] Markus, R. (2009). Video and Multimedia Transmissions over Cellular Networks Analysis, Modeling and Optimization in Live 3G Mobile Communications, John Wiley & Sons Ltd, Chichester England.

[8] Thomas, S., Stockhammer, T., and Wiegand, T. (2007). Mobile Video Transmission Using Scalable Video Coding, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No. 9.

[9] Hendrawan. (2010), Dasar Pengkodean dan Kompresi Citra/Video [Basic of Image/Video Coding and Compression], Bandung Institute Technology, Indonesia.

[10] Rosenbaum, R. (2006). Mobile Image Communication using JPEG2000, Faculty of Computer Science and Electrical Engineering, The University of Rostock.

[11] Khalid, S. (2003). Lossless Compression Handbook, Academic Press, Elsevier Science USA.

[12] Iain, E. R. (2003). H.264 and MPEG-4 Video Compression Video Coding for Next-generation Multimedia, John Wiley & Sons Ltd, Chichester, England.

The Spectral and Temporal Description of Javanese Gong Kempul

Matias H.W. Budhiantho[1] and Gunawan Dewantoro[2]
Department of Electronic and Computer Engineering[1,2]
Satya Wacana Christian University
Salatiga, Indonesia
matias@staff.uksw.edu[1]
gunawan.dewantoro@staff.uksw.edu[2]

***Abstract*— In Central Java, the Gong Kempul is one of eminent gamelan instrument, an ensemble of predominantly struck instruments found across Central Java and has deep philosophical meaning for Javanese. However, there lack of studies concerning on this particular instrument as a bridging means between scientific description and human artistic perception. This study aims to investigate the spectral and temporal properties as well as particularly look into the typical wave-like sound of the Gong Kempul. Acoustic measurements were conducted and analyzed using ARTA. Both frequency and time domain analyses were explored to better understand the nature of the Gong Kempul. It was found that the partials at 93.8 Hz, 187.5 Hz, 281.3 Hz, and 375 Hz are the strict integer multiples of the fundamental frequency. It can also be inferred that the fundamental frequency at 93.8 Hz decays much more slowly than the other harmonic. In addition, the pitch of the Gong Kempul slide upward as the time goes on. The wave-like sound of the Gong Kempul is due to the beat phenomenon between two partials that occur closely.**

Keywords—gong kempul; frequency spectrum; beat phenomenon; impulse response; partials

I. INTRODUCTION

Gongs are instruments in Gamelan which perform what is often termed a colotomic function in the music to mark off established time intervals [1-2]. The smallest gongs which hang in a crossbar supported by two side posts are the Gong Kempul, as shown in Figure 1, with 56 cm in diameter. This instrument has a very deep and distinctly pitched roar mimicking the rolling waves of the sea. Sacrilegious aspect of Javanese Gamelan is very obvious, as in [3], and ones need a certain extent of the scientific counterpart. There have been studies in the past that attempt to describe the spectrum of gamelan instruments. Suprapto et al. [4] proposed to construct gamelan models. The main target of his research was creating Gamelan Frequency Modeling. He proposed two Frequency Balungan Models, the first model was using average value, and the other was using average value in the most dense area. Tjahyanto et al. [5] proposed a method for the features extraction, selection, and identification of gamelan note and the proper instrument. It was an approach based on Fast Fourier Transform (FFT), and support vector machines (SVMs) for note and instrument identification. Four spectral features were selected (spectral centroid, two spectral roll-off, and fundamental frequency) as inputs for SVM. Latief et al. [6] analyzed and identified Gamelan Bonang sound spectrum. He showed a method for sound-recognition with artificial neural network back propagation concept. The artificial neural network used sigmoid activation function to all layers. Extraction was performed, as well as Fourier transform, allowing the system to do the identification of voice data Gamelan Bonang using artificial neural network back propagation concept.

Wulandari et al. [7] investigated the implementation of machine learning approach to understand statistical variations contained in gamelan signals which are relevant to onsets. The method used Elman Network which consists of one hidden layer. However, there have been lacks of studies in the past that attempt to describe the spectrum of gongs both theoretically and experimentally to better comprehend such analyses. Maclachlan [8] employed Finite Element Analysis to predict the effect of a range of variations of gong geometries on modal frequencies. The predicted frequencies of the Finite Element Analysis experiment for gong models did not match the acoustic spectra for these gongs. However, the results indicate how the near harmonic overtone spectra recorded for these gongs have been produced by the right combination of physical properties. In his concluding work [9], a series of musical gongs were designed with up to the first five overtones tuned to within 5% of the harmonic series using purposefully added prestresses generated by cold-forging small dimples. The frequencies of overtones were most sensitive to the depth of the dimples located near the maxima of modal stresses.

Ayers and Horner [10-11] analyzed the Gong Ageng, the largest gong used in Gamelan, using a phase vocoder spectral analysis to estimate the amplitudes and frequencies of the partials. The Gong Ageng has about a dozen prominent exponentially decaying partials, with some component frequency ratios that closely correspond to harmonics and other that are inharmonic. Many of the partials have a slow amplitude and frequency modulation of a few Hertz, and a faster modulation around 20 Hz resulting from multiple components falling into the same harmonic bin. The studies in [12-13] indicated that gongs show a pronounced nonlinear behavior. The pitch of the larger gong glides downward as

Fig. 1. The Gong Kempul

Fig. 2. Condenser microphone (left) and phantom power (right)

much as three semitones after striking, whereas that of the smaller gong glides upward by about two semitones. Rossing et al. [14] investigated some vibrational modes of the larger gongs. Some of the modes were confined pretty much to the flat inner portion, some to sloping shoulders, and some involve considerable motion in both parts. When the gong was hit near the center, the central modes clearly dominate the sound. When the gong was hit lightly on the shoulder, the lowest mode is heard.

This paper considers the spectral and temporal properties of Javanese Gong Kempul to examine the spectral harmonics and partials of theses gongs, and how they are put together to produce the continuous beats. It was also found that the Gong Kempul possesses a degree of harmonicity, making it sounds with targeted pitch.

II. Spectral Analysis

A. Measurement Setup

We recorded 48kHz sample tones of the Gong Kempul using impulse response measurement in ARTA. First, a measurement condenser microphone acquired the acoustic signal of the Gong Kempul, externally powered by a phantom power as shown in Figure 2. A sound card then interfaced and digitized this signal in order that computers are able to recognize.

B. Normal Modes of Membrane [15]

For any circular membrane with a fixed boundary at radius a, one possible approach to describe the free oscillations is given by:

$$\omega_{mn} = ck_{mn} = (c/a)j_{mn} \quad (1)$$

where ω=angular frequency, c=speed of sound, k=wave number, j_{mn}=n^{th} root of Bessel function J_m. These subscripts have the physical interpretation that m tells the number of nodal line going diametrically across the membrane, and n is the number of concentric nodal circles, as shown in Figure 3.

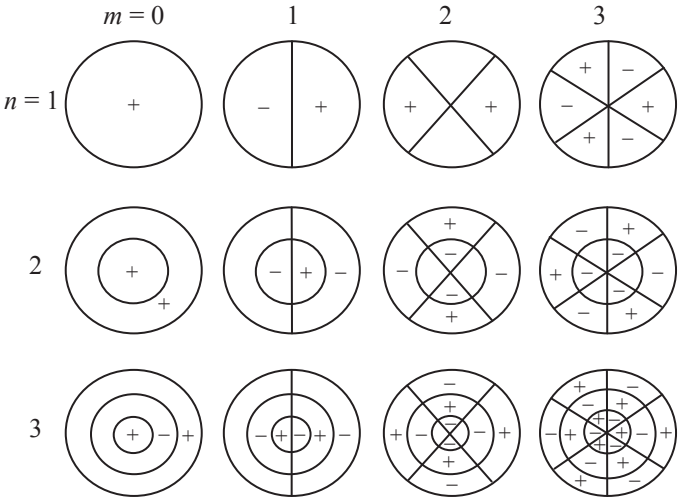


Fig. 3. Nodal lines and circles for some modes

The partials of the Gong Kempul were calculated using Eq. (1) with the total radius, including the gong shoulder, of 44 cm are shown in Table 1.

TABLE I. Calculated Partial Frequencies of the Gong Kempul

Partial at mode m,n	Frequency (Hz)
f_{01}	301.5 Hz
f_{11}	480.3 Hz
f_{21}	643.7 Hz
f_{02}	691.94 Hz
f_{12}	879.5 Hz
f_{22}	1055.25 Hz
f_{03}	1082.4 Hz
f_{13}	1275.3 Hz
f_{23}	1456.8 Hz

C. Spectral Properties

The plot of frequency spectrum of the Gong Kempul is shown in Figure 4 and 5. Hanning window was used to smooth out the spectrum profile. The spectral peaks indicate partials with significant strength. The log frequency scale was used in both smoothed and unsmoothed frequency spectrum analyses.

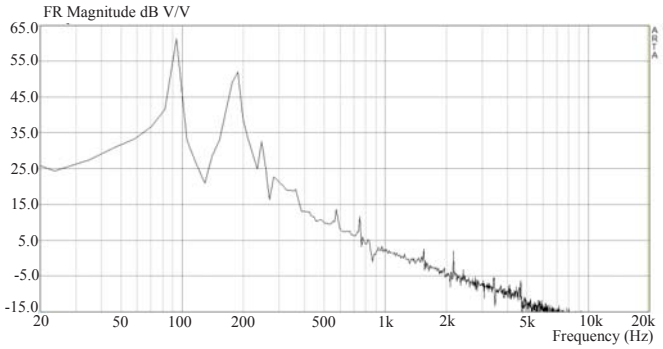


Fig. 4. Unsmoothed frequency spectrum of The Gong Kempul

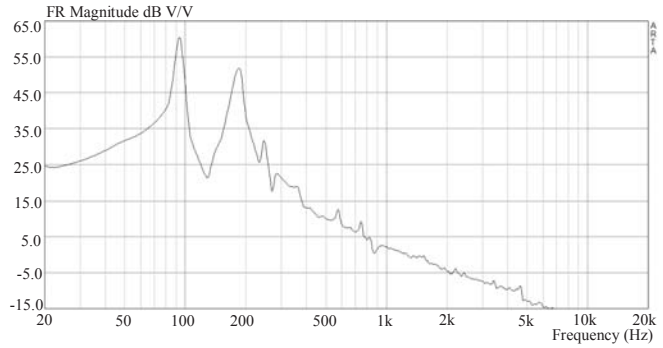


Fig. 5. Smoothed frequency spectrum of The Gong Kempul

TABLE II. MEASURED PARTIAL FREQUENCIES OF THE GONG KEMPUL

Partial #	Unsmoothed partial frequency (Hz)	Smoothed partial frequency (Hz)
1	93.8	93.8
2	187.5	187.5
3	246.1	246.2
4	281.3	281.3
5	339.8	339.8
6	375	375
7	457	457
8	574.2	574
9	750	750
10	832	831.6

Table 2 shows the ten first partial frequencies of the Gong Kempul, averaged from four trials. The partials in Table 1 do not match with that in Table 2 due to various restraining things. The measured partial frequencies in Table 2 are lower than the calculated partial frequencies in Table 1. Equation 1 perfectly holds for membranes whose tension is the only restoring force. This condition can only be satisfied in vacuum space. The decrease of frequencies might be caused by the mass of the gong and material stiffness, which in turn, give a low natural frequency as in mass-spring oscillation case. The irregularities of the Gong Kempul which stems from the manufacturing processes, including those caused by tapping the gong during tuning, also affect the partials. Furthermore, the massive and thick solid metal, in a sense, brings about inert properties of the Gong Kempul. However, the partials at 93.8 Hz, 187.5 Hz, 281.3, and 375 Hz are the strict integer multiples of the fundamental frequency. Hence, these partials are the first, second, third, and fourth harmonics, respectively. Such degree of harmonicity of the partials leads to the pitched sound of the Gong Kempul, differing it from that of Chinese gongs which sound like a crash. Therefore, the boss of the gong can bring the harmonic nature, as opposed to that calculated in Equation 1 which employs the non-strict integer multiples of roots of Bessel function. Nevertheless, some of the partials of Gong Kempul are not at strict multiples of the first harmonic. For instance, the inharmonic partial at 246.2 Hz falls between the second and third harmonics. The partials with the most significant magnitudes are the first and second harmonics, whereas the others do not seem to be significant. The harmonic and or inharmonic partials that occur in fairly close frequency beat together and form the roaring sound which is often associated with Bima's laughter (Bima is one of the Pandawa brothers in the story of Mahabarata).

Figure 6 show the burst decay graph to investigate the most prominent partials along a certain time intervals at the opening and tail of the sound, respectively. A complex Morlet wavelet analytic signal is used in convolution with system impulse response. The magnitude of that response, also known as *wavelet scalogram*, represents the envelope of the shaped burst response decay. A complex Morlet wavelet analytic signal is defined as:

$$w(t) = e^{-t^2/\tau^2}\left(\cos\omega_o t + j\sin\omega_o t\right) \quad (2)$$

The Fourier transform of the Morlet wavelet is equal to:

$$W(\omega) = e^{-(\omega-\omega_o)^2\tau^2/4} \quad (3)$$

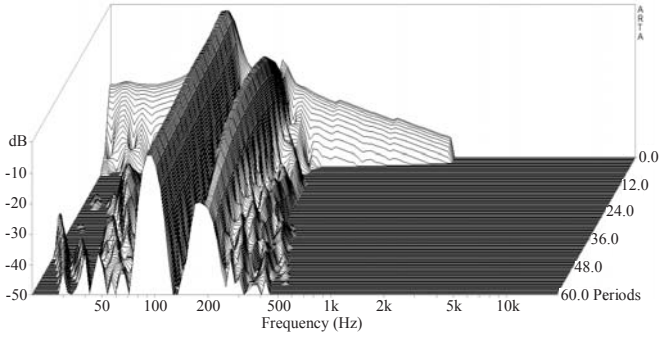


Fig. 6. The burst decay graph at the beginning of the sound

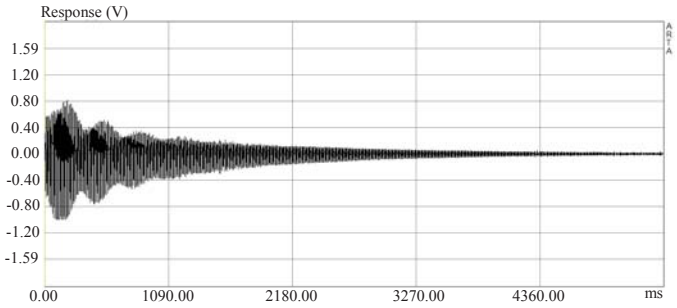


Fig. 7. Impulse Response

It can be seen that the fundamental frequency at 93.8 Hz decays much more slowly than the other harmonics and increases in dominance as the other partials weaken in decaying tone.

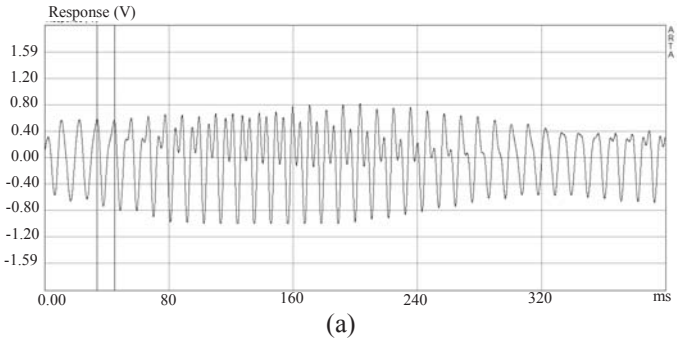
III. IMPULSE-GENERATED RESPONSE

A short impulse can be represented as the instantaneous and gentle strike on the Gong Kempul boss . Thus, an impulse-generated response was recorded to analyze in the time domain. Figure 7 shows the impulse response of the Gong Kempul within about 5400 ms in which the sound vanishes. The time plot specifies three-cycle beat frequency indicating three waves (Javanese: ombak telu).

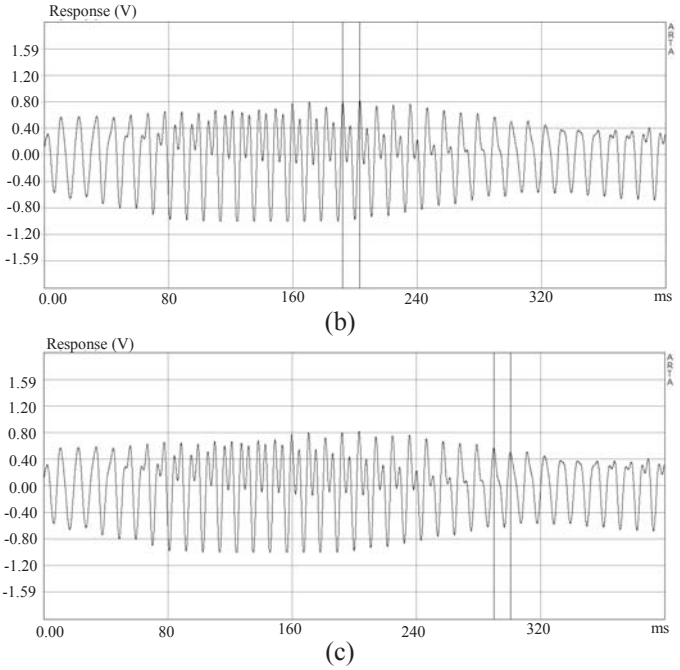
Upon zooming the time plot, as shown in Figure 8(a)-(c), we can observe that the Gong Kempul starts withs a repetitive wave of 11.33 ms period. After a while, the repetitive period decreases to 11.00 ms. Therefore, the gong starts with frequency of 88.26 Hz and subsequently the frequency shifts to 90.90 Hz. There is a difference of 2.64 Hz which is the origin of "ombak telu". At the tail of wave repetition, the period is 10.67 ms which coincides with the fundamental frequency, 93.8 Hz. In the other words, the sound fundamental wavelength can be derived by the following formula,

$$\lambda = \frac{343 + (T - 20) \times 0.6}{f} \tag{4}$$

where T is the ambient temperature during measurement (^{o}C), f=frequency(Hz), yielding the sound fundamental wavelength is 3.67 m with T=26 degree Celsius.



(a)



(b)

(c)

Fig. 8. Repetition waves with varying periods at the (a) beginning (b) middle and (c) end of the wave

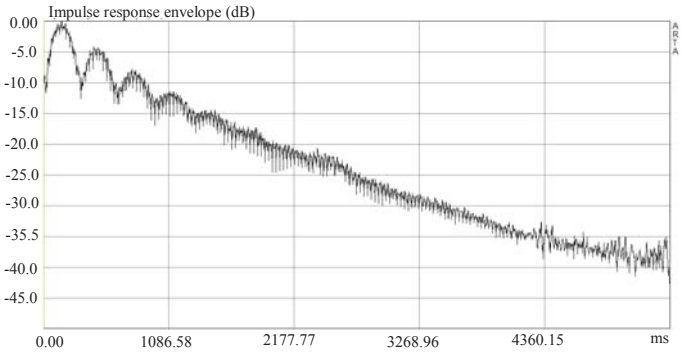


Fig. 9. Impulse response envelope

Figure 9 shows the impulse response envelope, also called Energy-Time Curve. The envelope of signal x(t) is an envelope of signal absolute values. Mathematically, it is defined by the following expression:

$$e(t) = \left[x^2(t) + \bar{x}^2(t)\right]^{1/2} \tag{5}$$

where is $\bar{x}(t)$ is the Hilbert transformation of $x(t)$, given by:

$$\bar{x}(t) = \frac{1}{\pi} x(t) \otimes \frac{1}{t} \tag{6}$$

The overall shape is a nearly-linear decay where the energy peak out at 151.938 ms. The beat phenomenon are apparent for about 1233 ms and such modulations are clearly found for the first five harmonics.

IV. CONCLUSION

The 56 cm Gong Kempul has a nature of harmonicity with fundamental frequency at 93.8 Hz, while the second, third, and fourth harmonics are, respectively, at 187.5 Hz, 281.3, and 375 Hz with several partials lie between each harmonics. It can also be inferred that the fundamental frequency at 93.8 Hz decays much more slowly than the other harmonics and increases in dominance as the other partials weaken. The wave-like sound of the Gong Kempul is due to the beat phenomenon between two partials that occur closely. The tuning of the Gong Kempul was carried by tapping the outer part of the gong which affects the vibration modes. In the future, gongs in different size and materials are about to be investigated.

ACKNOWLEDGMENT

This research was financially supported by Satya Wacana Christian University under grant no. 012/Penel./Rek/5/I/2013.

REFERENCES

[1] N. Sorrell, A Guide to The Gamelan, 1st ed. Portland: Amadeus Press, 1990.

[2] C. Capwell, "The Music of Indonesia", in Excursions In World Music, 4 ed., P.V. Bohlman, Upper Saddle River: Pearson Prentice Hall, 2004, pp. 141-170.

[3] H.S. Pranoto, "Sacrilegious aspect of Javanese gamelan: past and future," Harmonia. Semarang, vol. 13(1), pp. 55-68, 2013.

[4] Y.K. Suprapto, I.K.E. Purnama, M. Hariadi, M.H. Purnomo, T. Usagawa, "Sound Modeling of Javanese traditional Music Instrument," Proceedings of International Conference on Instrumentation, Communication, Information Technology, and Biomedical Engineering, Bandung, pp. 1-6, 2009.

[5] A. Tjahyanto, Y.K. Suprapto, M.H. Purnomo, D.P. Wulandari, "FFT-based features selection for Javanese music note and instrument identification using support vector machine," IEEE International Conference on Computer Science and Automation Engineering, Zhangjiajie, pp. 439-443, 2012.

[6] M.N. Latief, F.L. Gaol, B.H. Iswanto, "Analysis and Identification Gamelan Bonang Sound Spectrum," Proceeding of Second International Conference on Computational Intelligence, Modeling, and Simulation, Bali, pp. 335-338, 2010.

[7] D.P. Wulandari, Y.K. Suprapto, M.H. Purnomo, "Gamelan music onset detection using Elman Network", IEEE International Conference in Computational Intelligence for Measurement Systems and Applications, Tianjin, pp. 91-96, 2012.

[8] N. Maclachlan, "Finite element analysis and gong acoustics," Acoust. Aust. Sydney, vol. 25, pp. 103-107, 1997.

[9] N. Maclachlan, R. Adams, and C. Burvill, "Tuning natural modes of vibration by prestress in the design of harmonic gong," J. Acoust. Soc. Am. New York, vol 131(1), pp. 926-934, 2012.

[10] L. Ayers and A. Horner, "Synthesising Music in Csound with a Javanese Gong Ageng," Proceedings of Generate and Test, Australasian Computer Music Conference, Brisbane, Australia, pp. 11-14, 2005.

[11] L. Ayers and A. Horner, "Synthesizing a Javanese Gong Ageng," Proceedings of the 2005 International Computer Music Conference. Barcelona, Spain, pp. 644-647, 2005.

[12] T.D. Rossing and N.H. Fletcher, "Nonlinear vibrations in plates and gongs," J. Acoust. Soc. Am. New York, vol 73, pp. 345-351, 1983

[13] A. Chaigne, C. Touze, and O. Thomas, "Nonlinear axisymetric vibrations of gongs," Proceedings of ISMA. Perugia, Italy, pp. 147-152, 2001.

[14] T.D. Rossing, J. Yoo, and A. Morrison, "Acoustics of percussion instruments: An update," Acoust. Sci. & Tech. Japan, vol. 25(6), pp. 406-412, 2004.

[15] D.E. Hall, Basic Acoustics, 1st ed. New York: Harper & Row Publishers, 1987

Session 3

Power Systems

1 MWp Grid Connected PV Systems in the Village of Kayubihi Bali;

Review on Location's Characteristics and Its Technical Specifications

I N. Satya Kumara, W.G. Ariastina, I W. Sukerayasa, I. A. D. Giriantari
Department of Electrical Engineering, Faculty of Engineering Udayana University
Kampus Bukit Jimbaran, Bali
Telephone +62 361 703315
Email: satya.kumara@ee.unud.ac.id

***Abstract*—Bali is one of the small islands in Indonesia with total land area of 6,800 square kilometers and with population of around four million people. The island's economy is mainly driven by its tourism industry that leads to annual growth of 6.8%. The life and economy is supported by electric power system with 696 MW generating capacity but at peak time already operating at 679 MW which forces industries to operate their own captive powers. This condition requires swift and effective response from the authority to improve the generating capacity of the local power system so it can continue to support the regional development. Bali has no fossil based resources hence fuel for the power plants need to be externally sourced but with the dwindling capacity of the nation's conventional energy resources together with the prospect of global warming therefore solution has been directed on renewable generation. As tropical island which is in close vicinity of the equator line, Bali has abundant quantity of sun power. The sun insolation index is estimated between 5 to 6 kWh per square meter everyday. And also, photovoltaic system is suited for urgent power demand as it can be built in relatively short time and with advantage of being environmentally friendly. These general facts have led to to the installation of 1 MWp grid connected photovoltaic systems in the village of Kayubihi by the Ministry of Energy and Mineral Sources. The system is currently the largest PV system in Indonesia and connected to the 20 kV distribution network. The report presented on this paper is the review on natural and social characteristics of the location and technical specification of the hardware system. This information are useful in analyzing performance of the photovoltaic system particularly on energy production, reliability of system and components, tariff calculation and investment, as well as developing model for the running of remote PV system based on partnership between local government and university.**

Keywords— ***Bali renewable electricity, PV location characteristics, photovoltaic system, grid connected PV system***

I. INTRODUCTION

The utilization of photovoltaic system in Indonesia was started some time ago in the early 80-is. It was introduced to villagers in remote areas which could not be connected to PLN; a state owned utility company, grid due to geographical constraints. The photovoltaic system was built as a small PV system known as solar home system (SHS) to replace kerosene lamps for lightings. Until today, this system continue to be installed by the government as alternative for providing electricity access for people who live in remote islands of Indonesia. In fact, the SHS is also part of program to achieve the 2025 renewable target by distributing one million units of 50 Wp SHS to people living outside of PLN's grid. In addition, the government has also set to build 346 MW hybrid PV systems as energy source for remote areas, [1]. However, there has been little information about the progression of the two programs as to whether it will be able to meet the national target.

The Indonesian National Energy Policy states that by 2025, the contribution of new and renewable resources should be around 5% of the national energy mix for electricity generation. The renewable resources include hydropower, sun power, wind power, bio fuels, and nuclear. Total electric power demand by 2020 is estimated at around 50 GW [2] and by 2025 the figure will rise to between 60 to 70 GW. Currently, the total national power capacity is 30 GW. If the contribution of photovoltaic system is assumed 1% then it requires the development of 600 MW PV systems by 2025 within 12 years from today, which means that every year for the next 12 years, a total of 50 MW PV systems should be built. This figure is relatively large if it is compared with the current progress, particularly as the government had set a much higher target of 870 MW PV system by 2025 [3].

On February 27th, 2013 in the Village of Kayubihi, District of Bangli, Province of Bali, Minister of Energy and Mineral Resources officially inaugurated the operation of 1 MW grid connected photovoltaic system injecting power into PLN's 20 kV distribution networks. On his speech, the minister Mr. Jro Wacik clearly stated that the development of this photovoltaic system which currently is the largest grid connected PV system in the country should be disseminated to public, particularly the stakeholders of renewable energy in Indonesia. This pilot project should be considered as milestone on strengthening and increasing the national capacity of renewable electricity generation toward achieving 5% electricity from renewable resources by 2025 as stated on the National Energy Policy [4].

The nomination of Bali as location for this relatively large size PV system as national pilot project supports the regional government of Bali in the implementation of Bali Green and Clean Programs and also the fact that Bali is currently experiencing power deficits. Bali's electrical power system is supported by three power plants with total capacity of 696

MW, transmission lines of 662 kilometers long, 14 units of substations to serve over 800,000 consumers. The three power plants are fueled by high speed diesel oil (HSD). Bali is also interconnected with Java's grid using submarine cable and the power is supplied from Paiton coal power plants on East Java. It is clear that Bali is solely dependent on fossil based fuel for its power generation. On the other side, it is now better understood that the national reserved on oil, gas, and coal are estimated to only be available for the next 75 years. On the demand side, Bali's peak load has reached 679 MW with growth is estimated at 6.8% which largely due to extensive development of tourism industry. While villages' electrification has reached 100% which means all of the villages in Bali has had access to electricity but the more remote sub villages are still without it. Kumara [5] reported that there were more than 45 sub villages in Bali that had no access to electricity due geographical and topological characteristics of the areas which hinder expansion of PLN's distribution networks. Further, it was identified that local resources such as hydropower or sun power are alternatives for electricity generation for these sub villages.

The report presented here is an initial part of an on going research on the monitoring and performance analysis of the 1 MW Kayubihi PV System that covers energy production analysis, system and components reliability, energy tariff and investments, as well as developing model for operating a remote or rural medium size PV system by incorporating local government agencies and university. This paper will review the area or regional characteristics where the system is installed and the technical specifications of the PV hardware and system.

II. PROFILE OF PV LOCATION

A. Village of Kayubihi in District of Bangli

In order for PV system development project to generate maximum benefits for people and the stakeholders of renewable energy sectors, then various factors are considered in nominating the location of the PV system. This part will discuss natural and social characteristics of the region where the 1 MW PV systems is installed. The factors include the existing power grid, topography and demographics of the region, weather and climate, as well as the potential of sun power itself.

Bangli is one of the district governments of Bali with total area of 52,000 hectares and divided into four sub districts consisting 72 villages [6]. The region is spanned from 115^{o} 13' 48" to 115^{o} 27' 24" longitude and from latitude of 8^{o} 8' 30" to 8^{o} 31' 87". The location is at elevations from 100 to 2,152 meter above sea level. The population is around 216,000 with population density around 415 people per square kilometer [7].

Village of Kayubihi is within the administrative area of Sub district Bangli. It consists of four sub villages, namely Bangklet, Kayang, Kayubihi, and Undisan. The population of the village is around tens of thousands of people and the main livelihoods are dry crops farming, cattle breeding, bamboo and wood handicrafts production, as well as other general trades.

B. Electricity Grid in District of Bangli

From electrical power point of view, the regional area of Bangli is part of PLN's Eastern Bali Area Networks. The area is served by 20 kilometer 150 kV transmission lines, 20 kV distribution networks, 220 V low voltage networks. The customers are mainly residential type. All of the 72 villages in the district have been connected to the grid except some sub villagers that are located beyond the reach of PLN's distribution networks. Kumara [5] reported that in 2009 there were tens of sub villages that are not yet connected to the grid. Topographical characteristic of the areas such mountains and valleys are the constraints for PLN to reach for these people.

The 1 MWp Kayubihi PV system is a demonstration of grid connected distributed generation. The generated power is injected into the 20 kV distribution networks owned by PLN. The unit is located very close or it is inside the Kayubihi 20 kV feeder hence connection is straightforward. Observation on many existing renewable generation system in Bali shows that often the power plants are located far from the distribution lines hence extending the link is necessary and this will add into the initial investment and sometime could hinder the utilization of the natural resources.

C. Environmental Profile of Bangli

Output of PV system is affected by various natural factors such as the availability of sun, temperature, rainfall, shading, condition of atmosphere in the surrounding location, cleanliness of PV modules, site condition, wildlife and domestic animals, community villagers, etc. On this section, condition of those factors are presented and data are of secondary type are taken from local authorities.

Generally, the potential converting sun power into electricity in Bali is quite high, as it is shown by high sun's insolation index which are between four to six kilowatt hours per square meter every day. However, there has been little information on the actual energy production per square area of previously installed PV systems and perhaps comparing it to these figures to get more realistic data with respect to performance or energy production. It is clear that PV's output will be greatly affected by the actual availability of sun throughout the year. One factor that affects the availability of the sun is the number of rainfall and rain days. Based on Kopen weather classification, the area of Bangli belongs to wet weather condition [8]. At one time, the rainfall was recorded at 797 mm per year. Table I shows monthly rainfall and rain days on Bangli in 2010 [9]. The table shows that the number of rain days in the region is quite large where more than half of the month had rains which would directly affect the PV system performance with respect to energy production.

The district of Bangli has relatively cool daily temperature that varies from 15^{o}C to 30^{o}C with daily average of around 24.5^{o}C. Generally, the low temperature occurs from late afternoon to evening and night time until morning, while temperature is generally toward the high end during day time when the sun in shining. The average humidity is around 68% and wind speed at around 7 knot [7].

TABLE I. MONTHLY RAINFALL AND RAIN DAYS IN BANGLI IN 2010

Month	Rainfall (mm)	Rain days (day)
January	456.0	16.0
February	198.5	13.0
March	252.0	15.0
April	267.0	18.0
May	334.0	15.0
June	219.0	16.0
July	300.0	18.0
August	255.0	18.0
September	595.0	18.0
October	475.0	21.0
November	200.0	17.0
December	-	-

such as temperature, both ambient and module, wind speed, and solar irradiation level. This measurement will be useful in analyzing the performance of the PV systems in term of energy production and factors that affecting it.

A. *Photovoltaic array*

Photovoltaic panel is component that converts the energy of sunlight into electrical currents. The advance of PV system is greatly affected by the development of PV panel technology. The current commercially available panel has efficiency in the range of 14 to 16 percent. The PV panel used on this application is of monocrystaline type which manufactured by PT LEN Industry which capable of producing maximum power of 200 watt at standard condition test.

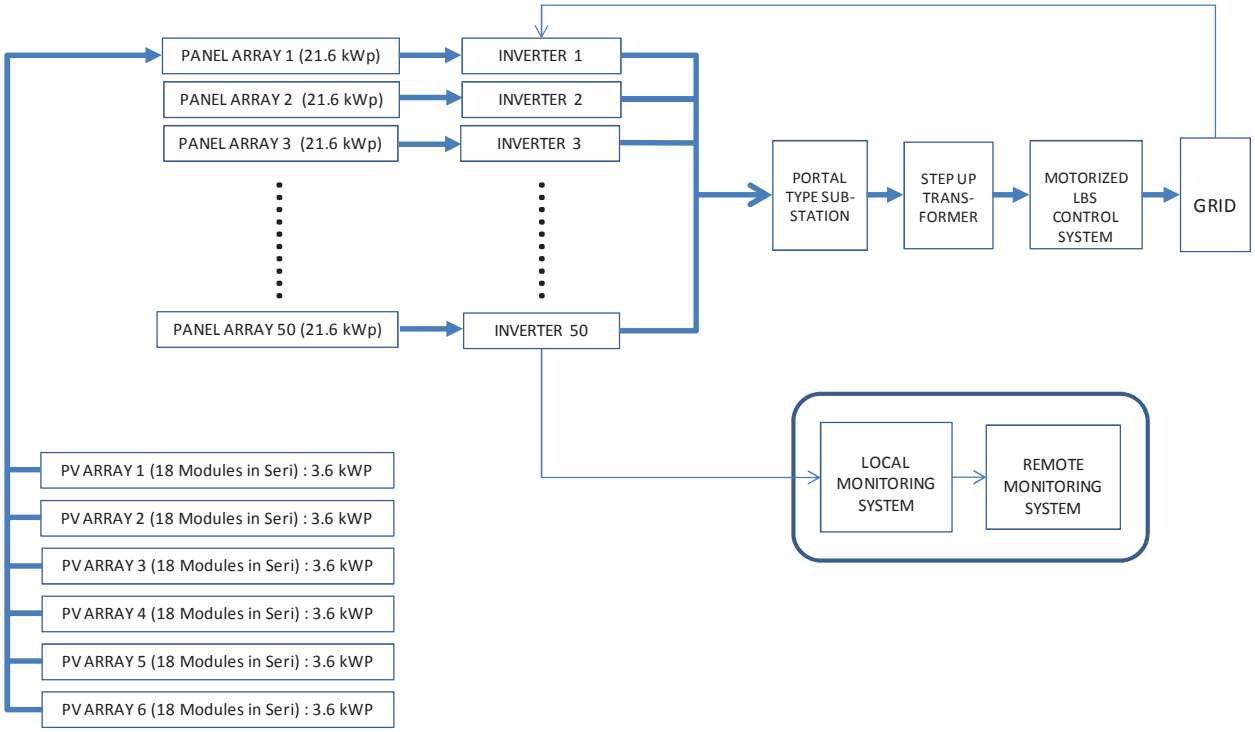


Fig. 1. Configuration of 1 MW Kayubihi PV system

III. TECHNICAL SPECIFICATIONS OF 1 MW PV SYSTEM

PV system converts sunlight into electricity and its output is affected both by the sun power and other environmental factors and the PV hardware. Typical components of PV systems are PV panel, battery, battery charger controller, inverter, control system, and other supporting components. Battery is used to store the energy during daytime and consume at night time, therefore for system that directly consume its output during daytime will not require energy storage. The Kayubihi system does not use battery as its output is directly connected to the medium voltage utility's grid. The schematic of the 1 MW Kayubihi PV system is shown in Fig. 1. [11]. The environmental monitoring systems monitor various variables

The complete technical specification of the LEN 200W-24V panel is shown on Table II [12]. To meet the requirement of the input voltage range of the inverter, 18 units of PV modules are connected in series to form an array with total nominal voltage of 18 x 24 volts or approximately 432 V dc voltage but maximum voltage could go up to 670 volts depending on the level of solar irradiation. Between five and six PV arrays is then paralleled to form a group and this group is served by one unit inverter of 20 kW. To meet the power output, 50 groups are built to make up the one megawatt output. The configuration of panels used on the 1 MW Kayubihi PV system is shown in Fig. 1. The total number of panels required is 5004 panels.

TABLE II. TECHNICAL SPECIFICATION OF LEN 200W–24V

Parameters	Symbol	Unit	Value
Panel efficiency	η	%	15
Nominal voltage	V	V	24
Maximum power (STC)	W_P	W	200
Voltage at maximum power	V_{max}	V	37.4
Current at maximum power	I_{max}	A	5.3
Short circuit currents	I_{SC}	A	5.5
Open circuit voltage	V_{OC}	V	44.2
Operating temperature	t	^oC	-40-+85
Temperature coefficient		%/^oC	0.44
Power tolerance		%	0.3
Weight		kg	16.5
Dimension	L x W x H	mm	806x1576x50

TABLE III. TECHNICAL SPECIFICATION OF SUNGROW INVERTER

Parameters	Symbol	Unit	Value
Output power	P_{OUT}	kW	20
Input voltage	V_{DC}	V	250
Input currents	I_{DC}	A	40
Output voltage	V_{OUT}	V	400
Output current	I_{OUT}	A	31
Output frequency	f	Hz	50
Waveform	Pure sinusoid		
Total harmonic distortion	THD	%	3
Number of phase			3
Operating temperature	T	^oC	25-60
Weight		kg	50
Dimension	L x W x H	mm	648x695x237

B. Inverter

An inverter converts the DC voltage generated by PV arrays into AC voltage. The inverter used for this application is built using IGBT three levels inverter made by SUN GROW China of the SG 20 KTI product series. The inverter has maximum DC power input of 21 kW and maximum AC output of 20 kW. The number of inverter units installed of this system is 50 units giving maximum AC power of 1 MW. The major subsystems of the inverter circuitry are shown in Fig. 2. and its technical specification is shown on Table III [13].

C. Connection to Grid

The 1 MW Kayubihi PV systems is connected to utility grid on the 20 kV networks. The schematic configuration of the connection is shown in Fig. 3. [11]. Output of SG 20 KTI inverter is connected to five units three phase 400/20KV, 50 Hz, 250 KVA step up transformers. The transformers installed are of oil immersed outdoor type transformer manufactured by Schneider. The connection or removal of the system to and from the grid is established via motorized load break switch (LBS).

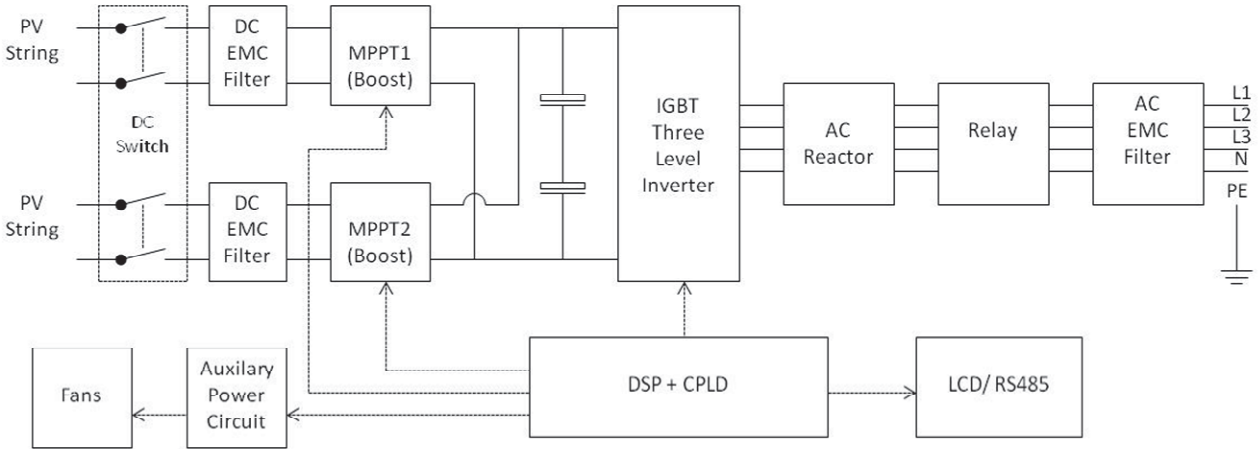


Fig. 2. Electronics systems of 1 MW Kayubihi PV system

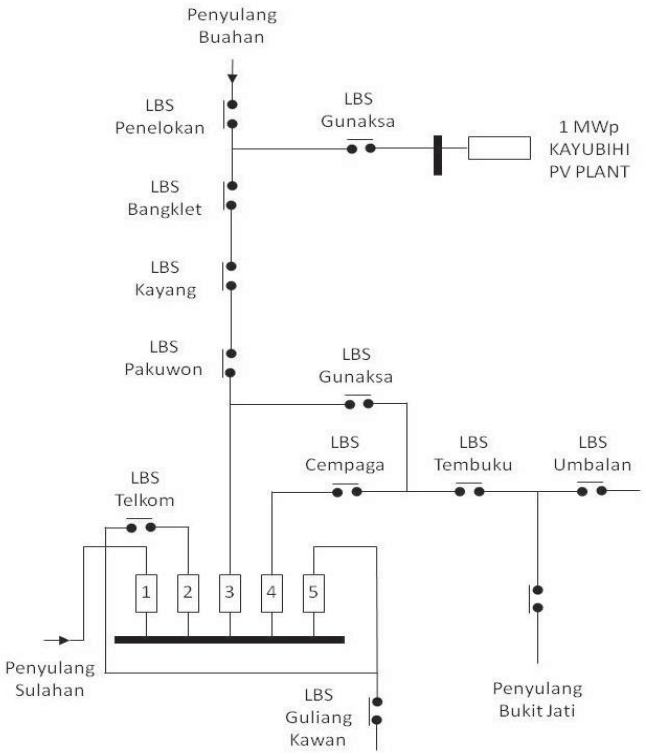


Fig. 3. PV system connection to 20 kV grids

D. *Grid and PV Synchronization*

The 1 MW Kayubihi PV systems is connected to utility grid on the 20 kV networks. Successful injection of power by the PV system depends on the synchronization of its output and grid's parameters such voltage, frequency, and phase sequence. The PV system or indeed any generating plant can only be connected to the grid if its output voltage, frequency and phase sequence of the plant match those of the grid. The synchronization is implemented using phase lock loop (PLL) in which grid parameters are used as reference to generate output of the inverters [11]. This also implies that the PV systems will only inject power when the grid is in operation, hence the reference is available. If the grid is off for any reasons then the inverter does not have any input references and consequently will not producing output. At this condition the control system will deactivate the power plant. This will ensure that during maintenance of the distribution networks, the PV system will not injecting power to the grid which may endanger the maintenance team.

IV. DISCUSSION

The location of the PV system is within a waste disposal site own by District Government of Bangli. The total land area required for the power plant is around 1.2 hectare. The area is used for PV arrays installation, grid connection through five units of step up transformers, control panel and buildings for housing control panels and computer based monitoring systems. Fig. 4. Shows photos of the PV power plant.

The power plant was developed by the Ministry of Energy and Mineral Resource and will be handed over to the local government after passing the commissioning period by the contractor. After the hand over, the operation of this PV system including the necessary maintenance will be carried out by the local government or agency set up for that purpose. Historically, many of the renewable power plants developed by the central government which then handed to the local government had experienced various problems due to lack of capacity of the operator's side and other factors. To avoid or minimize similar problems, the Ministry has initiated collaboration with local university to assist the local government in the operation and maintenance of the power plants.

Theoretically, the 1 MW Kayubihi PV system could generate maximum power of one million watt during day time. However, environmental factors such as temperature, rainfalls, rain days, wind condition, etc, will affect output of any PV systems. In addition, site condition such as shading, modules cleanliness, wildlife, and non technical aspects such as site security or access will affect the operation of the plant hence energy production of the Kayubihi PV plant.

Generally, the climate on the location of PV is a transition between wet climate of Kintamani and mild wet climate of Bangli which means relatively large rainfalls and many rain days. The relatively long duration of rains will directly affect the energy production of PV system. The design has adopted a daily effective sun hour (ESH) of 6.17 to estimate the system's energy production [11]. This number was obtained from NASA's publication and applies for Bali in general so it is not specifically for the location of the Kayubihi PV system. The value is relatively large considering the history of rain days and the amount of rainfall which is largely different from one area to the other in Bali. However, it remains to be confirmed by the actual energy production.

With respect to temperature, although at first it seems that data of local temperature is very close to STC value, with daily average of 24.5^{o}C, however, since there is little information on the actual daily temperature variation therefore this would need to be confirmed by actual measurement. Temperature needs to be monitored in order to better relate the effect of surrounding temperature on the energy production.

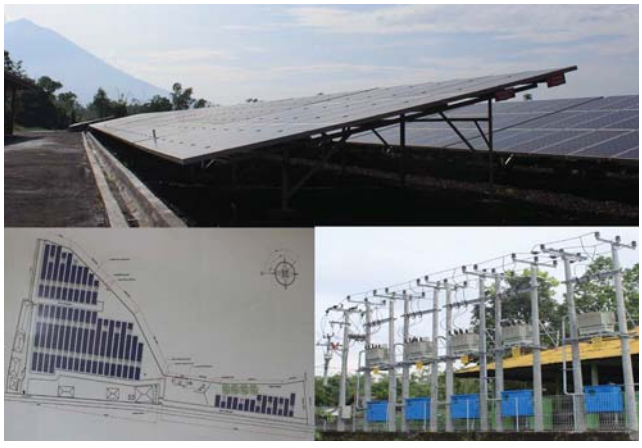


Fig. 4. Photo of the 1 MW Kayubihi PV Systems

Based on initial site observation, other factors that potentially affect the performance of the PV system is shading of the arrays as well as debris deposit on the modules' surface. Surrounding the PV installation, small vegetation such as weeds and tall grasses already growing which can affect the reception of sunlight. Big trees in the vicinity of the PV panels may also block sunlight. The positioning of the modules among themselves will likely to cause shading to other parts of the arrays and this need to be observed.

Wildlife in this area is mainly birds and also loose domestic dogs. Initial observation on site also found that many of the PV panels are already covered by the birds' manure and dry leaves or twigs. While this is minor but over time the accumulation of this debris could cause blockage of sunlight to the photovoltaic cells. Loose domestic dogs have been seen roaming the site and also tempering the installation and even climbing on the arrays.

The location of the PV power plant is part of an area designated for regional waste disposal site. Domestic wastes are collected by the local government unit and then dumped on the other part of the site. The surrounding atmosphere is likely to contain higher degree of debris or dusts which could drop and deposit on the surface of the PV module which occupies relatively large area of more than one hectare.

The site are visited by many people including those who work on the disposal site and people who scavenging used and scrap materials, therefore the power plant need to be secured from any tempering both for safe operation of the systems and equipments as well as safety for people.

V. CONCLUSION

This paper has presented information on the grid-connected 1 MW Kayubihi PV system with respect to environmental profile of the location and technical specifications of the installed PV system.

The environmental factors such as temperature, solar irradiation, weather conditions and their effects on the energy production will be analyzed using information recorded by the environmental monitoring systems.

To ensure maximum sunlight conversion, the cleanliness of the PV arrays should be given focus of attention as there are many factors on the surrounding site that could cause blockage of sunlight to the PV cells. Therefore, careful cleaning procedure and technique should be set up taking into account the potential amount of debris deposits, the delicate nature of the module surface and the way the modules are configured or positioned.

Disturbance on the site that potentially affect PV performance also come from animals that can get into the plant area relatively easy. Loose dogs have been seen tempering the wiring of the system and even climbing on the PV arrays. Measures should be taken to minimize unauthorized access and tempering the installation of the PV system.

Monitoring the operation of the PV systems and its performance are currently underway. The measurements will be used to discuss the energy production, reliability of PV hardware and systems, energy tariff and investments analysis, and development of model to operate this PV power plant, based on partnership between District Government of Bangli and Udayana University. The plan is aimed to enable sustainable operation of the Kayubihi PV systems and to provide better understanding on the actual electrical performance of large size grid connected PV system and other technical or operational aspects of the systems, and to provide benefits to wider community.

ACKNOWLEDGMENT

The authors would like to acknowledge the support given by LPPM Udayana University that provides the research grant under PUPT 2013 scheme. We also acknowledged the access to the 1 MW Kayubihi PV Power Plant provided by the Ministry of Energy and Mineral Resources and District Government of Bangli.

REFERENCES

[1] ESDM, Matahari Untuk Pembangkita Listrik Tenaga Surya di Indonesia, available on line at www.esdm.go.id

[2] Muchlis, M., Permana, A.D., Proyeksi Kebutuhan Listrik PLN Tahun 2010 Sampai Dengan 2020, BPPT

[3] LITBANG ESDM, Dukung Kebijakan Energi Nasional dengan Pemetaan Potensi Energi Baru Terbarukan, available online at www.esdm.go.id, LITBANG ESDM

[4] DEN, Kebijakan Energi Nasional 2010 2025, Dewan Energi Nasional, Jakarta, 2006

[5] Kumara, IN.S., Ariastina, W.G., Giriantari, IDA, Sukerayasa, I.W., Identifikasi Terhadap Wilayah Tanpa Listrik di Propinsi Bali dan Solusi Penyediaan Pasokan, Laporan Akhir Riset Strategis Nasional 2009, LPPM Unud, 2009

[6] BPS Bali, Data Curah Hujan dan Hari Hujan Kabupaten Bangli Tahun 2010, BPS Bali, 2011

[7] Pemerintah Kabupaten Bangli, Bangli Dalam Angka 2010, Pemkab Bangli, 2011

[8] Daryanto, Klasifikasi Koppen Cuaca untuk Bali, 2006

[9] Dirjen EBTKE, Petunjuk Operasi Dan Pemeliharaan Pembangkit Listrik Tenaga Surya (PLTS 1 MWp Terinterkoneksi Jaringan), Direktorat Jenderal Energi Baru Terbarukan Dan Konversi Energi Republik Indonesia, 2013

[10] PT LEN Industry, Len 200 Wp – 24V Monocrystalline Photovoltaice Moduls, PT LEN Industry Indonesia, 2013

[11] SUNGROW, PV Grid Connected Inverter User Manual: SG10KTL/ SG12KTL/ SG15KTL/ SG20KTL, SUNGROW China, 2012

A Novel Design of WACS Based Multi-Output Support Vector Machine (M-SVM) for Oscillation Damping on Power System

Muhammad Abdillah
Department of Electrical Engineering,
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
abdee.muhammad83@gmail.com

Mauridhi Hery Purnomo
Department of Electrical Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
hery@elect-eng.its.ac.id

Adi Soeprijanto
Department of Electrical Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
adisup@elect-eng.its.ac.id

Imam Wahyudi Farid
Department of Electrical Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
wahyu073@gmail.com

***Abstract*—This paper proposes a novel design of wide area control system (WACS). WACS is utilized to damp the oscillation on power system. WACS consists of wide area monitor (WAM) and wide area control (WAC). WAM is used to monitor the dynamic behavior of power system, while WAC is used as the additional controller on power system. The proposed method is called WACS based multi-output support vector machine (M-SVM). M-SVM used in this paper is M-SVM for regression. The input signal which used by WAM is the input signal to the AVR ΔV_{wi}, the mechanical power ΔP_{mi} and the electrical power ΔP_{ei} of each generator. The output of WAM is utilized to predict the speed deviation of the generator. WAC is using the input signal of mechanical power ΔP_{mi} and electrical power ΔP_{ei} from generator, while the output of WAC is the signal control which injected to the AVR ΔV_{wi}. A two-area-four-generator is utilized as a tested system to evaluate the performance of the proposed method. From the simulation results that has been conducted, the proposed method can reduce the overshoot and compress the settling time better than other methods that presented in this paper.**

Keywords—wide area control system; wide area monitor; wide area control; multi-output support vector machine

I. INTRODUCTION

Modern power systems are complex systems, multivariable and distributed in a wide geographical area. This led to the modern power system is difficult to be analyzed because it consists of a variety of load and power plants that work together to supply the load continuously and vulnerable from internal and external disturbances.

Wide area control system (WACS) is the alternative control method to solve the problem due to the WACS is using the input signal from each region of the distributed power systems and the output signal of WACS is injected to the AVR of generator in each region as additional controller on power system.

In the last decade, there are several methods that have been proposed in the area of wide area monitoring and control. Neural network based ACD for optimal wide area control scheme [1] and wide-area signals-based intelligent control of flexible AC transmission systems devices [2, 3].

This paper presents a novel design of WACS based multi-output support vector machine (M-SVM) for a two-area four-generator power system model. WACS is used to damp the oscillation on power system.

This paper are organized as follows, In section 2 a review of power systems modeling and wide area control system are discussed. The proposed method is mentioned in section 3. The performance of proposed method is presented in section 4. The last section is the conclusion.

II. POWER SYSTEM MODEL AND WIDE AREA CONTROL SYSTEM

A. Power System Model

Power system model used in this paper is a dynamic linear model. The model is used to analyze the behavior of system due to small perturbations, such as small load changes. When a small disturbance is occured, the system is towards the new operating point. The system of equation is linearized by fixing an initial condition for testing the behavior of the system during a small disturbance occured.

With a linearization method, a linear equation can be obtained from system of equation. When a new linear equations is obtained, thereafter only valid within domain

close to a fixed initial conditions. Dynamic model of a power system used in this paper is Kundur Power System [4] as shown in Figure 1. The system are consist of two symmetrical areas where each area supplied by two generators that connected with double circuit line 230kV and have distance about 220 km. Generator type which used in this paper is thermal generator with rate values 20kV/900MVA. The generator is connected to electrical power system grid through a transformer. Generators that present in all areas have the same parameters, except for the inertia value H = 6.5s for all generators in Area 1 and H = 6.175s for all generators in Area 2. The load on the system is assumed as a constant impedance.

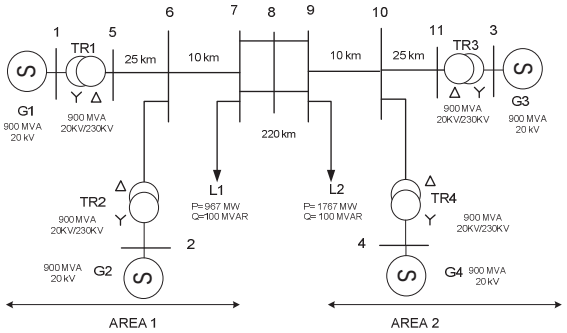


Fig.1. Two area power system model

B. Wide Area Control System (WACS)

Figure 2 is describe the wide area control system (WACS) that used as centralized control for a two-area four-generator power system. WACS in this paper consists of wide area monitor (WAM) and wide area control (WAC). WAM is used to monitor the dynamic behavior of power system while the WAC is used as the additional controller in the power system. WAM and WAC which are part of the WACS will describe on the next section.

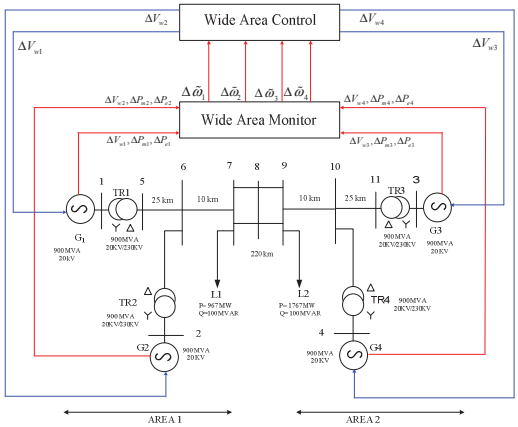


Fig.2. Two area power system model equipped with WAM and WAC

B.1. Wide Area Monitor (WAM)

Design of wide area monitor (WAM) on power system in this paper is using multi-output support vector machine (M-SVM) which consists of 36 inputs and 4 outputs. Kernel functions used in this design is radial basis function (RBF). The input signal used by the WAM are input signal to the excitation system [$\Delta V_{w1}(t\text{-}T)$, $\Delta V_{w1}(t\text{-}2T)$, $\Delta V_{w1}(t\text{-}3T)$, $\Delta V_{w2}(t\text{-}T)$, $\Delta V_{w2}(t\text{-}2T)$, $\Delta V_{w2}(t\text{-}3T)$, $\Delta V_{w3}(t\text{-}T)$, $\Delta V_{w3}(t\text{-}2T)$, $\Delta V_{w3}(t\text{-}3T)$, $\Delta V_{w4}(t\text{-}T)$, $\Delta V_{w4}(t\text{-}2T)$, $\Delta V_{w4}(t\text{-}3T)$], the mechanical power [$\Delta P_{m1}(t\text{-}T)$, $\Delta P_{m1}(t\text{-}2T)$, $\Delta P_{m1}(t\text{-}3T)$, $\Delta P_{m2}(t\text{-}T)$, $\Delta P_{m2}(t\text{-}2T)$, $\Delta P_{m2}(t\text{-}3T)$, $\Delta P_{m3}(t\text{-}T)$, $\Delta P_{m3}(t\text{-}2T)$, $\Delta P_{m3}(t\text{-}3T)$, $\Delta P_{m4}(t\text{-}T)$, $\Delta P_{m4}(t\text{-}2T)$, $\Delta P_{m4}(t\text{-}3T)$], and the electrical power [$\Delta P_{e1}(t\text{-}T)$, $\Delta P_{e1}(t\text{-}2T)$, $\Delta P_{e1}(t\text{-}3T)$, $\Delta P_{e2}(t\text{-}T)$, $\Delta P_{e2}(t\text{-}2T)$, $\Delta P_{e2}(t\text{-}3T)$, $\Delta P_{e3}(t\text{-}T)$, $\Delta P_{e3}(t\text{-}2T)$, $\Delta P_{e3}(t\text{-}3T)$, $\Delta P_{e4}(t\text{-}T)$, $\Delta P_{e4}(t\text{-}2T)$, $\Delta P_{e4}(t\text{-}3T)$] of Generator 1, 2, 3 and 4. Output signal of the WAM is prediction the speed deviation change of the generator change on power system. The structure of WAM for training process is shown in Figure 3.

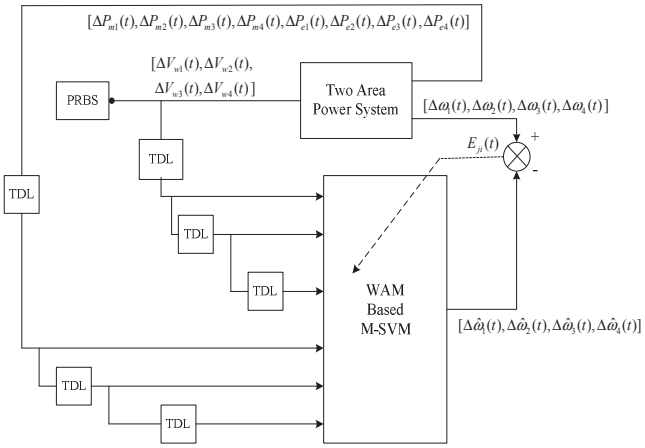


Fig.3. The structure of wide area monitor (WAM) based M-SVM

T is sampling period and the value of T used in this paper is 10 milliseconds or 0.01 seconds. To start the training process of WAM is applying the ΔV_w signal which is pseudo random binary signal signals (PRBS) to the excitation system of each generator. The value of PRBS is generated ±10% p.u. The application of PRBS signals to the system is allows the WAM to learn all possible the dynamic behavior of the systems. *Mean square error* (MSE) is used to evaluate the accuracy of the WAM prediction results. The MSE equation is describe on Equation (1).

$$MSE = \sum_{j=1}^{M} \frac{1}{N} \sum_{i=1}^{N} (\Delta\omega_{ji}(t) - \Delta\hat{\omega}_{ji}(t))^2 \quad (1)$$

where N is the number of training data, M is the number of generators.

B.2. Wide Area Control (WAC)

Design of wide area control (WAC) on power system in this paper also using M-SVM which composed of 24 inputs and 4 outputs. Kernel functions used in this design is radial basis function (RBF). The structure of WAC for training process is shown in Figure 7. The training process of WAC require assistance from WAM that has been trained in the cascade form with desired response predictors $\Delta\omega_d(t+T)$. At this stage, the input signal of the WAC are the mechanical power [$\Delta P_{m1}(t\text{-}T)$, $\Delta P_{m1}(t\text{-}2T)$, $\Delta P_{m1}(t\text{-}3T)$, $\Delta P_{m2}(t\text{-}T)$, $\Delta P_{m2}(t\text{-}2T)$, $\Delta P_{m2}(t\text{-}3T)$, $\Delta P_{m3}(t\text{-}T)$, $\Delta P_{m3}(t\text{-}2T)$, $\Delta P_{m3}(t\text{-}3T)$, $\Delta P_{m4}(t\text{-}T)$,

$\Delta P_{m4}(t\text{-}2T)$, $\Delta P_{m4}(t\text{-}3T)$], and the electrical power [$\Delta P_{e1}(t\text{-}T)$, $\Delta P_{e1}(t\text{-}2T)$, $\Delta P_{e1}(t\text{-}3T)$, $\Delta P_{e2}(t\text{-}T)$, $\Delta P_{e2}(t\text{-}2T)$, $\Delta P_{e2}(t\text{-}3T)$, $\Delta P_{e3}(t\text{-}T)$, $\Delta P_{e3}(t\text{-}2T)$, $\Delta P_{e3}(t\text{-}3T)$, $\Delta P_{e4}(t\text{-}T)$, $\Delta P_{e4}(t\text{-}2T)$, $\Delta P_{e4}(t\text{-}3T)$] of Generator 1, 2, 3 and 4, while the output signal of the WAC is the control signal $U'_j(t) = [\Delta V'_{w1}(t), \Delta V'_{w2}(t), \Delta V'_{w3}(t), \Delta V'_{w4}(t)]$. The input signal of the WAM is PRBS signal U_j = [$\Delta V_{w1}(t)$, $\Delta V_{w1}(t\text{-}T)$, $\Delta V_{w1}(t\text{-}2T)$, $\Delta V_{w2}(t)$, $\Delta V_{w2}(t\text{-}T)$, $\Delta V_{w2}(t\text{-}2T)$, $\Delta V_{w3}(t)$, $\Delta V_{w3}(t\text{-}T)$, $\Delta V_{w3}(t\text{-}2T)$, $\Delta V_{w4}(t)$, $\Delta V_{w4}(t\text{-}T)$, $\Delta V_{w4}(t\text{-}2T)$], the mechanical power [$\Delta P_{m1}(t)$, $\Delta P_{m1}(t\text{-}T)$, $\Delta P_{m1}(t\text{-}2T)$, $\Delta P_{m2}(t)$, $\Delta P_{m2}(t\text{-}T)$, $\Delta P_{m2}(t\text{-}2T)$, $\Delta P_{m3}(t)$, $\Delta P_{m3}(t\text{-}T)$, $\Delta P_{m3}(t\text{-}2T)$, $\Delta P_{m4}(t)$, $\Delta P_{m4}(t\text{-}T)$, $\Delta P_{m4}(t\text{-}2T)$], and the electrical power [$\Delta P_{e1}(t)$, $\Delta P_{e1}(t\text{-}T)$, $\Delta P_{e1}(t\text{-}2T)$, $\Delta P_{e2}(t)$, $\Delta P_{e2}(t\text{-}T)$, $\Delta P_{e2}(t\text{-}2T)$, $\Delta P_{e3}(t)$, $\Delta P_{e3}(t\text{-}T)$, $\Delta P_{e3}(t\text{-}2T)$, $\Delta P_{e4}(t)$, $\Delta P_{e4}(t\text{-}T)$, $\Delta P_{e4}(t\text{-}2T)$]. WAC is applied in parallel with power damping controller (PDC) during training process. The PDC is the aggregation of the external linear controller (ELC) output signals that installed on each generator during the training process. It was conducted because the WAC has not trained well. The structure of the PDC is shown in Figure 5. The output of PDC for Generator 1, 2, 3 and 4 are given by V_{PDCG1}, V_{PDCG2}, V_{PDCG3} and V_{PDCG4}. The maximum and minimum output signal values of PDC is limited $\pm 15\%$ p.u. The weighting factors K_{ij}, where i = 1, 2, ..., N and j = 1, 2, ..., M, to provide good damping on power systems and the PDC is the important part for designing the WAC. The exact value of weighting factor is indispensable in order to achieve overall good control performance. The equation of PDC in Figure 6 is shown in Equation (2).

$$V_{PDCG-j} = \sum_{i=1}^{N} \sum_{j=1}^{M} (K_{ii} V_{ELC-j} + K_{ij} V_{ELC-j}) \quad (2)$$

The structure of ELC that installed on each generator which is part of the PDC design is shown in Figure 5. The signal which used by ELC is the sum of the ELC which uses the mechanical power P_m input signal and the ELC which uses the electrical power P_e input signal of the generator. The maximum and minimum output signal values of ELC is limited $\pm 10\%$ p.u

The desired control signal is calculated via the WAM by comparing the desired response of the system. The changes in speed deviation of the generator is set to zero at any time due to changes in speed deviation of the generator is desired to be zero in every time $\Delta\omega_d(t+T) = 0$. The error that obtained from WAM training process is added to the control signal from the PDC to be target signal used to find the value of the desired control signals during the WAC training process. The desired control signal is shown in Equation (3).

$$U'_j(t) = U_{PDC-j}(t) + E_{ji}(t+T) \quad (3)$$

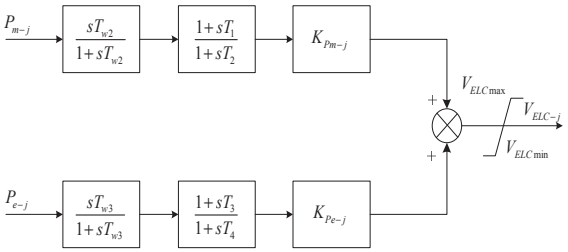


Fig.5. The structure of external linear controller (ELC)

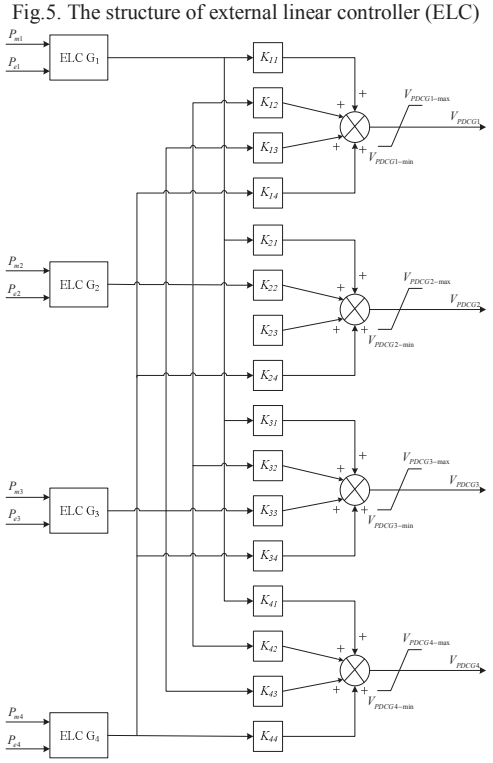


Fig.6. The structure of power damping controller (PDC)

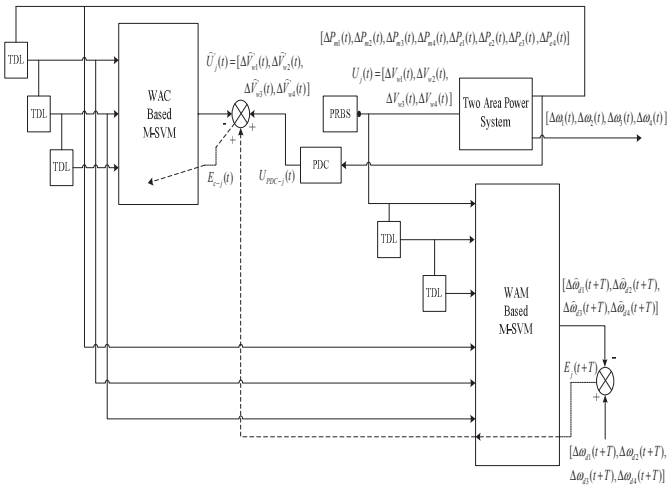


Fig.7. The structure of wide area control (WAC) based M-SVM

The parameter of WAM is unchange during the training process of WAC. The acccury criterion model for the WAC training process is MSE which shown in Equation (4).

$$MSE = \sum_{j=1}^{M} \frac{1}{N} \sum_{i=1}^{N} (U_{ji}^{'}(t) - \hat{U}_{ji}^{'}(t))^2 \quad (4)$$

The accuracy criterion model in Equation (1) and Equation (4) are used for training process of the WAM and the WAC. The WACS structure includes the WAM and the WAC is shown in Figure 8.

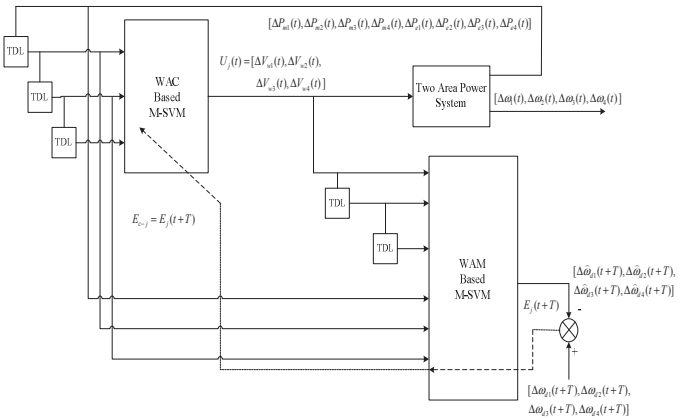


Fig.8. The structure of wide area control system (WACS) based M-SVM

The *comprehensive damping index* (CDI) value which is defined in Equation (6) is used to determine the overall system damping.

$$\zeta_i = \frac{-\sigma_i}{\sqrt{\sigma_i^2 + \omega_i^2}} \quad (5)$$

$$CDI = \sum_{i=1}^{n} \left(1 - \zeta_i\right) \quad (6)$$

where,

ζ_i = the *i*-th *damping ratio*

III. MULTI-OUTPUT SUPPORT VECTOR MACHINE (M-SVM)

The basic theory of multi output (M-SVM) used in this paper is defined as below [5]:

$$\min_{w,b} j(w,b) = \frac{1}{2}\left\|w^j\right\|^2 + \gamma \sum_{i=1}^{l} \xi_i \quad (7)$$

where,

$$\begin{cases} \left\|y_i - f(x_i)\right\| y_i \le \varepsilon + \xi_i, i = 1,2,...,l \\ \xi_i \ge 0, \qquad i = 1,2,...,l \end{cases} \quad (8)$$

From Equation (7) can be written as unconstrained optimization problem as below:

$$\min_{w,b} L_{MSVM}(w,b) = \frac{1}{2}\left\|w^j\right\|^2 + \gamma \sum_{i=1}^{l} L_\varepsilon(u_i) \quad (9)$$

The iterative re-weighted least-square (IRWLS) method is used to solve the optimization problem in Equation (9). To obtain the next solution (W^{t+1}, b^{t+1}) from previous solution (W^t, b^t) during the iteration process, $L_P(W,b)$ is using the first-order taylor expansion approach method.

$$L_P(W,b) \approx L_P^{''}(W,b) = \frac{1}{2}\sum_{j=1}^{k}\left\|w^j\right\|^2 + \frac{1}{2}\sum_{j=1}^{k}\alpha_i u_i^2 + \tau \quad (10)$$

where τ is sum of the constant that not depend on W or b; β_i provided by:

$$\alpha_i = \begin{cases} 0, & u_i \le \varepsilon \\ \dfrac{2\gamma(u_i - \varepsilon)}{u_i}, & u_i \le \varepsilon \end{cases} \quad (11)$$

The Equation (11) shows the problem to be solved using the IRWLS method. When (W^t, b^t) has been obtained, the optimization problem corresponding to $L_P(W, b)$ can be converted into searching the optimal solution of $L_p^{''}(W,b)$.

Based on the stationary point condition $\frac{\partial L_p^{''}(W,b)}{\partial w^j} = 0$ and $\frac{\partial L_p^{''}(W,b)}{\partial b^j} = 0$ can be obtained Equation (12).

$$\begin{bmatrix} \Phi^T D_\alpha \Phi & I\Phi^T \alpha \\ \alpha^T \Phi & \alpha^T 1 \end{bmatrix} \begin{bmatrix} W^j \\ b^j \end{bmatrix} = \begin{bmatrix} \Phi^T D_\alpha y^j \\ \alpha^T y^j \end{bmatrix}, \; j = 1,2,...,k \quad (12)$$

where $D_\alpha = diag(\alpha_1,...,\alpha_l)$, $\Phi = [\varphi(x_1),...,\varphi(x_l)]^T$, $\alpha = [\alpha_1,...,\alpha_l]^T$ dan $y^j = [y_{j1},...,y_{jl}]$.

The problem in machine learning can be expressed as a linear combination of training samples in the feature space,

$$w^j = \sum_{i=1}^{l} \beta_{ij}\varphi(x_i) = \Phi^T \beta^j, \; j = 1,2,...,k \quad (13)$$

where $\beta^j = [\beta_{j1},...,\beta_{jl}]^T$. If the Equation (13) is substituted into Equation (12) then It can be obtained the Equation (14).

$$\begin{bmatrix} K + D_\alpha^{-1} & 1 \\ \alpha^T K & 1^T \end{bmatrix} \begin{bmatrix} \beta^j \\ b^j \end{bmatrix} = \begin{bmatrix} y^j \\ \alpha^T y^j \end{bmatrix}, \; j = 1,2,...,k \quad (14)$$

where $(K)_{ij} = k\left(x_i, x_j\right) = \varphi(x_i)^T \varphi(x_j)$ is kernel matrix that must satisfy the Mercer condition $(K)_{ij}$ and match with the inner product of the feature space. The kernel functions used in this paper is shown in Equation (15).

$$K_{ij} = k\left(x_i, x_j\right) = \exp\left(-\frac{\left|x_i - x_j\right|}{\sigma^2}\right) \quad (15)$$

IV. SIMULATION, ANALYSIS AND RESULT

This section explained the design of simulation and analysis of simulation result as follow:

The simulation is performed using PC Pentium Intel (R) Core (TM) i3 CPU 380M@2.5 2.00 GHz GB RAM. The main purpose of this simulation is to evaluate the performance and accuracy of the proposed method. Simulation and analysis in this paper are conducted using MATLAB software. To evaluate the performance of WACS based MLS-SVM, two area power system-4 generator in Figure 1 is used as tested system. The data used in this system includes the generator data, governors, excitation systems, transformers and load can be seen in reference [4,6]. A block diagram of conventional PSS and its parameter that used in this paper can be seen in reference [7]. The data parameter of ELC and weigth factor of PDC are shown in Table I and II, respectively. The parameters of WAM based M-SVM and WAC based M-SVM are presented in Table III.

TABLE I. THE PARAMETERS OF ELC AT GENERATOR 1 TO GENERATOR 4

	K_{Pm}	T_{w2}	T_1	T_2	K_{Pe}	T_{w3}	T_3	T_4
Generator 1-4	5	10	0.05	0.02	1.462	10	0.03	0.01

TABLE II. THE WEIGTH FACTOR OF PDC FOR GEN.1 - GEN.4

K_{11}	=	0.95	K_{31}	=	0.02
K_{12}	=	0.13	K_{32}	=	0.04
K_{13}	=	0.07	K_{33}	=	1.07
K_{14}	=	0.23	K_{34}	=	0.02
K_{21}	=	0.12	K_{41}	=	0.12
K_{22}	=	1.15	K_{42}	=	0.17
K_{23}	=	0.24	K_{43}	=	0.21
K_{24}	=	0.16	K_{44}	=	1.05

TABLE III. THE WAM AND WAC PARAMETERS

	WAM	WAC
σ	750	90
γ	8	0.8

To evaluate and validate the robustness of the WACS application on power system, the power system is disturbed with small step signal 0.03 p.u in Generator 1. The prediction results of speed deviation for Generator 1 and 3 using the WAM based M-SVM in testing phase are shown in Figure 9 to Figure 10. The MSE value of WAM based M-SVM prediction result is 3.3743e-010.

The speed deviation of Generator 1 and 3 using the WAC based M-SVM in testing phase are shown in Figure 11 to Figure 12. The performance of speed deviation for Generator 1 and 3 are described in Figure 11 to Figure 12.

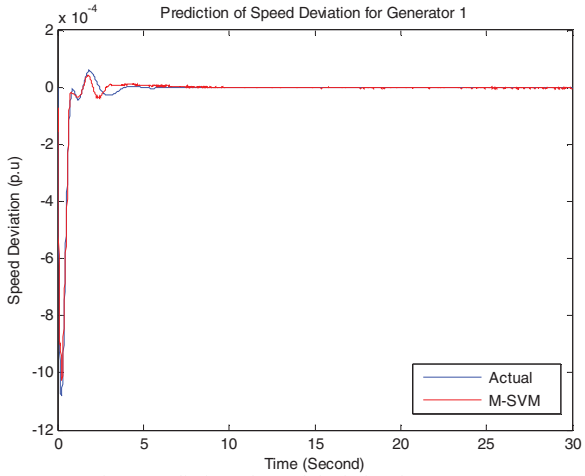


Fig.9. Prediction of speed deviation for Generator 1

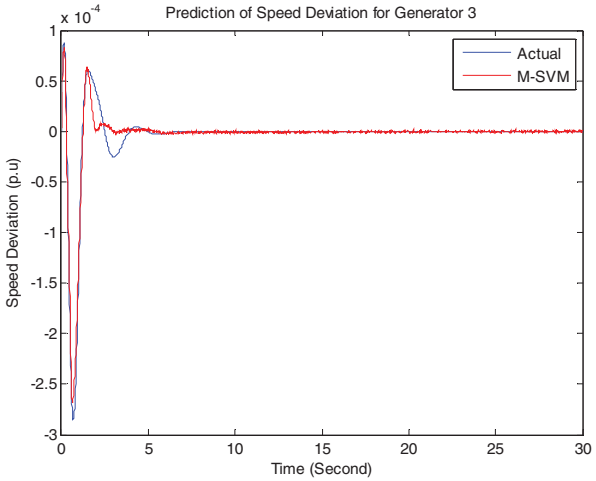


Fig.10. Prediction of speed deviation for Generator 3

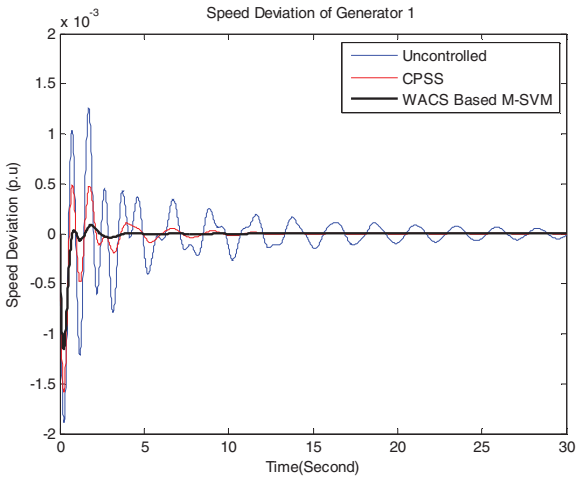


Fig.11. Performance of speed deviation for Generator 1

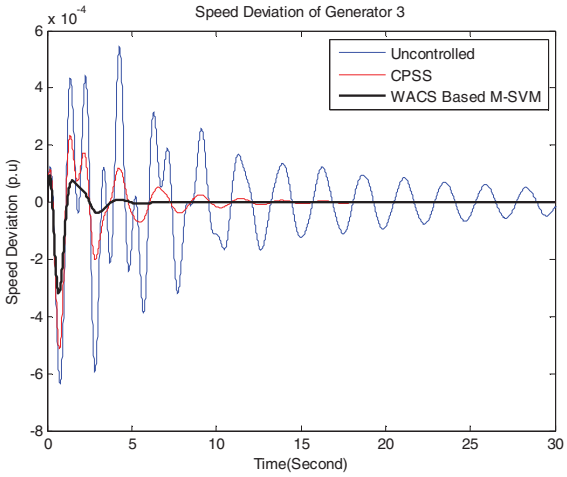


Fig.12. Performance of speed deviation for Generator 3

TABLE IV. OVERSHOOT OF GENERATOR SPEED DEVIATION (P.U)

	G_1	G_3
Uncontrolled	-0.0018590	0.0001212
CPSS	-0.0015610	0.0001141
WACS based M-SVM	-0.0011540	9.262e-005

TABLE V. SETTLING TIME OF GENERATOR SPEED DEVIATION (P.U)

	G_1	G_3
Uncontrolled	>30	>30
CPSS	13.09	15.52
WACS based M-SVM	5.39	7.3

From Figure 11 to Figure 12 it can be observed that the power system equipped with CPSS or WACS based M-SVM can reduce the overshoot oscillations and speed up the settling time of the system. Table IV and V show the overshoot and the settling time data result for the generator speed deviation from Figure 11 to Figure 12.

The comparison of comprehensive damping index (CDI) value of the system using CPSS, WAC based M-SVM and without controller are shown in Table VI.

TABLE VI. *COMPREHENSIVE DAMPING INDEX* (CDI)

	***Comprehensive Damping Index* (CDI)**
Uncontrolled	18.3157
CPSS	17.3724
WACS based M-SVM	15.3156

From Table VI it can be seen that the comprehensive damping index (CDI) value of uncontrolled system is 18.3157 to be 17.3724 using CPSS, the application of WACS based M-SVM is able to minimize the CDI value to be 15.3156. The WACS based M-SVM has a minimum value of CDI compared to another method which means WACS based M-SVM is able to reduce the oscillations that occur on power system better than other methods.

CONCLUSION

A new approach in designing the wide area control system (WACS) based M-SVM for oscillation damping on a two-area four-generator power system has been proposed in this paper. Wide area monitor (WAM), which is part of the WACS able to learn the dynamic behavior of the system properly and used as the reference for wide area control (WAC) in order to generate the appropriate control signals to damp the oscillations on power system. Based on the observation of speed deviation oscillations, WACS based M-SVM can reduce the oscillations, minimize the overshoot and accelerate the settling time of the speed deviation better than the aplication of CPSS and without controller.

ACKNOWLEDGMENT

The first author would like highly acknowledge gratitude to Indonesian Government especially the Directorate General of Higher Education (DIRJEN-DIKTI) for the Excellency Scholarship (Beasiswa Unggulan) that awarded during his study at the Graduate Program, Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia. The authors are grateful to The Laboratory of Instrumentation, Measurement, and Power System Identification and The Power System Simulation Laboratory, Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember (ITS), Surabaya for all facilities provided during this research.

REFERENCES

[1] Swakshar Ray a, Ganesh K. Venayagamoorthy, "A wide area measurement based neuro control for generation excitation systems", Engineering Applications of Artificial Intelligence vol.22, pp.473–481, 2009.

[2] Qiao. W., Venayagamoorthy G.K., Harley R.G.: "Optimal widearea monitoring and non-linear adaptive coordinating control of a power system with wind farm integration and multiple FACTS devices", Neural Netw., 2008, 21, (2–3), pp. 466–475.

[3] Mohagheghi s., Venayagamoorthy g.k., Harley r.g.: "Optimal wide area controller and state predictor for a power system', ieee trans. Power syst., 2007, 22, (2), pp. 693–705.

[4] Kundur, P., "Power system stability and control", EPRI. NewYork : McGraw-Hill, 1994.

[5] Dongjian Zheng, Lin Cheng, Tengfei Bao, Beibei Lv, Integrated parameter inversion analysis method of a CFRD based on multi-output support vector machines and the clonal selection algorithm, Computers and Geotechnics, vol.47, pp.68–77, 2012.

[6] Muhammad Abdillah, Adi Soeprijanto, Mauridhy Hery Purnomo, "Design of wide area monitoring on power system using least square support vector regression", JAVA Journal of Electrical and Electronics Engineering, Vol.10, No.1, April 2012.

[7] Sidhartha Panda and N. P. Padhy, "*Robust Power System Stabilizer Design using Particle Swarm Optimization Technique*", International Journal of Electrical and Electronics Engineering 1:1 2008.

A Novel Second-Order Model of Induction Motor Loads

Pichai Aree
Department of Electrical Engineering, Thammasat University
KlongLuang, Pathumthani, Thailand, 12120
E-mail: apichai@engr.tu.ac.th

***Abstract*—In power system dynamic studies, it is necessary to represent induction motor loads with a standard reduced order model in order to reduce computational requirements. A third-order model, neglecting the stator transients, is often used. In this paper, the third-order model is further simplified into two second-order models, called as slip-flux-magnitude and slip-flux-angle models. The model derivations are given in detail. The dynamic responses of both models are compared with the higher third-order model for the cases of small and large horse power induction motor loads. The study results show that the slip-flux-angle model gives better approximation of the third-order model.**

Keywords—Induction motor load; second-order model; Induction motor modeling;

I. Introduction

A major proportion of power system loads consists of induction motor loads. Therefore, an accurate representation of them is important in power system dynamic studies. Although the full fifth-order model of induction motor loads has been considered to be the most accurate model, the required computation time is very large if detailed induction motor loads are applied with large-scale power system studies. Hence, in power system simulation such as transient stability studies, it is desirable to represent induction motor loads by a standard reduced order model [1-5]. The order reduction is achieved by setting derivatives of stator flux linkages in direct (*d*) and quadrature (*q*) axes, which appear in the full order model, to zero in the stator differential equations [6-8]. The fact that the stator flux linkages change much faster than the rotor flux linkages is the basic argument of ignoring the stator electrical transients. This reduced order model is known as a third-order model. It has been widely used in power system transient stability analysis because it provides a good indication of transient active and reactive power flows following a voltage disturbance. Since the transients in transmission lines are usually ignored in this analysis, neglecting the stator's transient term is very useful.

In addition to the stator transient, electrical transients in the rotor windings are further neglected by setting the derivatives of rotor flux linkages in *d*- and *q*-axis to zero. This reduced order model is known as a first-order model where only differential equation left is the speed or slip dynamics [3-4, 6-8]. This model is suitable for long term power system dynamic studies.

In this paper, two second-order models of induction motor loads are further developed from the third-order model. The first model is called as slip-flux-magnitude model, where has firstly appeared in [6]. The second model, proposed by the author, is named as slip-flux-angle model. Both models are validated and compared with the higher third-order model through dynamic simulation responses. Accuracies of these two second-order models are also discussed.

II. Derivaration of Second-Order Model Induction Motor Load

The well-established representation of an induction motor is the two-axis fifth-order model. The derivation is based on the generalized theory in a common arbitrary frame of reference. In power system studies, the synchronously rotating reference frame is mostly used. The machine state variables are fully expressed in *d*- and *q*-axis, which consist of stator linkage flux dynamics (ψ_{ds} , ψ_{qs}) rotor linkage flux dynamics (ψ_{dr} , ψ_{qr}) and angular speed (ω_r) or slip (s). In a viewpoint of power system load modeling, the fifth-order model is often further simplified by ignoring the stator flux dynamics. This model is known as the third-order model [5], where differential and algebraic equations are:

$$\frac{d\bar{E}'_d}{dt} = s\omega_s \bar{E}'_q - \frac{1}{\tau'_m}\left(\bar{E}'_d + (\bar{X}_s - \bar{X}')\bar{I}_q\right) \tag{1}$$

$$\frac{d\bar{E}'_q}{dt} = -s\omega_s \bar{E}'_d - \frac{1}{\tau'_m}\left(\bar{E}'_q - (\bar{X}_s - \bar{X}')\bar{I}_d\right) \tag{2}$$

$$\frac{ds}{dt} = \frac{1}{2\bar{H}}\left(\bar{T}_m - \bar{T}_e\right) \tag{3}$$

$$\bar{V}_d = \bar{E}'_d - \bar{X}'\bar{I}_q + \bar{R}_s\bar{I}_d \tag{4}$$

$$\bar{V}_q = \bar{E}'_q + \bar{X}'\bar{I}_d + \bar{R}_s\bar{I}_q \tag{5}$$

Where,

$$\bar{T}_e = \bar{E}'_q\bar{I}_q + \bar{E}'_q\bar{I}_q \tag{6}$$

$$\bar{T}_m = \bar{T}_m^0\left(A(1-s)^2 + B(1-s) + C\right) \tag{7}$$

It should be noted that the topbar sign ($^-$) indicates per-unit quantities. $\bar{E}'_q$ and $\bar{E}'_d$ are per-unit transient electromotive

This work is sponsored by Thammasat University academic affairs, 2013.

force (emf) behind transient reactance $\bar{X}'$ in *d*- and *q*-axis, respectively. τ'_m is rotor open-circuit time constant. $\bar{X}_s$ is stator self reactance. In order to develop the motor's second-order models for power system load representation, the internal emf should be expressed in term of polar coordination since it can give a clear relationship between voltage and/or frequency and dynamics of electrical power supplied to the motor. The transient emf ($\bar{E}'_d$ and $\bar{E}'_q$) and bus voltage ($\bar{V}_d$ and $\bar{V}_q$) are then expressed in term of magnitudes ($\bar{E}'$ and $\bar{V}$) and phase angles (θ' and θ), respectively. For example, the transient emf can be converted into polar coordination using the following relationships,

$$\bar{E}'^2 = \bar{E}'^2_d + \bar{E}'^2_q \tag{8}$$

$$\theta' = \tan^{-1}\left(\bar{E}'_d / \bar{E}'_q\right) \tag{9}$$

Differentiating (8) and (9), we have,

$$\bar{E}'d\bar{E}' = \bar{E}'_d d\bar{E}'_d + \bar{E}'_q d\bar{E}'_q \tag{10}$$

$$\bar{E}'^2 d\theta' = \bar{E}'_q d\bar{E}'_d + \bar{E}'_d d\bar{E}'_q \tag{11}$$

Re-arranging (10) and (11) by making use of (1)-(5), the motor's internal emf dynamics, ignoring of the stator resistance, can be expressed in term of the polar coordination as,

$$\frac{d\bar{E}'}{dt} = -\frac{\bar{X}_s}{\tau'_m \bar{X}'}\bar{E}' + \frac{\bar{X}_s - \bar{X}'}{\tau'_m \bar{X}'}\cos(\theta - \theta')\bar{V} \tag{12}$$

$$\frac{d\theta'}{dt} = -s\omega_s + \frac{\bar{X}_s - \bar{X}'}{\tau'_m \bar{X}'}\frac{\bar{V}}{\bar{E}'}\sin(\theta - \theta') \tag{13}$$

Similarly, the electrical torque of induction motor in (6) can be expressed by,

$$\bar{T}_e = \frac{\bar{E}'\bar{V}}{\bar{X}'}\sin(\theta - \theta') \tag{14}$$

A. Slip-flux-magnitude model

The slip-flux-magnitude model takes motor's slip and rotor flux linkage magnitude directly related to the transient internal emf $\bar{E}'$ into account as state variables. The modification of this model is based on the assumption that the rotor flux angle varies more rapidly than the flux magnitude as mentioned in [6]. By setting the derivative term of rotor emf angle in (13) to zero, the motor slip can be then written by,

$$s\omega_s = \left((\bar{X}_s - \bar{X}')/(\tau'_m \bar{X}')\right)\left(\bar{V}/\bar{E}'\right)\sin(\theta - \theta') \tag{15}$$

By making use of (15), the rate of change of the internal emf magnitude in (12) can be modified as,

$$\frac{d\bar{E}'}{dt} = -\frac{\bar{X}_s}{\tau'_m \bar{X}'}\bar{E}' + \frac{\bar{X}_s - \bar{X}'}{\tau'_m \bar{X}'}\sqrt{1 - \left(\frac{s\omega_s \bar{X}'\tau'_m}{(\bar{X}_s - \bar{X}')}\frac{\bar{E}'}{\bar{V}}\right)^2} \tag{16}$$

The slip-flux-magnitude model in second-order form is represented by (3) together with (16). The electrical torque in (14) can be then modified making use of (15) as,

$$\bar{T}_e = \frac{s}{\bar{R}_r(\bar{X}_m / \bar{X}_r)^2}\bar{E}'^2 \tag{17}$$

B. Slip-flux-angle model

In contrast to the slip-flux-magnitude model, where the emf angle is assumed to decay rapidly, the slip-flux angle model considers a change in phase angle of the transient emf since it plays an important role on dynamic responses of the induction motor. This fact can be illustrated from the plots of magnitude and phase angle of the emf, as shown in Fig. 1 and 2, respectively. The responses consist of two study cases where small 50HP and very large 2250HP induction motors are considered. The mechanical shaft torques of both motor are perturbed starting at *t*=0.2sec for 0.6sec duration. It is noted that the full third-order model of induction motors is used to generate the responses in Fig. 1 and 2 using Matlab program. The motor's parameters are given in Appendix A.

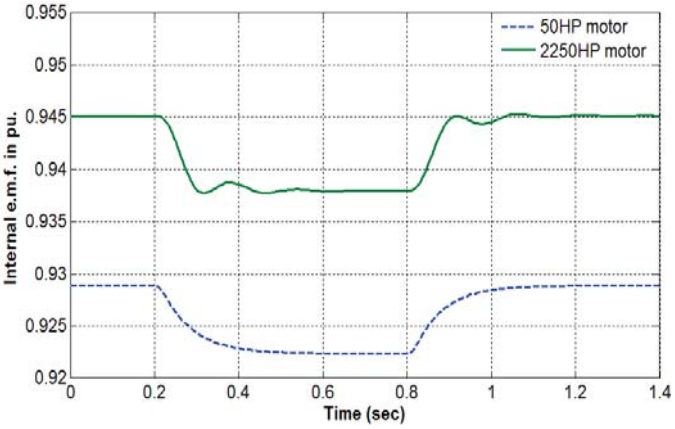


Fig. 1. Responses of emf's magnitude of 50HP and 2250HP motors

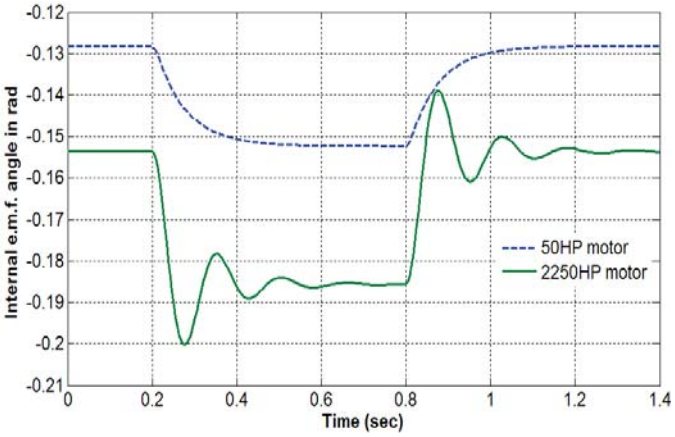


Fig. 2. Responses of emf's phase angle 50HP and 2250HP motors

It is evident that after applying an increase in the mechanical shaft torque via 0.2pu step change, the emf magnitudes of both motors suddenly drop with further decreases in the emf phase angles (They become more negative). For the case of small 50HP motor, the dynamic responses of emf magnitude and phase angle are very close to

the first-order approximation. On the other hand, the oscillation is severely pronounced for the case of large 2250HP motor. The oscillation is greatly exhibited through the emf phase angle response rather the magnitude. In order to preserve most of the dynamic integrity of induction motor load, the emf phase angle and slip are thus chosen as the state variables in this paper. The emf magnitude is then kept constant during transient period. The transient emf magnitude can be modified by setting the derivative term in (12) to zero as,

$$\bar{E}' = \frac{\bar{X}_s - \bar{X}'}{\bar{X}_s}\bar{V}\cos(\theta - \theta') \tag{18}$$

By making use of (18), the emf phase angle in (13) can be modified as,

$$\frac{d\theta'}{dt} = -s\omega_s + \frac{\bar{X}_s}{\tau_m' \bar{X}'}\tan(\theta - \theta') \tag{19}$$

Next, making the second derivative in (19) gives,

$$\frac{d\theta'^2}{dt^2} = -\omega_s \frac{ds}{dt} - \frac{\bar{X}_s}{\tau_m' \bar{X}'}\sec^2(\theta - \theta')\frac{d\theta'}{dt} \tag{20}$$

By rearranging (20) with making use of (3), the second-order model of induction motor load is newly arrived as,

$$\frac{d\theta'^2}{dt^2} + \frac{\bar{X}_s}{\tau_m' \bar{X}'}\sec^2(\theta - \theta')\frac{d\theta'}{dt} - \frac{\omega_s}{2\bar{H}}\left(\bar{T}_e - \bar{T}_m\right) = 0 \tag{21}$$

Where the electrical torque in (14) can be then modified making use of (18) as,

$$\bar{T}_e = 0.5\left(\frac{1}{\bar{X}'} - \frac{1}{\bar{X}_s}\right)\sin\left(2(\theta - \theta')\right)\bar{V}^2 \tag{22}$$

III. Results and Discussions

In this section, the dynamic responses of slip-flux-magnitude and slip-flux-angle models are verified and compared. The study considers a case where the induction motor load is directly connected to the supply. The small 50HP and large 2250HP induction motors are considered. After applying a step change in the mechanical shaft torque (0.2pu) for 0.6sec duration, the plots of electrical torque and slip are displayed in Fig. 3-6. It is apparent from Fig. 3 and 4 that transient responses of torque and slip obtained from both slip-flux-angle and slip-flux magnitude models agree very well with those of the higher third-order model, when both models are employed for representing the small 50HP motor. However, it is evident from Fig. 5 and 6 the slip-flux-magnitude model fails to capture the dynamic responses of the particular large 2250HP motor. Only slip-flux-angle model gives the electrical torque and slip responses almost the same as the third-order model does. Since the input active power and air-gap torque of motors are changed in coresponding to a variation in the angles between the internal emf and terminal voltage, their electromechanical oscillations could greatly be related to this varition. Thus, preserving the flux angle as state variable through the slip-flux-angle model could be more relevant in oder to get more accurarte dynamic profiles.

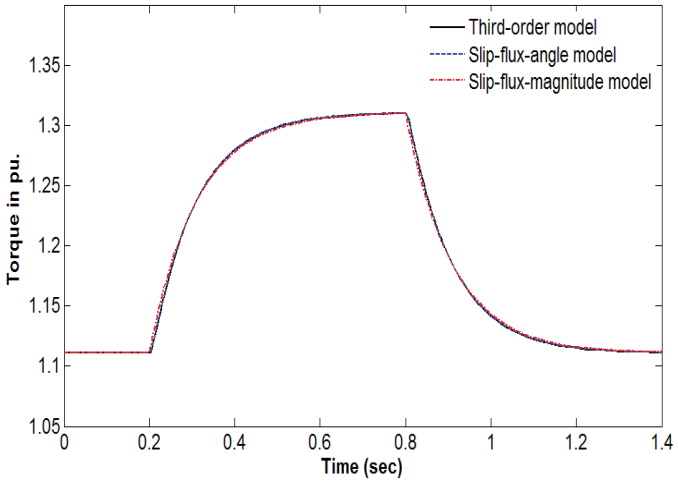


Fig. 3. Electical torque responses of 50HP motor

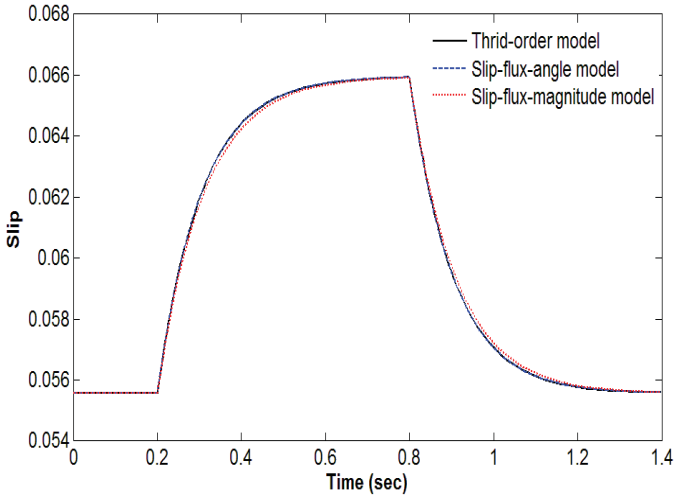


Fig. 4. Slip responses of 50HP motor

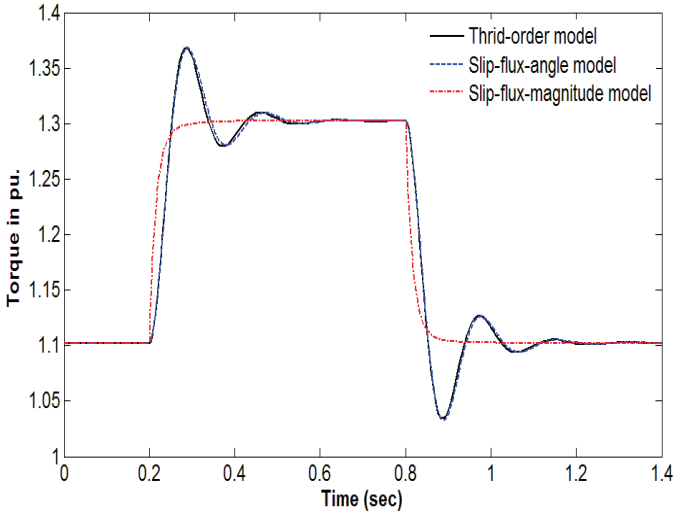


Fig. 5. Electrical torque responses of 2250HP motor

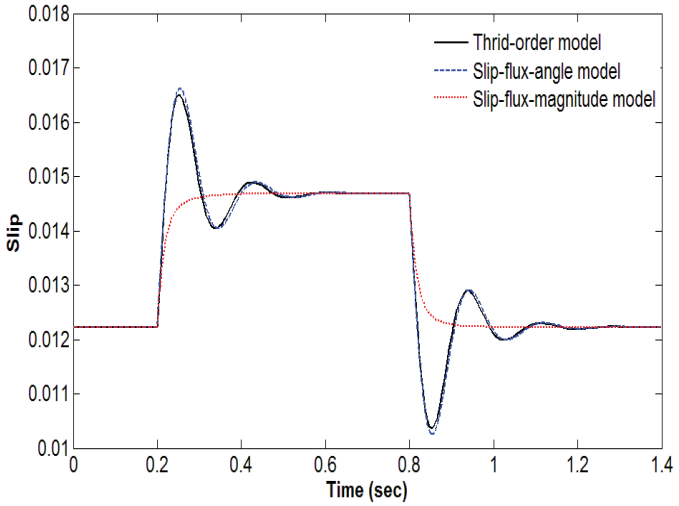


Fig. 6. Slip responses of 2250HP motor

IV. Conclusions

In this paper, a novel second-order slip-flux-angle model of induction motor is newly derived in addition to the slip-flux- magnitude model. The models are verified and compared with the third-order model. The results indicate that the slip-flux-magnitude model provides an accurate dynamic response only for small induction motor load representation. It fails to capture dynamic responses of the particular large HP motors. The slip-flux-angle model alternately gives a closer approximation to the higher third-order model. Therefore, the slip-flux-angle model could be a candidate for power system dynamic load representation when the large HP induction motor loads are the main of concern.

Appendix A: Parameters of 50HP and 2250HP motors in SI unit [9]

HP	R_s	R_r	L_{ss}	L_{rr}	L_m	f	V
50	0.087	0.228	0.0355	0.0355	0.0347	60	460
2250	0.029	0.022	0.0352	0.0352	0.0346	60	2300

References

[1] T. Aboul-Seoud, and J. Jatskevich, "Dynamic modeling of induction motor loads for transient voltage stability studies," EPEC Conference, pp. 1-5, October 2008.

[2] A. Borghetti, R. Caldon, A. Mari, and A. A. Nucci, "One dynamic load models for voltage stability studies," IEEE Trans. Power Sys., vol. 12, pp. 293-303, Febuary 1997.

[3] C. D. Vournas, and G. A. Manos, "Modelling of stalling motors during voltage stability studies," IEEE Trans. Power Sys., vol. 13, pp. 775-781, August 1998.

[4] M. A. Merkle, and A. M. Miri, "Modelling of industrial loads for voltage stability studies in power systems," Canadian Conference on Electrical and Computer Engieering, vol. 2, pp. 881-886, May 2001.

[5] P. Kundur, Power System Stability and Control. McGraw-Hill, 1994.

[6] G. G. Richards, and O. T. Tan, "Simplifed model for inudction machine trasients under balanced and unbalanced condition," IEEE Trans. Ind. Appl, vol. IA-17, pp. 15-21, January/Febuary 1981.

[7] N. Derbel, M. B. A. Kamoun, and M. Poloujadoff, "On the order reduction of induction machine during start-up," IEEE Trans. Energy Convers., vol. 10, pp. 655–660, December 1995.

[8] T. Thiringer, and J. Luomi, "Comparison of reduced-order dynamic models of induction machines," IEEE Trans. Power Sys., vol. 16, pp. 119-126, Febuary 2001.

[9] J. J. Cathey, R. K. Cavin, and A. K. Ayoub, "Transient load model of an induction machine," IEEE Trans. Power App. Syst., vol. 92, pp. 1399-1406, July 1973.

Characteristics of Electric Field Change Preceding Negative First Return Stroke Produced by Preliminary Breakdown

Ariadi Hazmi, Zulka Hendri[1], Syaifa Mulyadi[1], Dasrinal Tesal[1]
Dept. of Electrical Engineering
Atmospheric Electricity Research Division
[1]Graduate School of Electrical Engineering
Andalas University
Padang, Indonesia
ariadi@ft.unand.ac.id

Daohong Wang, Nobuyuki Takagi
Dept. of Electrical, Electronic and Computer Engineering
Gifu University
Gifu City, Japan

***Abstract*—Understanding the initiation process of preliminary breakdown during negative CG lightning discharges to the ground is important to design a lightning protection system. We observed electric field change preceding negative cloud to ground lightning flashes related to preliminary breakdown in tropical region, Indonesia. The characteristics of electric field change of negative first return stroke produced by preliminary breakdown process were studied. The results showed that arithmetic mean and geometric mean of pre-return stroke separation is 57 ms and 32 ms, respectively. Other statistical parameters also are shown in this paper.**

Keywords—electric field change; return stroke; negative cloud to ground; preliminary breakdown, leader

I. Introduction

The lightning discharge processes can be studied by the electric field change observation. The electric field change observation show that almost 90% of all negative cloud-to-ground (CG) lightning flashes that are initiated in the cloud or intracloud, initially develop in downward leader, and transport negative charges to ground. The electric field changes preceding first return stroke in the negative CG lightning flashes have been published by many researchers.

Clarence and Malan [1] reported that the preliminary breakdown process was a result of a vertical discharge between the main negative charge center and the lower positive charge center inside the thundercloud, which has a duration in range 2–10 ms. However, other researchers found that more than 100 ms of electric field changes preceded the first return stroke had a horizontal channel during the development of preliminary breakdown process preceding a step leader [2-4]. The preliminary breakdown (B) is followed by the stepped leader (L) either immediately or after an intermediate (I) stage, which may have the duration of up to 400 ms. Furthermore, the intermediate stage was the charging process of the negative charge from the vertical channel of the preliminary breakdown until the electric field at the bottom of the channel was sufficient to initiate a stepped leader [5]. Understanding the initiation process of preliminary breakdown and stepped leader during negative CG lightning discharges to the ground is important to design a lightning protection system and study the lightning physic.

In this paper, we aimed to report the characteristics of the electric field changes preceding negative first return stroke (RS) especially on the relation to electric field changes in the preliminary breakdown process in equator area.

II. Observation

The observation has being aimed at understanding the initiation of negative first return stroke of lightning discharges in equator area (Latitude: $0^o54'$ S, Longitude: 100^o E). Our observation site is located in Padang City, Indonesia. We have recorded the electric field change of lightning flashes by using a parallel antenna to sense fast electric field change with the time constant is 100 ms. In addition, we have recorded the thunderstorm electric field with an electric field mill. The gap and diameter of parallel flat plate antenna is 0.1 and 0.3 m, respectively. The flat plate antenna and the field mill is to sense electric field change are located above roof level at a height 16 m from the ground. Thus, unlike electric field antenna, which respond to fast transients in the electromagnetic field generated by lightning, electric field mill detect the electrostatic field that the electric field changes is relatively slow. The observation of electric fields generated by the negative CG flashes were recorded from January to March 2013. Our measurement station altitude is about 317 m above sea level and 13 km away from the Padang Beach. The signals

from fast electric field parallel antenna were recorded with 12 bit, 20 MHz digital recorder, the sampling rate was set to 1 MS/s with the total length recorded was 1 s. The digital recorder was operated in 300 ms in pre-trigger mode. The trigger type of the digital recorder was window type that was set to be able to capture such the signals of both polarities of electric field change.

III. Results and Discussion

We examined the data set with a total of 94 electric fields of negative CG flashes which satisfied the criteria for the determination of initial state of electric field changes associated to the preliminary breakdown and negative first return stroke. The close distance of the negative ground flashes with a distance less than 50 km were selected with a criteria that all the lightning data have no zero crossing on negative first return stroke [6].

From an example of the data recorded by the electric field mill (20130312), the thunderstorm activity lasts for two hours and the lightning frequency is about 1 per a minute is shown in Figure 1. It is a typical thunderstorm in tropical area. The thunderstorm took place between 14:00 and 16:00, Padang Time, March 12, 2013. The peak value of surface of electric field was about 20 kV/m during the thunderstorm.

The electric field change produced by negative CG lightning flashes were detected by fast antenna system. The waveforms recorded by the fast antenna are all negative

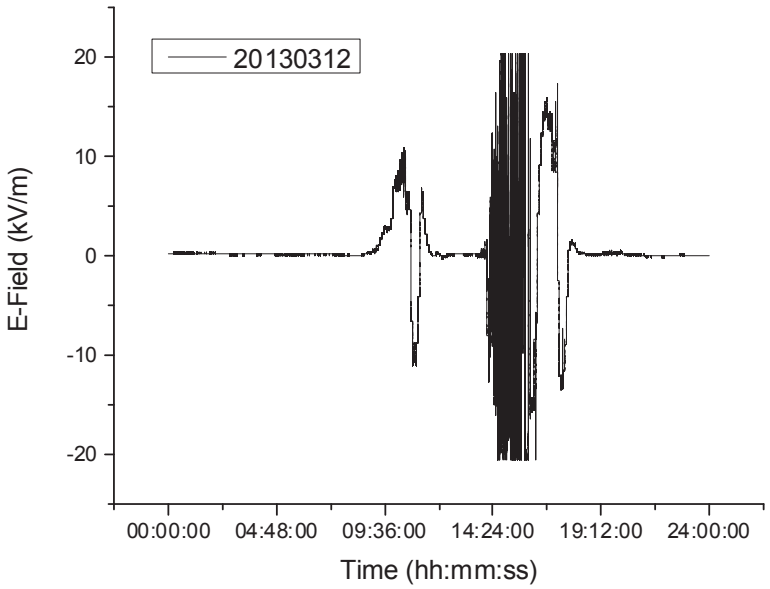


Fig. 1. Thunderstorm electric field on March 12, 2013 at Padang City

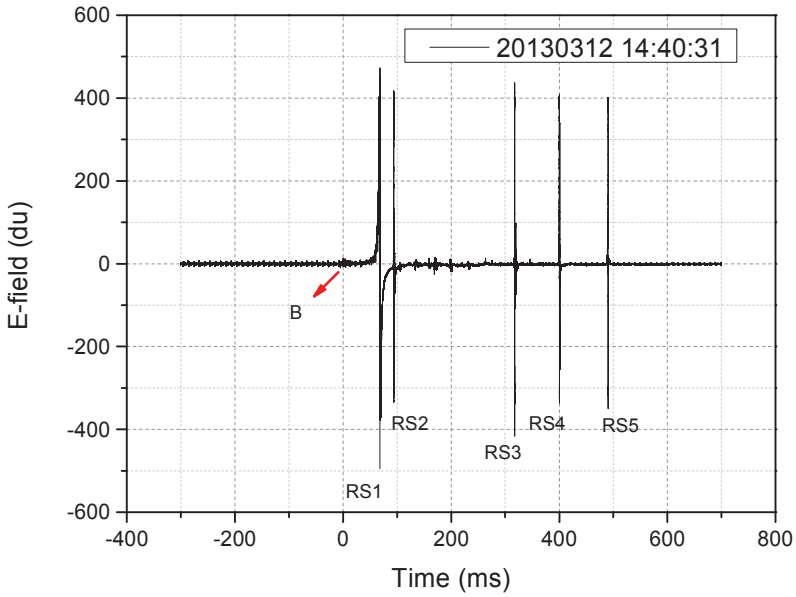


Fig. 2. Electric field change produced by negative CG lightning flash

lightning flashes with a preliminary breakdown process, a stepped leader followed by a first return stroke. Negative CG discharge usually consist of multiple return strokes to ground. One of the negative CG waveforms (20130312 14:40:31) struck to ground estimated within 8 kilometers from the observation site was shown in Figure 2. Five return strokes are marked with RS1, RS2, RS3, RS4, and RS5, with time intervals variation from 25 ms to 223 ms, respectively. The time expanded waveform of Figure 2 displayed in Figure 3, which also showed the electric field change preceding negative first return stroke RS1 produced by preliminary breakdown (B) and followed by rapid electric field change of stepped leader. The negative CG lightning flashes on March 12, 2013 caused five farmers died immediately.

In this study, we used the parameter of pre-return stroke separation to determine the time interval between the peaks of first preliminary breakdown pulse and first return stroke. Ninety four negative CG lightning discharges to ground with obvious impulsive electric field changes preceding the first return strokes have been analyzed. Figure 4 shows a histogram of the distribution of pre-return stroke separation about 70% is less than or equal 50 ms, lies in range from 5 ms to 50 ms. Pre-return stroke separation varies from 5 ms to 328 ms with arithmetic mean (AM), geometric mean (GM) and median values are 57 ms, 32 ms, and 27 ms, respectively. The arithmetic mean in this study support [7]. The result indicated that the characteristic of electric field change both country showed the similarity. The summary of statistical result of pre-

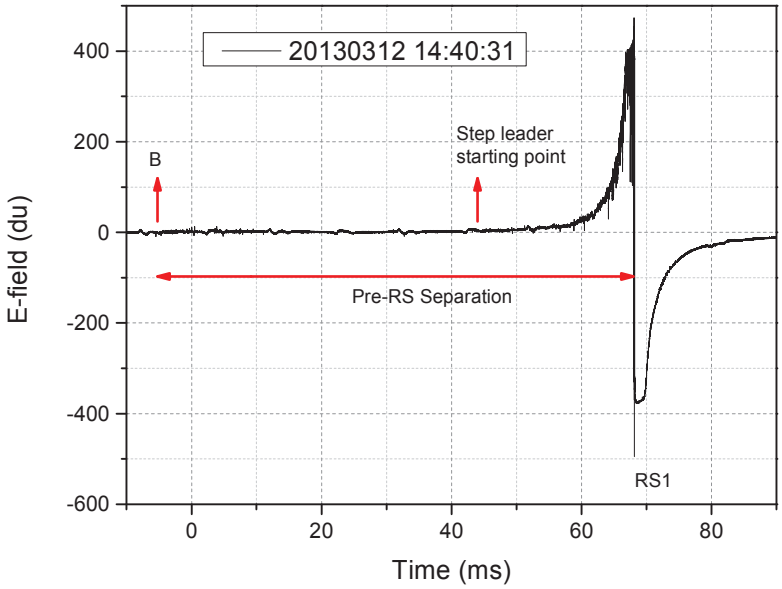


Fig. 3. The expanded electric field change of Fig. 2

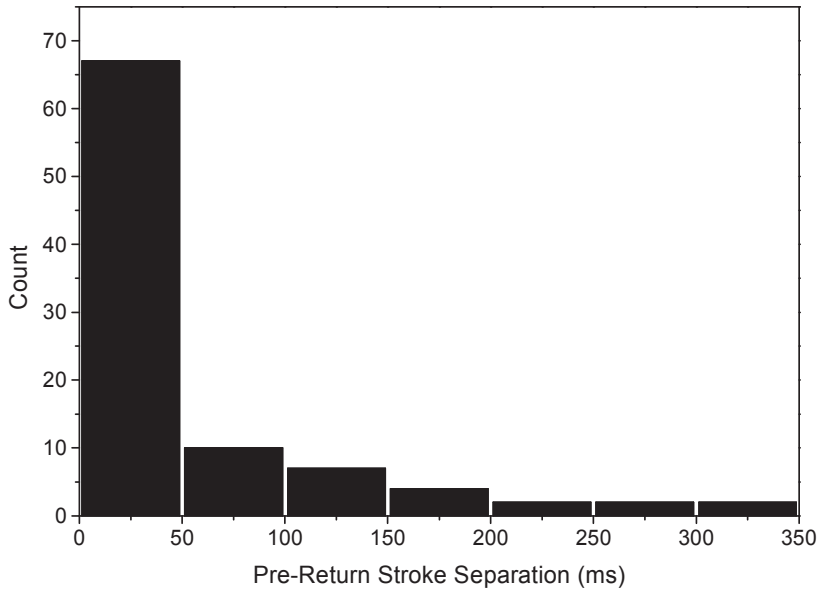


Fig. 4. The distribution of pre-return stroke separation

return stroke separation can be seen from Table 1.

TABLE I. SUMMARY OF STATISTICAL RESULTS OF PRE-RETURN STROKE SEPARATION

N	Min (ms)	Max (ms)	AM (ms)	GM (ms)	Median (ms)
94	5	328	57	32	27

However, the different locations, meteorological conditions, geographical region, and latitude, played a significant role, which indicated differences in characteristics of electric field change preceding negative first return stroke produced by preliminary breakdown processes.

IV. Conclusion

Characteristic of electric field change preceding negative CG lightning flashes produced by preliminary breakdown has been examined. The arithmetic mean and geometric mean of pre-return stroke separation are 57 ms and 32 ms, respectively. Pre-return stroke separation varies from 5 ms to 328 ms. Further studies should be conducted by using other statistical parameters such as pulse train duration, individual pulse duration, inter pulse duration, stepped leader, and time interval variation between return stroke to better understand the characteristics of electric field change preceding negative first return stroke produced by preliminary breakdown process.

Acknowledgment

The authors would like to thank to State Electricity Company (PT PLN) supported this work (Grand number, PLN Sumbar, No. 019/202.SPK/WSB/2012). The authors also thank to our laboratory members for making this work possible.

References

[1] Clarence, N.D., Malan, D.J., "Preliminary discharge processes in lightning flashes to ground," Quarterly Journal of the Royal Meteorological Society 83, 161–172, 1957.

[2] Krehbiel, P.R., Brook, M., McCrory, R., "An analysis of the charge structure of lightning discharges to the ground," Journal Geophysics Research 84, 2432–2456, 1979.

[3] Proctor, D.E., Uytenbogaardt, R., Meredith, B.M., "VHF radio pictures of lightning flashes to ground," Journal of Geophysical Research 93, 12,683–12,727, 1988.

[4] Kitagawa, N., Brook, M., "A comparison of intracloud and cloud-to-ground lightning discharges," Journal of Geophysical Research 65, 1189–1201, 1960.

[5] Baharudin, Z.A., M. Fernando, N. A. Ahmad, J.S. Makel, M. Rahman, and V. Cooray, "Electric field changes generated by the preliminary breakdown for the negative cloud-to-ground lightning flashes in Malaysia and Sweden," Journal of Atmospheric and Solar-Terrestrial Physics, Vol. 84–85, , pp. 15-24, August 2012.

[6] Lin, Y. T., M. A. Uman, J. A. Tiller, R. D. Brantley, W. H. Beasley, E. P. Krider, and C. D. Weidman, "Characterization of lightning return stroke electric and magnetic fields from simultaneous two station measurements," J. Geophys. Res., 84, 6307–6314, doi:10.1029 /JC084iC10p06307, 1979.

[7] Baharudin, Z.A., N. A. Ahmad, M. Fernando, V. Cooray, and J. S. Makel, "Comparative study on preliminary breakdown pulse trains observed in Johor, Malaysia and Florida, USA," Atmospheric Research, Vol. 117, pp. 111-121, 2012.

Comparative Study of Electric Generator Drive Engine Performance by Various Types of Fuel

Yandri and Seno D. Panjaitan
Department of Electrical Engineering
Tanjungpura University
Pontianak, Indonesia
e-mail: yandri.hasan@gmail.com, senopanjaitan@gmail.com

***Abstract*— This paper discusses the performance of gasoline generator set (genset) by using gasoline, LPG, and biogas as the fuel. Combustion engine of the gasoline genset was modified in order to be able to operate using LPG and biogas fuels. A 850 W, 220 V, 50 Hz, single phase generator drive engine has been used in the experiment using 100-W and 200-W bulbs as the electrical loads. Five variables were compared based on the types of fuel and variations of load: fuel consumption (FC), specific fuel consumption (SFC), input energy (thermal energy), output energy (electrical energy), and thermal efficiency. As the results, SFC of the biogas was lower than gasoline and LPG. For 100-W bulb as the load, SFC of the gasoline, LPG, and biogas were 4.5294 kg/kWh, 2 kg/kWh, and 0.68 kg/kWh respectively. While 200-W bulb, SFC of the gasoline, LPG, and biogas were 2.2883 kg/kWh, 1.0742 kg/kWh, and 0.6476 kg/kWh respectively. In addition, the thermal efficiency of biogas was higher than gasoline and LPG. For 100-W bulb as the load, thermal efficiency by using gasoline, LPG, and biogas were 1.6741 %, 3.5884 %, and 10.5710 % respectively. For 200-W bulb, the thermal efficiency by using gasoline, LPG, and biogas were 3.3137 %, 6.6811 %, and 11.0996 % respectively.**

Keywords—renewable energy; genset; biogas; fuel consumption; specific fuel consumption; thermal energy; thermal efficiency

I. INTRODUCTION

Gasoline, diesel, and Liquid Petroleum Gas (LPG) fuels are non-renewable energy so that its implementation in combustion engine to drive the electric generator needs to be gradually minimized. The use of biogas as an alternative fuel derived from municipal solid waste (organic waste) can be used as a solution.

Gasoline is a hydrocarbon fuel (mixture of hydrogen and carbon), refined from petroleum. Petroleum is a dark, thick liquid that is extracted from the earth by oil wells. LPG may be propane, butane, or a mixture of both. Properly designed fuel systems will allow the use of LPG with no appreciable loss of horsepower, as compared to a similar engine burning gasoline. LPG burns cleanly and leaves few combustion chamber deposits. In addition, LPG burns slower than gasoline because it has higher ignition temperatures [1].

Biogas originates from bacteria in the process of biodegradation of organic material under anaerobic conditions. It consists of a varying proportion of CH_4 (methane) and CO_2 (carbon dioxide) and traces of H2S, N, CO, O, etc. The content of CH4 and CO2 is a function of the matter digested and the process conditions like temperature, C/N ratio, etc. CH4 is the most valuable component under the aspect of using biogas as a fuel; the other components do not contribute to the calorific ("heating") value and are often "washed out" in purification plants in order to obtain a gas with almost 100 % CH4. The useful part of the energy of biogas is the calorific value of its CH4 content. The other components are the energy content also but they do not participate in a combustion process [2].

A gasoline-fueled engine is a mechanism designed to transform the chemical energy of burning fuel into mechanical energy. A gasoline engine is an internal combustion engine. Gasoline is combined with air and burned inside the engine.

Some previous researchers have studied the biogas and others as the fuel of engine to drive the electric generator. Reference [3] discusses the performance of biogas run petrol engine for a small scale power generation. The biogas was produced synthetically using a mixture of line supply of natural gas and regulated flow of CO2 from a cylinder. The engine performance was studied using proportional mixtures of the two gases having various percentages of CH4 contents by volume that simulated a range of biogas sources. The performance was only comparable with petrol when the CH4 content of biogas was at least about 60 % by volume. A study on power generation system using biogas generated from the waste of pig farm is discussed in [4]. Spark ignition (SI) gasoline engine was modified to use biogas fuel. An electronic speed regulation unit was also developed to keep the system at certain speed. Experimental investigations have been carried out to examine the performance characteristics of power generation system (such as system frequency and phase output voltage). Other paper with regard to the design and study on the mixture control system of the biogas-gasoline dual-fuel engine is discussed as in [5]. The characteristic experiments those consisted of two major parts (full-load speed characteristic experiment and load characteristic experiment) have been performed to test the function of electronically controlled system. While the component, physical & chemical characteristics of biogas and the problems of engines running on it are introduced in [6]. Two kinds of biogas engines: biogas-diesel dual fuel engine generators and SI biogas engine generators are developed. In addition, experiment study on

This research funded by Indonesian Ministry of Research and Technology (KEMENRISTEK) in INSINAS Research program (2012-2013).

miniature biogas generating system is discussed as in [7]. It involved the design of pressure stabilizer and mixer and the methods and results of testing. Besides, a quasi-dimensional SI engine cycle model is used to predict the combustion and exhaust emissions of a gas engine driven generator for the cases of using biogas and LPG-biogas blends and discussed as in [8], while field-measured results and dynamic stability analysis of a commercial 100-kW biogas generation system consisted of a gas engine and an induction generator is presented in [9]. Furthermore, migration from gasoline to gaseous fuels is presented in [10] while the potency of biogas from municipal solid waste to be converted into electricity is discussed in [11].

II. PERFORMANCE PARAMETERS OF GENSET

A. Fuel Consumption (FC)

FC is a measure of fuel consumed by an engine, usually measured in unit of volume of fuel consumed per unit of time. It can also be defined as the amount of fuel consumed by the engine to run the engine for a certain time.

FC is determined as follows,

$$FC = \frac{V_f}{t} \quad (1)$$

where V_f is volume of fuel used (liter) and t is operating time (hour).

B. Specific Fuel Consumption (SFC)

SFC is the engine performance parameter directly related to the economic value of the engine. This value is used to determine the amount of fuel required to produce the amount of power in a certain time interval. *SFC* allows all different sizes of engines to be compared which is the most fuel efficient. It allows manufacturers to see which engine will use the least fuel while still producing a high amount of power.

SFC is determined as follows,

$$SFC = \frac{m_f}{E} \quad (2)$$

where m_f is mass of the used fuel (kg) and $m_f = V_f \cdot \rho$, ρ is fuel density (kg/liter), and E is electrical energy (kWh).

C. Input Energy (Thermal Energy) of Fuel

The internal combustion engine is a heat engine that converts chemical energy in a fuel into mechanical energy, usually made available on rotating output shaft. Chemical energy of the fuel is first converted to thermal energy by means of combustion or oxidation with air inside the engine. This thermal energy raises the temperature and pressure of the gases within the engine and the high-pressure gas then expands against the mechanical mechanism of the engine. This expansion is converted by the mechanical linkages of the engine to a rotating crankshaft, which is the output of the engine. The crankshaft, in turn, is connected to a transmission and/or power train to transmit the rotating mechanical energy to the desired final use. For engines this will often be the propulsion of a vehicle (i.e., automobile, truck, locomotive, marine vessel, or airplane). Other applications include stationary engines to drive generators or pumps, and portable engines for things like chain saws and lawn mowers [12].

Input energy (thermal energy) of fuel can be determined as follows,

$$Q_f = m_f \, . \, LHV \quad (3)$$

where Q_f is the amount of fuel heat (kcal) and LHV is lower heating value (kcal/kg).

LHV is a property of a fuel and defined as the amount of heat released by combusting a specified quantity (initially at 25 ºC or another reference state) and returning the temperature of the combustion products to 150 ºC. The *LHV* assumes that the latent heat of vaporization of water in the fuel and the reaction products is not recovered. It is useful in comparing fuels where condensation of the combustion products is impractical, or heat at a temperature below 150 ºC cannot be put to use.

D. Output Energy (Electrical Energy)

Electrical energy is energy stored in charged particles within an electric field. In other words, electrical energy results from the movement of an electrical charge.

Electrical energy (*E*) can be determined as follows,

$$E = P \, . \, t \quad (4)$$

where P is electrical power of load (kW) and t is operating time (hour).

E. Thermal Efficiency

Thermal efficiency shows how much power generated by a number of the rate of heat generated by the combustion of fuel. Heat release rate itself is the result of the fuel flow rate and the value of fuel combustion. Thermal efficiency of the engine output power can be increased by increasing the fuel flow rate or using fuels with high combustion value.

Thermal efficiency (η_{th}) of the genset expresses the effectiveness value of fuel energy supplied to the combustion chamber to produce work and is determined in (5).

$$\eta_{th} = \frac{859.845 \, E}{Q_f} \times 100\,\% \quad (5)$$

III. EXPERIMENTAL METHOD

In the experiment, some equipment and materials had been used: (a) A 850 W, 220 V, 50 Hz, single phase gasoline-fueled genset, (b) some types of fuel, such as gasoline, LPG, and biogas, (c) conversion kit and mixer. Conversion kit serves to lower the gas pressure of the tube in accordance with the engine operating pressure, while the mixer serves to mix gas and air, (d) gas tube and accessories, used especially for the placement of LPG or biogas fuel which then through a hose connected to the inlet of the genset, (e) compressor, used for distributing biogas from temporary biogas storage appliance to the gas tube, (f) digital scales, used to measure gas

consumption to operate the genset using either LPG or biogas in ten minutes, (g) hose, which serves to drain the gas from the tube to the engine, (h) tachometer, which serves to measure engine speed, (i) 100-W and 200-W bulbs (incandescent lamps) those used as the generator load, (j) power quality analyzer (PQA) which serves to measure the electrical quantities.

The used genset for the experiments was from the similar vendor which was originally a gasoline generator. Experiment using gasoline did not need any modification in combustion system. In order to operate genset using LPG and biogas, a modification at the combustion system had been done at the original gasoline genset. The experiment using LPG and biogas used the same generator using the gas combustion system.

The genset tests were performed on some variations of load in accordance with the amount of fuel used and the duration of the operation, either gasoline, LPG, or biogas. Active power of the various loads was also recorded based on PQA readings. Besides, other quantities were also calculated based on the data obtained in the experiment. They involved *FC*, *SFC*, input energy of fuel (thermal energy), output energy (electrical energy) and thermal efficiency. Furthermore, the value of each quantity (variable) for various types of fuel was compared and analyzed.

IV. RESULTS AND DISCUSSION

A. Test Results of Genset

Table I, II, and III show the test results of genset using gasoline, LPG, and biogas as fuel. Active power in each table was measured by using PQA.

TABLE I. TEST RESULT OF GENSET USING GASOLINE

Bulb specification (W)	Active power (W)	Fuel used (ml)	Duration of genset operation (minute : second)
100	87	200	23:26
200	176	200	22:56

TABLE II. TEST RESULT OF GENSET USING LPG

Bulb specification (W)	Active power (W)	Fuel used (gram/minute)
100	91	3
200	183	3.33

TABLE III. TEST RESULT OF GENSET USING BIOGAS

Bulb specification (W)	Active power (W)	Fuel used (gram/minute)
100	91	1.5
200	127	2

B. Calculation of genset performance parameters

1) Genset operation using gasoline as fuel

a) 100 W bulb as the load

Density of gasoline is 770 kg/m^3 = 0.77 kg/liter, therefore :

$$m_f = V_f \,.\, \rho$$
$$= 0.2\ liter\ .\ 0.77\ kg/liter$$
$$= 0.1540\ kg$$

$$E = P\,.\,t$$
$$= 0.087\ kW\ .\ 0.3906\ hour$$
$$= 0.0340\ kWh$$

$$FC = \frac{V_f}{t}$$
$$= \frac{0.2\ liter}{0.3906\ hour}$$
$$= 0.5120\ liter/hour$$

$$SFC = \frac{m_f}{E}$$
$$= \frac{0.1540\ kg}{0.0340\ kWh}$$
$$= 4.5294\ kg/kWh$$

LHV of the gasoline is 47.4 MJ/kg = 47,400 kJ/kg. Because 1 kcal = 4.18 kJ, so 47,400 kJ = 11339.7129 kcal and 47,400 kJ/kg = 11339.7129 kcal/kg. Therefore,

$$Q_f = m_f\,.\,LHV$$
$$= 0.1540\ kg\ .\ 11339.7129\ kcal/kg$$
$$= 1746.3158\ kcal$$

And 1 kWh = 859.845 kcal, so :

$$\eta_{th} = \frac{859.845\,.\,E}{Q_f} \times 100\,\%$$
$$= \frac{(859.845\,.\,0.0340)\ kcal}{1746.3158\ kcal} \times 100\,\%$$
$$= 1.6741\,\%$$

b) 200 W bulb as the load

By doing the same steps as in (*a*), it is obtained m_f = 0.1540 *kg*, *E* = 0.0673 *kWh*, *FC* = 0.5233 *liter/hour*, *SFC* = 2.2883 *kg/kWh*, Q_f = 1746.3158 *kcal*, and η_{th} = 3.3137 %.

2) Genset operation using LPG as fuel

a) 100 W bulb as the load

Density of LPG (by assuming 70 % propane and 30 % butane) is 522.2 kg/m^3 = 0.5222 kg/liter, therefore :

$$V_f = \frac{m_f}{\rho} = \frac{0.003\ kg}{0.5222\ kg/liter} = 0.0057\ liter$$

$$E = P\,.\,t = 0.091\ kW\ .\ 0.0167\ hour = 0.0015\ kWh$$

$$FC = \frac{V_f}{t} = \frac{0.0057\ liter}{0.0167\ hour} = 0.3413\ liter/hour$$

$$SFC = \frac{m_f}{E} = \frac{0.003\ kg}{0.0015\ kWh} = 2\ kg/kWh$$

LHV of LPG is 50.08 MJ/kg = 50,080 kJ/kg. Because 1 kcal = 4.18 kJ, so 50,080 kJ = 11980.8612 kcal and 50,080 kJ/kg = 11980.8612 kcal/kg. Therefore,

$$Q_f = m_f\ .\ LHV = 0.003\ kg\ .\ 11980.8612\ kcal/kg = 35.9426\ kcal$$

And 1 kWh = 859.845 kcal, so :

$$\eta_{th} = \frac{859.845\ .\ E}{Q_f} \times 100\ \% = \frac{(859.845\ .\ 0.0015)\ kcal}{35.9426\ kcal} \times 100\ \% = 3.5884\ \%$$

b) 200 W bulb as the load

By doing the same steps as in (*a*), it is obtained V_f = 0.0064 *liter*, *E* = 0.0031 *kWh*, *FC* = 0.3832 *liter/hour*, *SFC* = 1.0742 *kg/kWh*, Q_f = 39.8963 *kcal*, and η_{th} = 6.6811 %.

3) Genset operation using biogas as fuel

a) 100 W bulb as the load

CH_4 content in the biogas after purification was 68 %. The density of biogas at standard conditions (i.e. at temperature of 273 ºK and ambient pressure of 1013 mbar) is 0.72 kg/m^3 = 0.00072 kg/liter. Therefore,

$$m_f = 0.68\ .\ 0.0015\ kg = 0.00102\ kg$$

$$V_f = \frac{m_f}{\rho} = \frac{0.00102\ kg}{0.00072\ kg/liter} = 1.4167\ liter$$

$$E = P\,.\,t = 0.091\ kW\ .\ 0.0167\ hour = 0.0015\ kWh$$

$$FC = \frac{V_f}{t} = \frac{1.4167\ liter}{0.0167\ hour} = 84.8323\ liter/hour$$

$$SFC = \frac{m_f}{E} = \frac{0.00102\ kg}{0.0015\ kWh} = 0.68\ kg/kWh$$

LHV of biogas at standard conditions is 50 MJ/kg = 50,000 kJ/kg. Because 1 kcal = 4.18 kJ, so 50,000 kJ = 11961.7225 kcal and 50,000 kJ/kg = 11961.7225 kcal/kg. Therefore,

$$Q_f = m_f\ .\ LHV = 0.00102\ kg\ .\ 11961.7225\ kcal/kg = 12.2010\ kcal$$

and 1 kWh = 859.845 kcal, therefore :

$$\eta_{th} = \frac{859.845\ .\ E}{Q_f} \times 100\ \% = \frac{(859.845\ .\ 0.0015)\ kcal}{12.2010\ kcal} \times 100\ \% = 10.5710\ \%$$

b) 200 W bulb as the load

By doing the same steps as in (*a*), it is obtained $m_f = 0.00136$ *kg*, $V_f = 1.8889$ *liter*, $E = 0.0021$ *kWh*, $FC = 113.1078$ *liter/hour*, $SFC = 0.6476$ *kg/kWh*, $Q_f = 16.2679$ *kcal*, and $\eta_{th} = 11.0996$ %.

C. Discussion

Table IV-VI show performance parameters of genset using gasoline, LPG, and biogas as fuel based on the calculation. The calculation was based on the determined loads. Characteristics of *FC*, *E*, *SFC*, Q_f , and η_{th} vs load are described based on table IV - VI and depicted in Fig. 1 - 5.

From Table IV – VI and Fig. 3 shows that SFC will decrease with increasing load. SFC is the biggest on gasoline, followed by LPG fuel, and the smallest is powered by biogas. With increasing load (from 100 W to 200 W) the decrease in SFC value for gasoline, LPG, and biogas respectively 49.48 %, 46.29 %, and 4.76 %. Unlike the case with FC, where its value is getting bigger with increasing load (Fig. 1).

From Fig. 2, gasoline takes the uppermost place in producing electrical energy (output energy), followed by LPG and biogas. With increasing load (from 100 W to 200 W), the increase in electrical energy for the gasoline, LPG, and biogas respectively 97.94 %, 106.67 %, and 40 %.

From Fig. 4, gasoline also takes the uppermost place in producing thermal energy (input energy), followed by LPG and biogas. With increasing load (from 100 W to 200 W), the increase in the input energy only occurs on LPG and biogas which are respectively 11 % and 33.33 %.

From Fig. 5, biogas is the highest in thermal efficiency, followed by LPG and gasoline. With increasing load (from 100 W to 200 W), the increase in thermal efficiency for the gasoline, LPG, and biogas respectively 97.94 %, 86.19 %, and 5 %.

Because there are only two samples of electrical load tested in this experiment (100-W and 200-W bulbs), therefore a linear approach is used for graphic presentation and can be seen from the line connecting the two points (Fig. 1-5). The accurate curves will be obtained for more samples of electrical load.

TABLE IV. PERFORMANCE PARAMETERS OF GENSET USING GASOLINE

Bulb specification (W)	*FC* (liter/hour)	*E* (kWh)	*SFC* (kg/kWh)	Q_f (kcal)	η_{th} (%)
100	0.5120	0.0340	4.5294	1746.32	1.6741
200	0.5233	0.0673	2.2883	1746.32	3.3137

TABLE V. PERFORMANCE PARAMETERS OF GENSET USING LPG

Bulb specification (W)	*FC* (liter/hour)	*E* (kWh)	*SFC* (kg/kWh)	Q_f (kcal)	η_{th} (%)
100	0.3413	0.0015	2.0000	35.9426	3.5884
200	0.3832	0.0031	1.0742	39.8963	6.6811

TABLE VI. PERFORMANCE PARAMETERS OF GENSET USING BIOGAS

Bulb specification (W)	*FC* (liter/hour)	*E* (kWh)	*SFC* (kg/kWh)	Q_f (kcal)	η_{th} (%)
100	84.8323	0.0015	0.6800	12.2010	10.5710
200	113.1078	0.0021	0.6476	16.2679	11.0996

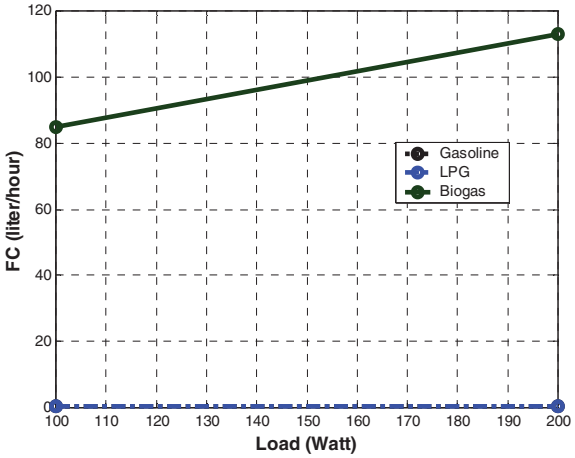


Fig. 1. FC – load characteristic

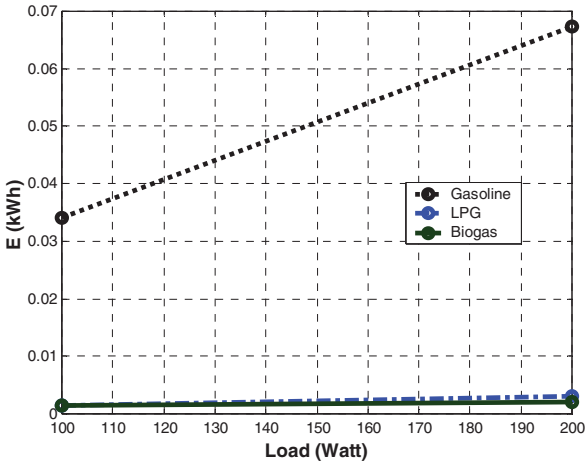


Fig. 2. Electrical energy – load characteristic

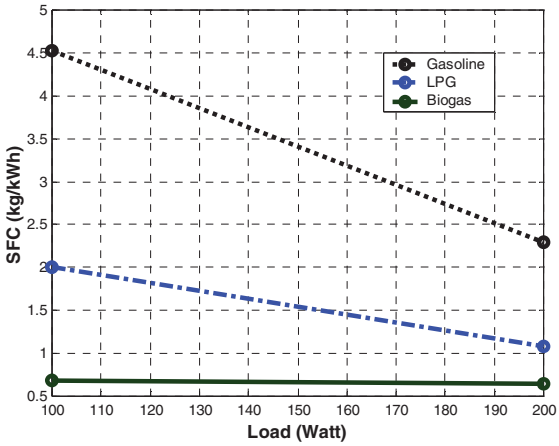


Fig. 3. SFC – load characteristic

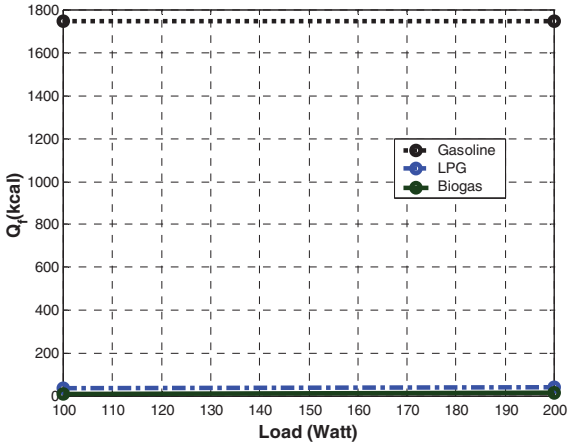


Fig. 4. Thermal energy – load characteristic

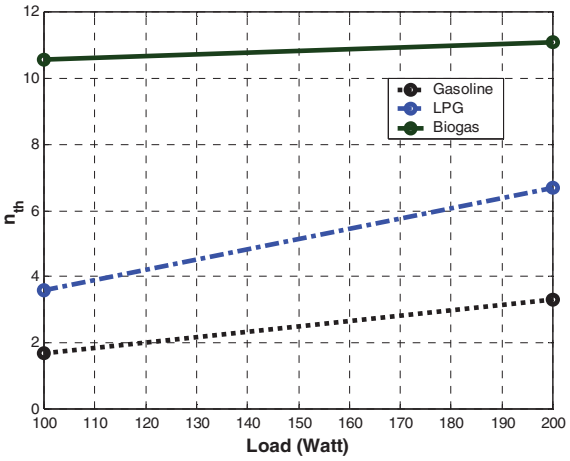


Fig. 5. Thermal efficiency – load characteristic

V. CONCLUSIONS

Biogas has several advantages over gasoline and LPG. This refers to the value of biogas SFC which is lower than LPG and gasoline. In addition, thermal efficiency of biogas is also greater than LPG and gasoline. Therefore abundant source of biogas from organic waste can be used as a solution to cope with the electricity originated from municipal solid waste. Besides, using waste as the source of electrical generation system by converting them into biogas will bring two benefits: renewable energy source and environmental friendly technology.

ACKNOWLEDGMENT

The authors would like to thank head of research center of Tanjungpura University who facilitated the administration for the research operational and students in the control systems laboratory who helped in the preparation for data measurement in the research.

REFERENCES

[1] A.C. Roth and R.J. Baird, Small Gas Engine, South Holland, Illinois: The Goodheart-Willcox Company, Inc., 1992.

[2] K.V. Mitzlaff, Energies for Biogas, Deutsche: GTZ, 1988.

[3] Md. Ehsan and N. Naznin, "Performance of Biogas Run Petrol Engine for Small Scale Power Generation," Journal of Energy & Environment 4, 2005, pp. 1-9.

[4] Tran Minh Tien, Pham Xuan Mai, Nguyen Dinh Hung, and Huynh Thanh Cong,"A Study on Power Generation System Using Biogas Generated from the Waste of Pig Farm," International Forum on Strategic Technology (IFOST), 2010, pp. 203-207.

[5] Su Chu-wang, Ye Shu-jian, Wei Chao-yi, and Xie Mei-zhi, "The Design and Study on the Mixture Control System of the Biogas-Gasoline Dual-fuel Engine," IEEE International Conference on Materials for Renewable Energy & Environment (ICMREE), Volume 1, 2011, pp. 517-520.

[6] Jiang Yao-hua, Xiong Shu-seng, Shi Wei, He Wen-hua, Zhang Tian, Lin Xian-ke, Gu Yun, Lu Yin-ding, Qian Xiao-jun, Ye Zong-yin, Wang Chong-ming, and Wang Bei, "Research of Biogas as Fuel for Internal Combustion Engine," Asia-Pacific Power and Energy Engineering Conference (APPEEC), 2009, pp. 1-4.

[7] Shi Ling, Rang Guowei, Zhang Tianhui, and Zhang Rukun, "Experimental Study on Miniature Biogas Generating System," Energy Procedia 11, 2011, Elsevier, pp. 2615-2623.

[8] Jiasheng Guo, Chaokui Qin, Wanneng Dai, and Jixu Tang, "Numerical Investigation on the Effects of LPG-Biogas Blends on Engine Combustion and Emission," International Conference on the Environmental Science and Information Application Technology (ESIAT), Volume 3, 2010, pp. 179-182.

[9] Li Wang and Ping-Yi Lin, "Analysis of a Commercial Biogas Generation System Using a Gas Engine-Induction Generator Set," IEEE Transactions on Energy Conversion, Volume 24, 2009, pp. 230-239.

[10] Seno D. Panjaitan, Yandri, Sukandar, Berlian Sitorus, "Migration from Gasoline to Gaseous Fuel for Small-scale Electricity Generation Systems", Journal Telkomnika, Vol. 11, No. 1, 2013, pp. 1-8.

[11] Seno D. Panjaitan, Sukandar, Berlian Sitorus, Yandri, "Potency of Biogas from Municipal Solid Waste for Electricity Generation: Case Study of A Small Scale Landfill in Pontianak", Proceeding of the 7th Asian Pacific Landfill Symposium – Sustainable Solid Waste Management for a Better Life, 2012.

[12] W.W. Pulkrabek, Engineering Fundamentals of the Internal Combustion Engine, New Jersey: Prentice Hall, 1997.

Contingency Analysis on 500kV Jawa-Bali Transmission Line System Based on Power Load Performance Index

Lesnanto Multa Putranto, Julian Widya Perdana, M. Isnaeni
Department of Electrical Engineering and Information Technology UGM
Yogyakarta, Indonesia
lesnanto@ugm.ac.id, julian_te09@mail.ugm.ac.id, isnaeni@te.ugm.ac.id

***Abstract*—The line outage of a power system may cause overloads on the lines, overvoltage on buses, and undervoltage on buses that can threat the power system security itself. The effect of the contingency of each power system element is vary. Power Load Performance Indexes as one of the contingency indexes can be implemented to rank the contingency level of each power system lines. N-1 contingency line is applied at 500 kV Jawa Bali Transmission System to rank the severe of contingency and to evaluate the effect of contingency of each element. Tanjung Jati-Ungaran line has the biggest contingency index which means it would be the worst scenario of the line outage.**

Keywords: Power Load Performance Indexes, Power System Security, Contingency

I. INTRODUCTION

Electric power system is a dynamic system. Voltage, power flow, and current magnitude can fluctuate according to the condition of the components. Those components are generators, transmission lines, and interbus transformer as loads. Changing conditions occurred on those components will result in a change the parameters of voltage, power flow, and current magnitude in the other components. [1]

The disturbances occurred in the generation components may be influenced to the generator itself so the power from generator can not supplied to the power system. Because of these conditions, the lack of power supply from this generator should be handled by generating units of other bus. [1]

Disturbances that may occur in the transmission components could be a line outage from the system so that the power should be carried to the load through other lines.

Disturbance in the system can lead to changes in the flow of power on the power system. The different changes that occur - depending on the location of the point of interruption. [2] When the power flow changes, there would be also change in the current of the line and change the voltage on the bus. Bus component and line resistance is rated for a certain voltage and current magnitude. Therefore, analysis is required to test the bus and the most critical line interference.

Contingency analysis is an attempt to determine the operating state of the power system as a component of impaired or even out of the system. With the contingency analysis, power system operation is expected to continue to run safely despite an interruption in the system. [3] The end result of this analysis in the form of contingency indices indicating the order of transmission line of the most dangerous to the security of power system operation in the event of disruption the line location.

II. POWER FLOW PERFORMANCE INDEX

Contingency index of a bus or line is a variable that indicates whether safe or not to operate the power system after contingency occurred. Contingency index is to indicate the transmission lines that can lead to the most critical condition of systems when they are out from the system.

To determine the contingency index, there are some methods, such as Power Flow Performance Index [4]. Using contingency index, the worst scenario of line outage can be obtained which would cause the most dangerous effect of continuous operation of electric power systems thereafter.

$$PI_{(LF,j)} = \sum_{k=1}^{nl} W_k \left(\frac{|I_{ij-k}|}{I_{ij-k}^{limit}} \right)^{2m} \quad (1)$$

W_k = weighting factor of line k
I_{i-k} = current flow from bus i to bus k (p.u.)
nl = number of line

Contingency index calculation method of comparison with the current magnitude of the comparison is done by summing the currents are flowing in the event of a contingency k lines with the line current-carrying capability.

$$PI_{(v,j)} = \sum_{k=1}^{nb} W_k \left(\frac{V_k - \overleftarrow{V_k}}{\Delta V_k} \right)^{2m} \quad (2)$$

W_k = weighting factor of bus k
V_k = voltage at bus k (p.u.)
nb = number of bus

Contingency index calculation with voltage comparison method comparison performed by summing the difference in bus voltage magnitude is going on contingency (Vk) against the mean of the bus at the time of contingency (Vk) of the

difference between the maximum and minimum voltage on the bus (ΔVk). [4]

$$Index = \sum W_{pi}\left(\frac{S_{ipc}}{S_{iLimit}}\right)^{2m} \quad (3)$$

W_{pi} = weighting factor
S_{ipc} = power flow on post-contingency (MVA)
S_{iLimit} = power flow limit (MVA)

The Power flow Performance Index of line k contingency sum the comparison between the apparent power flow of each line post contingency and the power flow limit of each line [1]. The weighting factor indicates the importance level of each line rank by the power system utility. The exponential variable gives the significant degree among the lines.

III. SYSTEM SETUP

The object of this work is 500 kV Jawa- Bali transmission system. The load operating condition is used the peak load profile at 19.00 on 21 March 2013.

The data of transmission subsystem 500kV Jawa-Bali, fig 1, is taken from the Transmission and Load Control Center (P3B). The transmission system has 25 buses, 51 generating units, and 51 lines are connected form a single circuit and double circuit. In this research, a system load interbus loading on transformers (IBT), which is connected to high voltage substation (GITET).

Generating units are connected to the 500kV subsystem is Suralaya, Muaratawar, Cirata, Saguling, Tanjung Jati, Gresik, Grati, and Paiton with power generation of 8168MVar and 13452MW.

Load connected to the 500kV divided on 25 stations with loading of 13072MW and 4217MVar. Those stations are Suralaya, Cilegon, New Balaraja, Kembangan, Gandul, Bekasi, Cibinong, Cawang, Depok, Cibatu, Muaratawar, Mandirancan, Saguling, Cirata, Tanjung Jati, Bandung Selatan, Tasikmalaya, Unggaran, Pedan, Gresik, Surabaya Barat, Ngimbang, Kediri, Grati, and Paiton.

Jawa-Bali 500kV subsystem uses 2 types of conductors, namely Dove and Gannet. Dove conductor type has a capacity of 1718 MVA power Transmission. Conductor type Gannet has a capacity of 2078 MVA power Transmission. 500kV transmission subsystem Jawa-Bali consists of 51 lines, 42 lines of a double circuit line and 9 lines of a single circuit.

Fig 1. Transmission line 500kV Jawa-Bali

IV. METHODOLOGY

In this work, the contingency analysis is done by following steps described in Fig 2. The power system model and simulation are built using MATLAB software.

The previous contingency analysis's scenario determined manually by a variety of considerations so the simulated scenario maybe not the worst scenario. Contingency index is required to rank the effect of an outage of the element of power system operating conditions on the post-contingency condition.

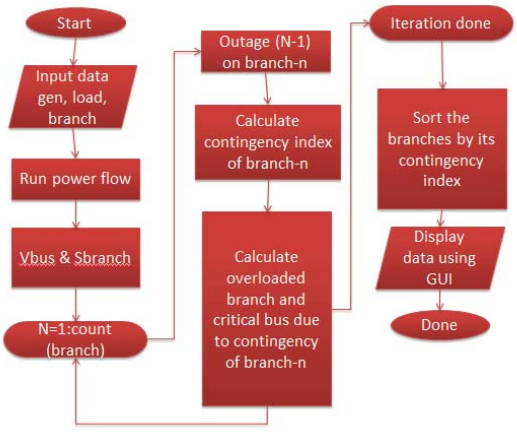


Fig 2. Flowchart of contingency index calculation

In this work, the contingency analysis is done by performing load flow on the normal and post contingency condition. The outage is N-1 contingency scenario. The transient condition when fault occurred is neglected. All line contingency is simulated so that the rank of contingency can be done. Power flow calculation is used to predict the variable value at post contingency index.

Powerflow Performance Index is used to evaluate the contingency effect of each line outage in the power system. Determination of contingency index is calculated from the flow of power flowing on each line (MVA) and Transmission capacity (MVA) for every contingency conditions [1]

In the calculation, all the lines are considered equally important to factor in determining the priority line made contingency index 1 for all lines. Exponential variables so that the result is multiplied by 2 contingency index is positive. [1] The variables exponentially large contingency index serves to demonstrate any significant line with a value between lines can be causing critical systems and lines that have little effect on the system. In this study, exponential variables (m) is set at 15. [1]

In this work, interface system is proposed to ease observing contingency effect on the whole power system. The interface system can help power system operators (PT PLN Persero Indonesia) to observe the contingency condition of each line outage in the power system. There are 52 options of line outage that the effects of each line outage are shown in the user interface.

.

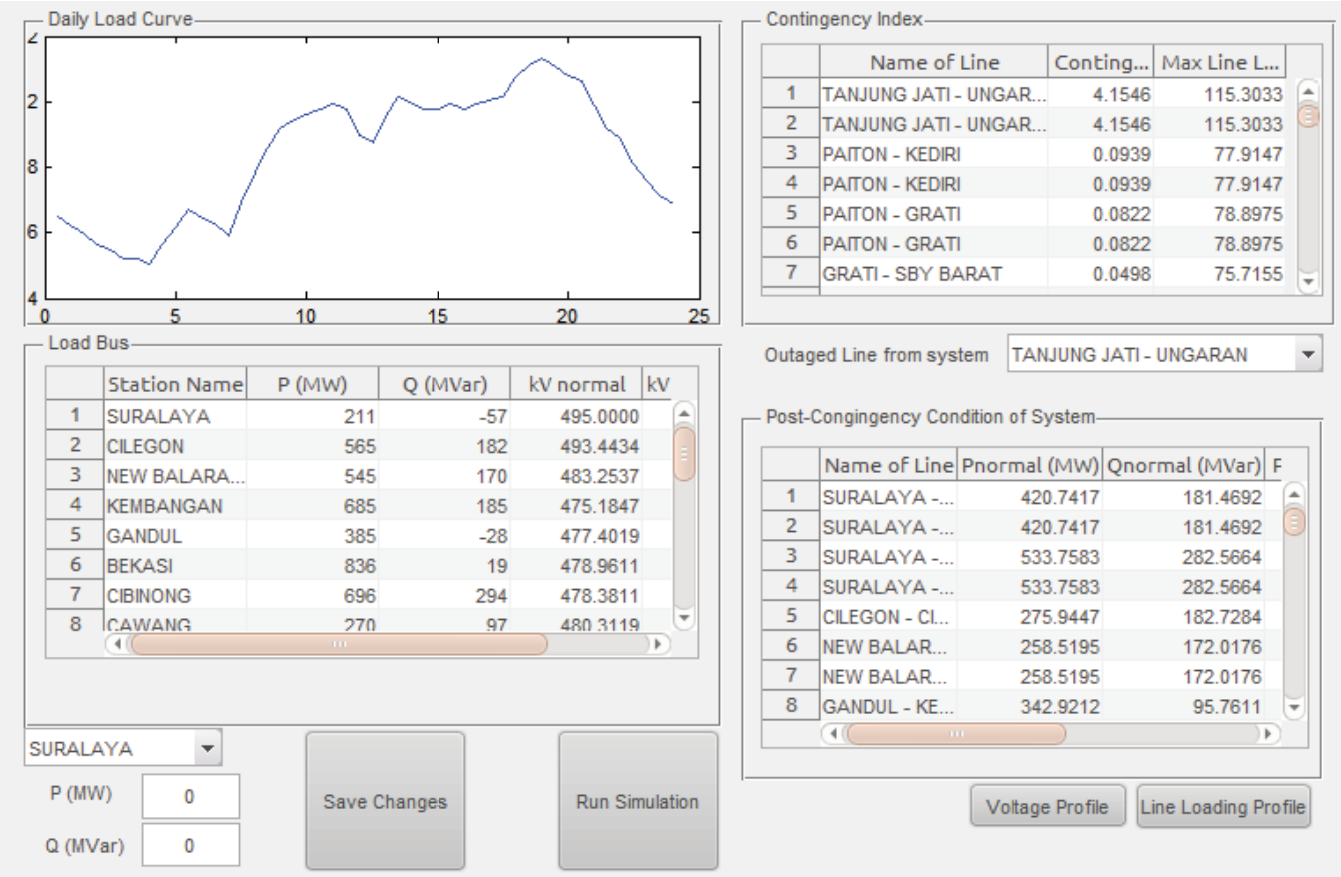


Fig 3. Interface for contingency monitoring

User Interface system at fig 3 combined shows the result of the contingency index calculation. User Interface system has the menu to choose contingency line and shows the actual line loading condition before the contingency occurred, predictable line loading condition and contingency index.

In the previous work [1], the limit of violation has not been reflected in the contingency index value. In this work, the limit of violation will be determined based on the desired power system

According to Fig.3, it can be seen that the user interface of monitoring system can show contingency index of each line outage and the maximum line loading of each contingency. Furthermore, the voltage and line powerflow in both pre and post contingency condition of each line can be shown.

The user interface can also compare the voltage profile of each bus and line loading profile of each line.

According to Fig.4 and Fig.5, there are differences in voltage and line loading in system between pre and post contingency condition of one line outage.

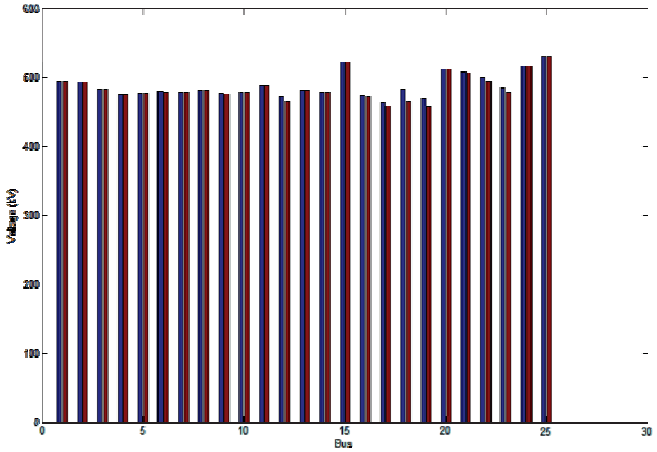


Fig 4. Interface of voltage profile

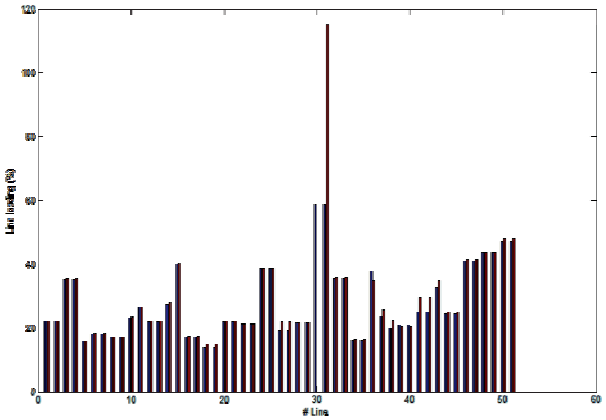


Fig 5. Interface of line loading profile

V. Test Result And Analysis

A. Overview of Contingency Index Calculation

From the contingency index, transmission lines can be ranked. From fig 4, Tanjungjati – Ungaran line has the highest contingency index with the value 71.65. This line is double circuit. The contingency of this line makes a line in the system loaded 115.3 % so there is a violation in the operation.

The contingency index of other line is below 1 and there is no violation from any line due to loading condition.

TABLE I. Contingency index of transmission line 500kV Jawa-Bali

Branch	Contingency Index	Maximum branch loading (%)	Overloaded branch due to contingency
Tanjung Jati - Ungaran	71.6552000	115.30	Tanjung Jati – Ungaran
Paiton - Grati	0.0008163	78.90	-
Paiton - Kediri	0.0005610	77.91	-
Grati - Sby Barat	0.0002372	75.72	-
Cawang - Muaratawar	0.0000031	65.65	-
Mandirancan - Ungaran	0.0000025	64.82	-
Cibinong - Bekasi	0.0000023	65.01	-
Bandung Selatan - Mandirancan	0.0000019	64.62	-
Sby Barat - Ngimbang	0.0000007	61.01	-
Ungaran - Sby Barat	0.0000004	59.99	-

B. Overview of Tanjung Jati –Ungaran Outage

Tanjung Jati – Ungaran line is a double circuit line. Outage of one circuit of this line is ranked on the second order. The changes of loading condition of the other line are shown in Table 2.

TABLE II. Change of transmission line's power loading due to post-contingency of Tanjung Jati - Ungaran line

Branch	Maximum branch loading (%) in normal condition	Maximum branch loading (%) in post-contingency
Tanjung Jati - Ungaran	58.73	115.30
Gresik - Sby Barat	24.89	29.67
Gresik - Sby Barat	24.89	29.67
Saguling - Bandung Selatan	37.73	34.89
Saguling - Bandung Selatan	19.10	21.88
Sby Barat - Ngimbang	19.10	21.88
Paiton - Kediri	19.88	22.39
Paiton - Kediri	32.62	34.80
Ungaran - Ngimbang	23.60	25.73

The outage of a line Tanjung Jati - Unggaran results the other line of Tanjung Jati - Unggaran experiencing overload with power load 115.3% above its MVA rating. Once analyzed, the effect of this overload caused by several things:

a. The power delivered by the two lines is only delivered by a line only.
b. Double circuit line Tanjung Jati - Unggaran radial form, as shown in Fig 2. Tanjung Jati station only supplies power to Unggaran station. Tanjung Jati station is generating units supplying power for 3237MVA or 16% of total generation to the system.

TABLE III. Change of Station voltage magnitude due to Post-Contingency of Tanjung Jati - Ungaran line

High Voltage Substation	V_{normal}	$V_{post\text{-}contingency}$
Pedan	469.39	457.85
Tasikmalaya	464.23	458.17
Ungaran	482.51	464.75
Mandirancan	472.31	464.84
Bandung Selatan	474.43	472.58
Kembangan	475.18	474.54
Depok	477.19	476.51
Gandul	477.40	476.76

Tanjung Jati – Ungaran contingency cause voltage decrement at Unggaran station 17.76 kV. Power flow changes to Unggaran is decreased the voltage at Pedan, Tasikmalaya, and Madirancan station.

C. Overview of Paiton-Grati Outage

Paiton-Grati line is also a double circuit line. Outage of one circuit of this line is ranked on the second order. The changes of loading condition of the other line are shown in Table 4.

TABLE IV. Change of transmission line's power loading due to Post-Contingency of Paiton-Grati line

Branch	Maximum branch loading (%) in normal condition	Maximum branch loading (%) in post-contingency
Paiton - Grati	43.72	78.90
Paiton - Kediri	46.98	50.23
Paiton - Kediri	46.98	50.23
Kediri - Pedan	24.58	27.42
Kediri - Pedan	24.58	27.42
Tasikmalaya - Depok	16.20	16.48
Tasikmalaya - Depok	16.20	16.48
Pedan - Tasikmalaya	21.00	21.26
Pedan - Tasikmalaya	21.00	21.26

If the contingency occurred in single circuit of Paiton-Grati line, the loading condition of the other lines is increasing. It can be shown in Table 4. The highest increment is on the other line of Paiton Grati (double circuit) with the loading value 78.90 %.

The contingency of one line in Paiton-Grati also gives effect on the station voltage. The five most influenced stations are Kediri, Pedan, Tasikmalaya, Ngimbang, Surabaya Barat, and Depok. The station voltages of Kediri, Pedan,

Tasikmalaya, and Depok are decreased due to the powerflow increasing to those buses. On the other hand, the station voltages of Surabaya Barat and Ngimbang are increased.

TABLE V. CHANGE OF STATION VOLTAGE MAGNITUDE DUE TO POST-CONTINGENCY OF PAITON-GRATI LINE

High Voltage Substation	V_{normal}	$V_{post-contingency}$
Kediri	484.4182	482.6147
Pedan	469.3938	468.4043
Tasikmalaya	464.2322	463.5413
Ngimbang	499.8889	500.1514
Sby Barat	507.8145	507.9229
Depok	477.1927	477.1063

D. Limit of Contingency Index

The m variable in Power Flow Performance Index formulation shows the significance degree of the contingency index as shown in Table 6. The bigger value of m shows more different between contingency indexes of each line outage.

The application of Power Flow Performance Index in the previous work [1] is to range between the line contingency effects, but the limit of violation is not presented yet.

From Table 6, it can be seen that when the contingency index is below 1, there is no violation in line loading. From this phenomenon, the limit of contingency index for Subsystem 500kV Jawa-Bali is 1.

TABLE VI. CONTINGENCY INDEX OF VARYING EXPONENSIAL VARIABLE

Name of Line	Value of variable m					Maximum line loading (%)
	2	3	5	10	15	
Tanjung Jati - Ungaran	1.995708	2.3874361	4.15463	17.25169	71.65519	115.30
Paiton - Grati	0.900247	0.345729	0.082239	0.008775	0.000816	78.90
Paiton - Kediri	0.744193	0.258894	0.093899	0.006881	0.000561	77.91
Grati - Sby Barat	0.678347	0.192994	0.049839	0.003884	0.000237	75.72
Cawang - Muaratawar	0.674016	0.151481	0.005179	0.000266	0.000003	65.65

According to Table 6, it can be seen that the increase of variable (m) may cause higher index value of worse scenario and lower index value of less dangerous scenario.

CONCLUSION

From this work there are few points related to the contingency N-1:

a. Globally Jawa - Bali 500 kV Transmission Line System is strong enough to face line contingency except the contingency at line Tanjung Jati - Ungaran
b. Power flow Performance Index can effectively rank the line contingency by considering the apparent power flow in each line.
c. To increase the system security, the installation of line connecting Tanjung Jati station to the other station is needed.
d. The limit of contingency index for Subsystem 500kV Jawa-Bali is 1.

ACKNOWLEDGEMENT

Authors thank the Power System Laboratory of Electrical Engineering and Information Technology Department for supporting the funding, equipment and tool for this research. Authors also appreciate Department of Analysis and Evaluation, PT. PLN P3B Indonesia for supporting the data for this research.

REFERENCES

[1] Mohamed, Salah E.G. (2012). "Power System Contingency Analysis to detect Network Weaknesses". *Zaytoonah University International Engineering Conference on Design and Innovation in Infrastructure 2012 (ZEC Infrastructure 2012)*
[2] Stevenson Jr, William D. 1983. *Analisis Sistem Tenaga Listrik edisi keempat*. Jakarta: Erlangga.
[3] Susilo, Yusuf (2005). *Analisis Kontingensi Sistem Transmisi 500kV Jawa Bali menggunakan EDSA*. *Skripsi S1,* Jurusan Teknik Elektro dan Teknologi Informasi FT UGM.
[4] Wood, Alen J. 1995. *Power Generation, Operation, and Control*. New York: JOHN WILEY & SONS, INC.
[5] Zimmerman, R.D., 2011. Matpower: Steady-State Operations, Planning and Analysis Tools for Power Systems Research and Education. Power Systems, IEEE Transactions, 26, pp.12-19.
[6] PLN P3B. 2011. *Rencana Operasi Tenaga Listrik Jawa-Bali*. Depok:PLN Persero

Control of TCSC and SVC Using Least Square Support Vector Regression (LS-SVR) to Improve Voltage Stability

Rony Seto Wibowo, Bahrowi AW, K Anam,
M Abdillah, Adi Soeprijanto, Ontoseno Penangsang
Institut Teknologi Sepuluh Nopember, Indonesia
e-mail: ronyseto@ee.its.ac.id

Naoto Yorino
Hiroshima University, Japan

Abstract — **This paper proposes the application of *Least Square Support Vector Regression* (LS-SVR) for controlling *Flexible AC Transmission Systems* (FACTS) in order to meet voltage stability requirement. Transient voltage stability is a very fast phenomenon. Therefore, the proposed approach is aimed to provide a quick response to prevent voltage collapse. Generally, time response consists of two parts. Firstly, control center receives signals from the field and then process those signals to determine the appropriate setting of FACTS devices according to load level and location of fault. Secondly, FACTS devices react based on the signals sent by control center to prevent voltage collapse. The total response time should be shorter than the time to voltage collapse. Two kinds of FACTS devices, *Thyristor Controlled Series Capacitor* (TCSC) and *Static VAR Compensator* (SVC), are used to represent series and shunt type devices, respectively. To prove the effectiveness of the proposed approach, IEEE 14 buses is used as test system. In addition, comparison study between application of LS-SVR and Extreme Learning Machine (ELM) is also presented.**

Keywords; Contingency; Extreme Learning Machines; Support Vector Machines; Voltage Stability.

I. INTRODUCTION

IN recent years, power system demand increases significantly. On the other hand, transmission line capacity expansion is difficult to do due to some reasons such as limited space and environment concern. In this situation, voltage stability becomes a vital issue as system operation shifts closer to collapse point due to lack of reactive power and system capacity.

Voltage collapse can occur very fast. As a result, immediate actions are required to avoid it. Flexible AC Transmission systems (FACTS) is one of technologies able to handle voltage instability quickly by injecting reactive power or modifying power flow [1,2] as it is a power electronic based technology.

Appropriate setting of FACTS devices can be determined using optimal power flow solved by sequential quadratic programming [3]. Nevertheless, it may take a long computation time. The larger is the considered system, the longer is the computation time. For that reason, intelligent computation system may become a promising solution. *Least Square Support Vector Regression* (LS-SVR) is one of recent intelligent approach. This approach is an extension of neural network which is able to predict some output values based on given inputs. The application of LS-SVR has been used in predicting the parts weight in plastic injection molding [4] and aero-engine modeling [5].

This paper presents the application of multi output LS-SVR as a controller of Thyristor Controlled Series Capacitor (TCSC) and Static VAR Compensator (SVC). Those devices are not only used under normal state but also under contingency state. Input of the controller are load level and system configuration while output of controller are suitable setting of FACTS devices to maintain required voltage stability index under both normal and contingency circumstances. Load margin is the utilized index to measure the level of voltage stability in this paper. To show the effectiveness of the proposed approach, IEEE 14 bus is used as test system.

II. VOLTAGE STABILITY

A. *Definition and Time Scale*

Study of voltage stability can be done via bifurcation analysis. Bifurcation curve is developed by changing system parameters slowly. In power system, load is usually used as changed parameter. Naturally, system becomes unstable as the load raises. The illustration of bifurcation curve can be seen in fig. 1. In a saddle-node bifurcation, a stable operating equilibrium disappears as parameters change. As a result, system states dynamically collapse. In a computation point of view, collapse is found when computation cannot converge. A critical operating point is defined as a state slightly before collapse point. In general, voltage stability is measured by an index. One of voltage stability indexes widely used is load margin. This index measures the distance of current operating point to the collapse point.

Time scale of voltage stability is generally divided into two categories, transient voltage stability and long term voltage stability. Power electronic based devices such as TCSC and SVC are suitable to handle transient voltage stability while load tap changer and generation re-scheduling are only capable to tackle long term voltage stability. The detailed time scales of voltage stability and theirs related devices are described in fig. 2 [6].

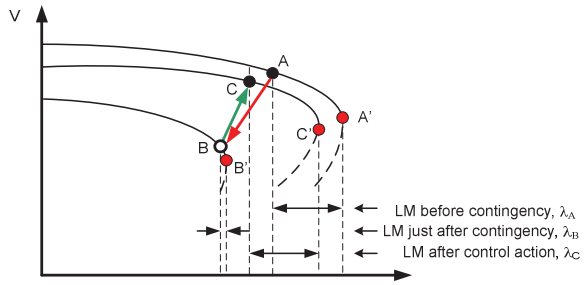


Figure 1. System state transitions

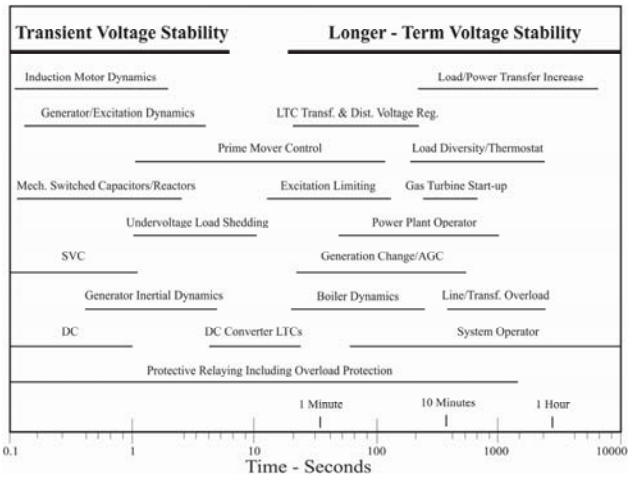


Figure 2. Time scales of voltage stability

B. *FACTS devices to improve Voltage Stability*

FACTS devices are valuable for improving voltage stability. In this paper, two kinds of FACTS devices are used i.e. TCSC and SVC. These devices represent series and shunt types, respectively. TCSC is modeled as series variable capacitor and installed in series with transmission line. Changing the impedance of the capacitor will automatically change the impedance of the transmission line. As a result, power flow will consequently be modified. In this paper, impedance of capacitor is chosen as control variable to improve voltage stability.

Fig. 3 illustrates the installed TCSC. From that figure, we can obtain :

$$X_{ij} = X_{Line} + X_{TCSC}, \quad (1)$$

where :

X_{ij} Modified transmission line impedance
X_{line} Transmission line impedance
X_{TCSC} TCSC impedance

SVC is able to inject reactive power to the system and installed in parallel with bus. In this paper, SVC is modeled as generator with no active power. Figure 4 displays model of SVC.

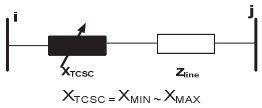


Figure 3. TCSC Model

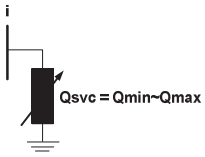


Figure 4. SVC Model

The application of FACTS devices to improve voltage stability under contingency state is illustrated in fig. 1. It is assumed that contingency is the outage of one line of double track transmission line. Point A is the operation point under normal circumstances with load margin 0.1. If contingency take places, operation point moves to point B that may be closer to collapse point B' with load margin less than 0.1. Therefore, control actions should be taken to improve load margin to prevent voltage collapse. The actions are controlling FACTS devices, re-scheduling power generation and shedding load. In this case, just after contingency, FACTS devices quickly react to prevent voltage collapse as well as minimize generation re-scheduling and load shedding. After FACTS devices respond the contingency, the PV curve changes and the operation point move to point C with adequate load margin which is equal to or greater than 0.1.

III. LEAST SQUARE SUPPORT VECTOR REGRESSION (LSSVR)

In this paper, multi output LS-SVR is applied. This method is the extension of single output LS-SVR. Before going to the detailed formulation, the following definitions and functions [7] are required to understand.

- Radial Basis Function Kernel is denoted and formulated as $k(x,y) = exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$, with σ^2 as kernel parameter.
- $trace(\boldsymbol{A})$ is summation of diagonal elements of matrix $\boldsymbol{A}$.
- $repmat(\boldsymbol{A},m,n)$ or $repmat(\boldsymbol{x},m,n)$ is to develop replica matrix of $\boldsymbol{A}$ or x of m x n dimension.
- $blockdiag(\boldsymbol{A}_1,\boldsymbol{A}_2,..,\boldsymbol{A}_n)$ or $blockdiag\ (\boldsymbol{x}_1, \boldsymbol{x}_2,..,\boldsymbol{x}_n)$ is to develop diagonal matrix with $\boldsymbol{A}$ **or** $\boldsymbol{x}$ as element diagonal.

Lets,

$$D_n = (x_i,y_i)\ldots(x_j,y_j);\ \ \boldsymbol{x_i} \in \mathbb{R}^d,\ \boldsymbol{y^i} \in \mathbb{R}^m \quad (2)$$

D_n is a set of data with *x* as input data and *y* as target output data. Training of LS-SVR was conducted by solving the following formulation.

$$\min_{\boldsymbol{w}\in\mathbb{R}^{n_h xm},\ \boldsymbol{b}\in\mathbb{R}^m} \Im(\boldsymbol{w},\boldsymbol{\xi}) = \frac{1}{2}trace(\boldsymbol{w}^T\boldsymbol{w}) + \frac{1}{2}\gamma trace(\boldsymbol{\xi}^T\boldsymbol{\xi}) \quad (3)$$

Subject to equality constraints:

$$\boldsymbol{Y} = \boldsymbol{Z}^T\boldsymbol{w} + repmat(\boldsymbol{b}^T, l, \boldsymbol{1}) + \boldsymbol{\xi} \quad (4)$$

Where $\boldsymbol{Z} = \big(\varphi(\boldsymbol{x}_1), \varphi(\boldsymbol{x}_2), \dots \varphi(\boldsymbol{x}_l)\big)\mathbb{R}^{n_h xl}$; $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}^{n_h}$ is transformation from input to the higher dimension feature space or Hilbert space, $\mathcal{H}$, with n_h dimension; $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_m) \in \mathbb{R}_+^{lxm}$ is a vector consisting of slack variable and $\gamma \in \mathbb{R}_+$ is regularized parameter with positive real number.

Equation (3,4) is then solved using Lagrangian Function and formulated as follow:

$$\mathcal{L}(\boldsymbol{w_0}, \boldsymbol{b}, \boldsymbol{\xi}, \boldsymbol{A}) = \Im(\boldsymbol{w},\boldsymbol{\xi}) - trace(\boldsymbol{A}^T(\boldsymbol{Z}^T\boldsymbol{w} + repmat(\boldsymbol{b}^T, l, \boldsymbol{1}) + \boldsymbol{\xi} - \boldsymbol{Y})) \quad (5)$$

Where $\boldsymbol{A} = (\alpha_1, \alpha_2, \dots, \alpha_m) \in \mathbb{R}^{lxm}$ is Lagrange multipliers .

By satisfying Karush-Kuhn-Tucker (KKT) condition, the following linier functions are obtained.

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w_0}} = 0 \rightarrow \boldsymbol{w_0} = \sum_{i=1}^m \boldsymbol{ZA} \quad (6)$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{b}} = 0 \rightarrow \boldsymbol{A}^T\boldsymbol{1}_l = \boldsymbol{0}_l \quad (7)$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\xi}} = 0 \rightarrow \boldsymbol{A} = \gamma\boldsymbol{\xi} \quad (8)$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{A}} = 0 \rightarrow \boldsymbol{Z}^T\boldsymbol{w} + repmat(\boldsymbol{b}^T, l, \boldsymbol{1}) + \boldsymbol{\xi} - \boldsymbol{Y}) = \boldsymbol{0}_{lxm} \quad (9)$$

Parameters $\boldsymbol{w}$ and $\boldsymbol{b}$ are initial *weight* and *bias* with n_h dimension, respectively, Where $\boldsymbol{w} = w_1, w_2, \dots, w_m \in \mathbb{R}^{n_h xm}$ and $\boldsymbol{b} = (b_1, b_2, \dots b_m)^T \in \mathbb{R}^m$.

Similar to *single output regression*, the solution of equations (6-9) can be rewritten as the following term.

$$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{1}^T \\ \boldsymbol{1}^T & \boldsymbol{\Omega} + \boldsymbol{\gamma}^{-1}\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_l \\ \boldsymbol{\alpha}_l \end{bmatrix} = \begin{bmatrix} \boldsymbol{0}^T \\ \boldsymbol{Y}_l \end{bmatrix} \quad (10)$$

Where kernel matrix $\boldsymbol{\Omega}_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$;
$\boldsymbol{1} = [1,1, \dots, 1]^T$, $\boldsymbol{0} = [0,0, \dots, 0]^T$;

$$\boldsymbol{Y_l} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_l] = \begin{bmatrix} \boldsymbol{y}_{11} & \cdots & \boldsymbol{y}_{1l} \\ \vdots & \ddots & \vdots \\ \boldsymbol{y}_{n1} & \cdots & \boldsymbol{y}_{nl} \end{bmatrix};$$

$$\boldsymbol{\alpha}_l = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_l] = \begin{bmatrix} \boldsymbol{\alpha}_{11} & \cdots & \boldsymbol{\alpha}_{1l} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\alpha}_{n1} & \cdots & \boldsymbol{\alpha}_{nl} \end{bmatrix}$$

From equations above, Output of multi output LS-SVR can be attained through the following equation :

$$\hat{y}(\boldsymbol{x}) = \sum_{i=1}^n \boldsymbol{\alpha}_{ij}\boldsymbol{k}(\boldsymbol{x}, \boldsymbol{x}_i) + \boldsymbol{b}_j; \quad j = 1,2, \dots, l \quad (11)$$

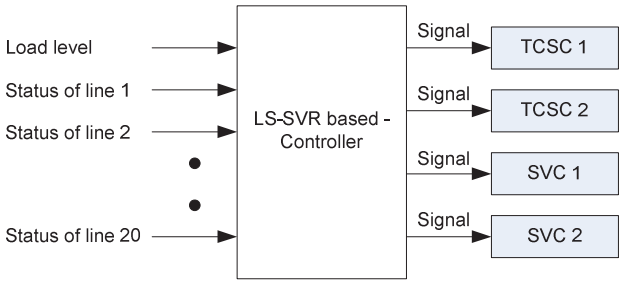


Figure 5. Implementation scheme

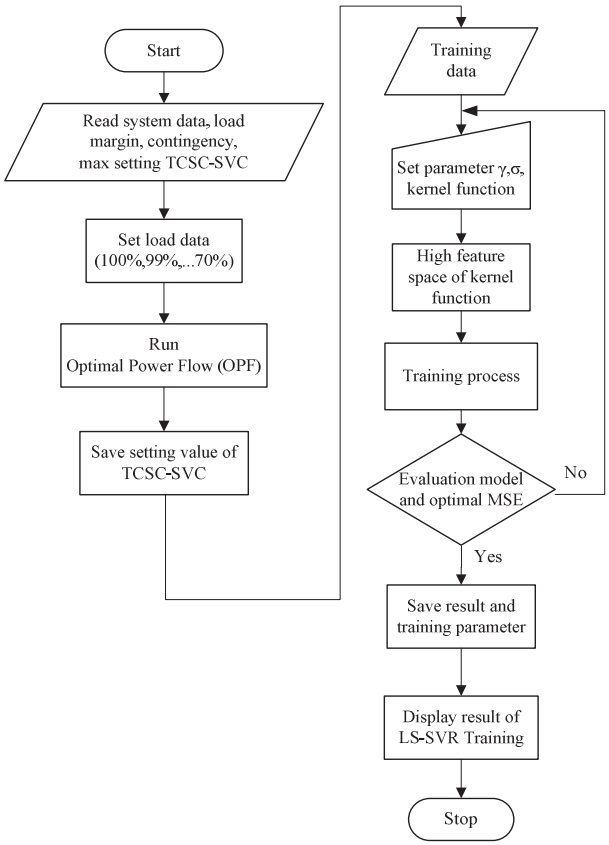
IV. IMPLEMENTATION SCHEME

The implementation scheme of the proposed approach can be seen in fig. 5. The input data of the controller are load level and status of transmission lines. For each load level, it is assumed that generators are loaded in economic manner. On the other word, power output of generators result from optimal power flow considering voltage stability [3]. Line status is equal to "1" if the line transmission is in service and equal to "0" if the transmission line is outage. Thus, location of line outage is known by receiving status signal from the field. The status of line can also be represented by Circuit breaker (CB) status of the corresponding line. After receiving data from the field, controller will process the data and then send signals to TCSC and SVC. In this case, signals represent equivalent capacitance of TCSC and reactive power that should be injected by SVC to restore the load margin to specified value i.e. 0.1.

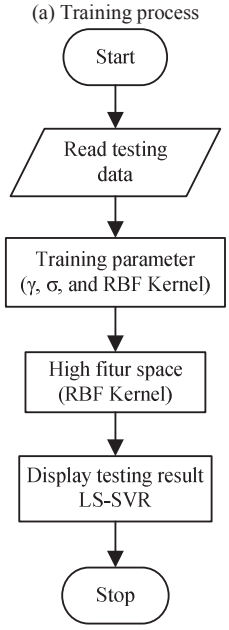
V. SOLUTION ALGORITHM

In this section, application of multi output LS-SVR for determining required setting of TCSC-SVC is described. Flowchart proposed approach is displayed in fig. 6 while procedures are described as follow:

1. Read data consisting of power system data, load margin, contingency lines, TCSC and SVC capacities.
2. Determine several load levels that will be considered in the training process of LS-SVR. Along with contingency lines, these load levels become input data of training process.
3. Run Optimal Power Flow (OPF) for several load levels in step 2 using [3] to obtain appropriate settings of TCSC and SVC that will be used as target data.
4. Set γ and σ used to find optimal high feature space during training process.
5. Run multi output LS-SVR as training process.
6. After completing training process, the result is evaluated using *Mean Square Error* (MSE) as criterion.
7. If optimal MSE is achieved, save parameters of LS-SVR, then testing process is conducted. Otherwise, back to step 4.
8. Save testing result according to equation (11) and process is end.



(a) Training process



(b) Testing process

Figure 6. Flowchart of controlling TCSC-SVC by multi output LS-SVR

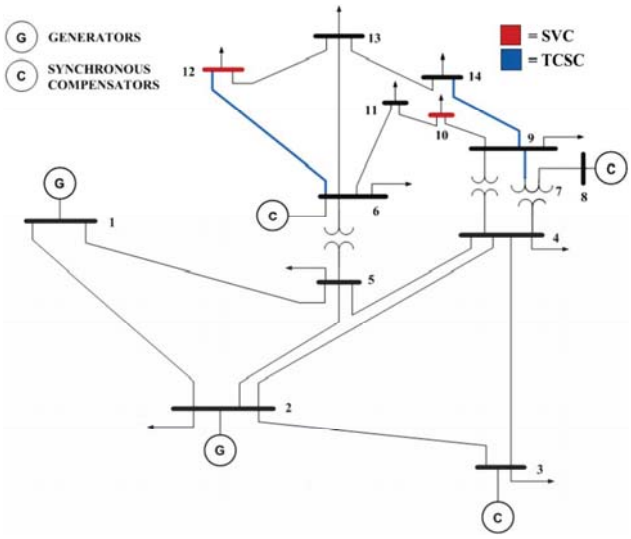


Figure 7. Locations of TCSC and SVC

VI. CASE STUDY

The proposed approach is applied to IEEE 14 bus to show the effectiveness of the approach in determining settings of TCSC and SVC in order to maintain load margin greater than 0.1 under normal state and all considered contingency states. Fig. 7 displays the IEEE 14 bus. Location of TCSC is indicated by blue lines while SVC is indicated by red lines. Those locations are already optimized based on [3].

The used kernel function is *Radial Base Function* (RBF) while parameters γ and σ is determined by *trial and error.* Performance of training with different values of γ and σ are shown in table 1. To evaluate the performance, Mean Square Error (MSE) is used as criteria. Fig. 8 and fig. 9 display settings of TCSC 1 under various load levels using γ = 1000 and σ = 0.01 as training and testing results, respectively. Fig. 8 shows that result of training process is in line with target data while fig. 9 indicates that result of testing is almost similar to target data.

Table 2 indicates that LS-SVR is better than ELM in term of MSE and MAPE training. On the other hand, ELM is better than LS-SVR in term of training time. In addition, as seen in table 3, LS-SVR is also better than ELM in term of accuracy of testing results. The detailed testing output under normal and several contingency states using both LS-SVR and ELM are shown in table 4 while table 5 gives detailed error of testing using both ELM and LS-SVR. With load margin 0.1 and very small error testing results, it can be guaranteed that voltage stability can be maintained.

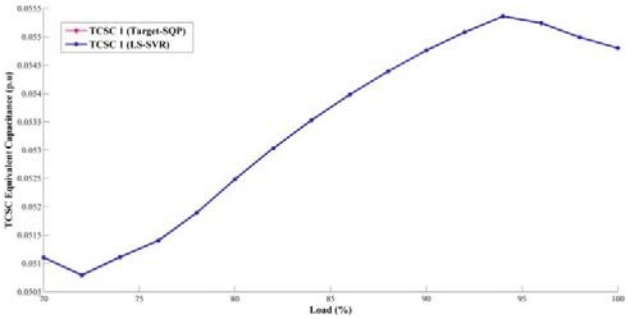


Figure 8. Result of LS-SVR training with γ = 1000 and σ = 0,01 (TCSC 1, in normal state)

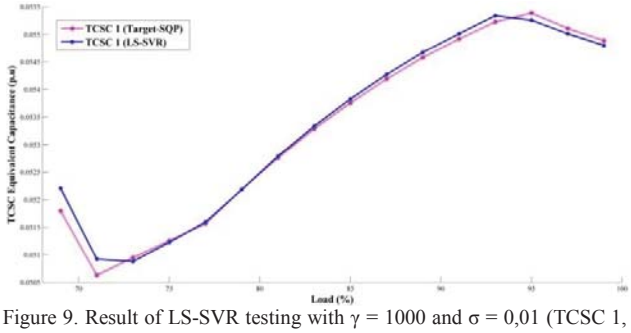


Figure 9. Result of LS-SVR testing with γ = 1000 and σ = 0,01 (TCSC 1, in normal state)

TABLE I. PERFORMANCE OF KERNEL PARAMETER IN LS-SVR

RBF Kernel Parameters		MSE Training	Training Time (second)
γ = 100	σ = 10	0.00670062	0.0624004
	σ = 1	0.00174889	0.0624004
	σ = 0	0.00174889	0.0780005
	σ = 0.1	0.0001755	0.6396041
	σ = 0.01	4.2414e-06	0.5928038
γ = 1000	σ = 10	0.00499541	0.3744024
	σ = 1	0.0013413	0.2184014
	σ = 0	0.00373178	0.0780005
	σ = 0.1	9.7008E-05	0.3432022
	σ = 0.01	4.2955e-08	0.0780005

TABLE II. ERROR TRAINING AND TRAINING COMPUTATION TIME

Method	MSE Training	MAPE Training	Training time (second)
ELM	8.7556e-06	0.044124	0.0469
LS-SVR	4.2955e-08	10e-07	0.0780

TABLE III. ERROR TESTING AND TESTING COMPUTATION TIME

Method	MSE Testing	MAPE Testing	Testing time (second)
ELM	0.00079182	0.000125686	0.0312002
LS-SVR	0.00010343	1.641768e-05	0.0468003

TABLE IV. DETAILED SETTING OF FACTS DEVICES FOR IEEE 14 BUS SYSTEM USING LS-SVR AND ELM

Load levels (MW)	Contingency lines		Target					Output Using LSSVR					Output Using ELM				
			TCSC equivalent Capacitance (pu)			Injected reactive power of SVC (MVAr)		TCSC equivalent Capacitance (pu)			Injected reactive power of SVC (MVAr)		TCSC equivalent Capacitance (pu)			Injected reactive power of SVC (MVAr)	
	From	To	Line 6-12	Line 7-9	Line 9-14	Bus 10	Bus 12	Line 6-12	Line 7-9	Line 9-14	Bus 10	Bus 12	Line 6-12	Line 7-9	Line 9-14	Bus 10	Bus 12
249.084			0.0549	0.077	0.1892	20	20	0.0548	0.077	0.1892	20	19.9964	0.0550	0.0766	0.1938	20	20.0343
244.052	Normal State		0.0551	0.077	0.1892	20	20	0.0550	0.077	0.1891	20	19.9757	0.0551	0.0772	0.1879	20	20.0042
239.020			0.0554	0.077	0.1892	20	20	0.0553	0.077	0.1892	20	20.0002	0.0553	0.0769	0.1899	20	19.9494
249.084			0.0550	0.077	0.1892	20	20	0.0550	0.077	0.1892	20	19.9964	0.0539	0.0769	0.1696	20	20.0056
244.052	1	5	0.0552	0.077	0.1892	20	20	0.0551	0.077	0.1891	20	19.9749	0.0555	0.0770	0.1946	20	19.9904
239.020			0.0555	0.077	0.1892	20	20	0.0553	0.077	0.1892	20	20.0084	0.0553	0.0770	0.1865	20	20.0181
218.892			0.0534	0.077	0.1892	20	18.160	0.0535	0.077	0.1892	20	18.2701	0.0533	0.0770	0.1873	20	18.1880
213.860	2	4	0.0530	0.077	0.1892	20	17.534	0.0530	0.077	0.1892	20	17.6124	0.0532	0.0771	0.1899	20	17.5886
208.828			0.0524	0.077	0.1892	20	16.900	0.0525	0.077	0.1891	20	16.9470	0.0526	0.0770	0.1909	20	16.9650
218.892			0.0528	0.077	0.1892	20	17.780	0.0529	0.077	0.1892	20	17.8870	0.0526	0.0770	0.1869	20	17.8275
213.860	3	4	0.0524	0.077	0.1892	20	17.171	0.0524	0.077	0.1892	20	17.2487	0.0525	0.0770	0.1922	20	17.2415
208.828			0.0519	0.077	0.1892	20	16.553	0.0519	0.077	0.1891	20	16.6002	0.0520	0.0770	0.1909	20	16.6122
193.732			0.0417	0.077	0.1892	20	14.005	0.0391	0.077	0.1760	20	14.0549	0.0380	0.0771	0.1531	20	14.1254
188.700	6	11	0.0360	0.077	0.1892	20	13.536	0.0355	0.077	0.1909	20	13.4555	0.0364	0.0770	0.1993	20	13.6694
183.668			0.0362	0.077	0.1892	20	13.043	0.0364	0.077	0.1888	20	12.9657	0.0363	0.0769	0.1959	20	13.2053
193.732			0.0131	0.077	0.1892	20	15.263	0.0137	0.077	0.1891	20	15.2178	0.0106	0.0770	0.1881	20	15.3440
188.700	6	13	0.0129	0.077	0.1892	20	14.601	0.0129	0.077	0.1892	20	14.5188	0.0132	0.0770	0.1889	20	14.7407
183.668			0.0132	0.077	0.1892	20	14.101	0.0134	0.077	0.1892	20	14.0199	0.0146	0.0770	0.1902	20	14.1959

TABLE V. DETAILED TESTING ERRORS USING LS-SVR AND ELM FOR IEEE-14 BUS SYSTEM

Load levels (MW)	Contingency lines		Error Testing (%) LSSVR					Error Testing (%) ELM				
			TCSC			SVC		TCSC			SVC	
	From	To	Line 6-12	Line 7-9	Line 9-14	Bus 10	Bus 12	Line 6-12	Line 7-9	Line 9-14	Bus 10	Bus 12
249.084	Normal State		0.130	0.011	0.075	0	0.179	8.037	0.254	7.962	0	0.071
244.052			0.153	0.014	0.100	0	0.237	2.187	0.104	2.163	0	0.064
249.084	1	5	0.096	0.011	0.075	0	0.179	0.325	0.498	3.349	0	0.148
244.052			0.124	0.014	0.100	0	0.241	0.086	0.154	0.577	0	0.102
218.892	2	4	0.215	0.011	0.075	0	0.523	0.172	0.054	1.148	0	0.175
213.86			0.182	0.011	0.079	0	0.383	0.195	0.066	0.869	0	0.274
218.892	3	4	0.227	0.011	0.075	0	0.530	0.249	0.073	0.578	0	0.308
213.860			0.187	0.011	0.079	0	0.400	0.623	0.004	0.012	0	0.390
193.732	6	11	5.778	0.012	7.067	0	0.422	8.165	0.061	17.931	0.001	0.948
188.700			1.050	0.011	0.854	0	0.496	2.208	0.034	6.654	0.001	1.089
193.732	6	13	5.864	0.012	0.081	0	0.279	13.889	0.036	0.053	0.001	0.622
188.700			2.229	0.011	0.079	0	0.514	5.745	0.053	0.091	0.001	1.035

VII. CONCLUSIONS

Determining settings of TCSC and SVC is a crucial step in handling voltage stability. Accurate and rapid response is greatly required in order to prevent voltage collapse. This paper has presented the application of multi output LS-SVR to control TCSC and SVC. It is observed that the method results in very accurate output with fast response. In addition, the results of simulations indicate that LS-SVR is more accurate than ELM.

For further research, to enhance the applicability of the proposed approach, type of contingency can be extended by considering generator outage. In addition, the online generator may be used as data input of training and testing process.

VIII. REFERENCES

[1] Y. H. Song and A. T. Johns, "Flexible ac transmission systems (FACTS)," in *IEE Power and Energy Series*, U.K., 1999.

[2] C. A. Caizares, Editor, "Voltage stability assessment: concepts, practices and tools", IEEE-PES Power System Stability Subcommittee Special Publication, SP101PSS, May 2003.

[3] R. S. Wibowo, N. Yorino, M. Eghbal, Y. Zoka, and Y. Sasaki, "Facts Devices Allocation with Control Coordination Considering Congestion Relief and Voltage Stability", IEEE Trans. Power Syst., vol. 26, no. 4, pp 2302-2310, Nov. 2011.

[4] Feng, Lu ; Jinquan, Huang ; Xiaojie, Qiu, Application of multi-outputs LSSVR by PSO to the aero-engine model", Systems Engineering and Electronics, Journal of Volume: 20 , Issue: 5 Publication Year: 2009 , Page(s): 1153 – 1158

[5] Xiaoli Li ; Bin Hu ; Ruxu Du, "Predicting the Parts Weight in Plastic Injection Molding Using Least Squares Support Vector Regression", Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on Volume: 38 , Issue: 6, 2008 , Page(s): 827 – 833

[6] Carson W Taylor, "Power System Voltage Stability", McGraw-hill, Singapore, 1994

[7] Shou Xu, Xiaodong Qiao, L Zhu, Lin Li, "Multi-Output Least Square Support Vector Regression Machines", Elsevier, Pattern Recognition, 2013

Design of a Portable Pico Linear Permanent Magnet Generator for Wave Energy Conversion

M. A. F. M. Hamim, T. Ibrahim and N. M. Nor
Department of Electrical and Electronic Engineering,
Universiti Teknologi PETRONAS
31750, Tronoh, Perak, Malaysia
taibib@petronas.com.my

Abstract— This paper presents the design and modeling of a portable pico linear permanent magnet generator for the application of wave energy conversion system. Two types of generator technology namely the rotary and linear are identified but due to the high maintenance and low efficiency of rotary, linear generator is preferable. Three linear permanent magnet generator designs with different types of permanent magnet shape have been proposed. Preliminary results for the air gap distribution and open circuit test are demonstrated and discussed are shown to be as expected.

Keywords-Wave energy conversion; slotless ; linear permanent magnet generator; air gap distribution; open circuit test

I. INTRODUCTION

Renewable energy sources were introduced in 1973 during the highly installed capacities and in 1983 during the petroleum crisis [1]. During these years, rapid decrement of global fuel stock is moving proportionally to the increment in demand of electrical energy. The uses of non-renewable fuels, such as fossil fuels (oil, coal and gas) may lead to pollution, acid rain and global warming. These fossil resources are depleting at a very fast rate along with the increasing fuel prices. All the mentioned circumstances lead the engineers and scientist to discover and implement a new alternative technology that uses alternative resources.

Sea wave has been the least utilized source amongst renewable energy sources [2]. One of the reasons for this is the difficulty to find a suitable system to convert the wave energy into electrical energy. However, the sea or ocean wave is a potential renewable source in term of competitive economics. Total energy harvested from the waves in the world's coasts is estimated of 10^6 MW and only 2 % from the amount is enough to supply the whole energy demand [3].

Various interest in the wave energy against other forms are that the waves is predictable, poses higher energy density, reduces in visual impact and provide more power extracted to the energy converter with a small volume of wave [4]. Other than that, wave energy also gives more concentrated and persistent compared with wind and solar energy [5].

Recently, direct drive linear generator becomes the most popular type of generator investigated by the researchers for the wave energy conversion (WEC). Linear generator is useful for the application when the motion is linear. Linear generator has a different topology from the rotary machine but has similar working principle. There are several advantages and disadvantages of linear generator as compared to rotary generator as shown in Table I. Linear generator have several types [6] for WEC system. However in this paper only one type will be considered, that is linear permanent magnet generator.

TABLE I. Comparison between linear and rotary generator [1], [4], [6-8]

Linear Generator	Rotary Generator
Provide lifetime 20 years or longer	Short lifetime due to the mechanical error
Direct drive from linear motion	Need gear box to convert linear motion to rotary
Simple structure and robustness	Small in size but complicated structure
High efficiency, capability to convert calm wave	Low efficiency and low reliability
Large up-front cost and huge size	Small initial cost but need frequent maintenance
Voltage and frequency variation	Large air-gap diameter; good in mechanically

The objective of this work is to design an economical topology for used in linear permanent magnet generator based on different shape of permanent magnet. The desired generator is planned to be used for the fisherman or outdoor activity that need low power consumption and easy to carry. Hence, the design have to achieve certain criteria such as cheap, light in weight, portable and can produce power output up to 100 W.

II. RELATED WORKS

Bracco et al. [9] discussed and proposed the tubular, slottless, iron-cored and axial magnetized magnet for the design generator. Among all linear generators, tubular is chosen because of the highest power density produced. Existing design usually used slotted rather than slottless where slotted is famous for reducing air gap length between stator and translator and therefore will produce higher power output or achieve higher flux density. The main concern of applying slotted in the design is to introduce cogging force due to the present of stator teeth. Comparison has been made in term of performance between single phase and three phase and clearly shows that three phase is almost 10 % higher than the single

This work is supported in part by Universiti Teknologi PETRONAS (UTP).

phase. However, slotted design required utilization of copper and this can be reduced if slotless design is used since the devoid of stator.

Oprea et al. [10] investigated the different between tubular air-cored and iron-cored in term of the suitable parameter for used in the design of the generator. Iron-cored has higher magnetic flux density for both the poles and slot compared to air-cored. But, the problem of iron-cored is the present of electromagnetic force in the generator. Besides that, iron-cored also induced and generated 14 times higher voltage and power than air-cored. Fast flux variation need to be removed as it can produce iron losses where prevention need to be taken by laminating the stator magnetic circuit. Meanwhile, the different between tubular and four-sided generator is that the tubular has constant magnetic flux for entire air gap whilst four-sided experience lower and drop in magnetic flux toward the pole corner. However, induced voltages of the latter are similar.

Cheng et al. [11] presented the development of direct drive Slotless Tubular Linear Generator (STLG) for WEC application. The generator used is the quasi-halbach magnetized magnet where the conventional STLG used is axial and radial magnetizations. Radial and axial magnetizations require ferromagnetic material to guide the desire flux path. Besides, operational temperature and eddy current in the generator produced additional power losses and hence, reduce the performance and output. This paper also discussed on optimization process using Taguchi's method.

III. PROPOSED DESIGN

Based on the reviews, tubular type is selected over planar as it provides constant air gap and higher power density [12]. Iron-cored is used for the stator part where it is superior in term of magnetic flux density and induced voltage [13]. For translator part, moving magnet is used as it gives high force capability even with different magnetization [14]. Slotless stator provides minimal cogging force and reduces the material cost [12]. Halbach magnetization is implemented in the design as no ferromagnetic material is required due to the self-shielding magnetized [14].

The basis of Faraday's Law in electromagnetic induction is shown in eq. (1)

$$e = \frac{d\lambda}{dt} \quad (1)$$

where voltage e is induced in the coil if there are flux linkage λ varies with the time t .

Meanwhile, induced voltage in the moving magnet can be determined as in eq. (2)

$$E = 4.44 f B_m A_g N \quad (2)$$

where f = frequency of the moving magnet
A_g = surface area of magnet
N = total number of turns in stator
B_m = maximum air gap flux density

Three different designs are proposed for the slotless linear permanent magnet generator is shown in Fig. 1 to 3. The common criteria of designing the generator are to reduce the weight and cost by maintaining the desire output power. To achieve the target, the generator will be designed by focusing on the permanent magnet. This is because permanent magnet is expensive and heavy. These three proposed designs differ in term of the permanent magnet shape.

The first proposed design as shown in Fig. 1 is known as rectangular layout and used as reference to the other layout. The advantage of this design is the shape of the permanent magnet is easy to construct using full rectangular layout.

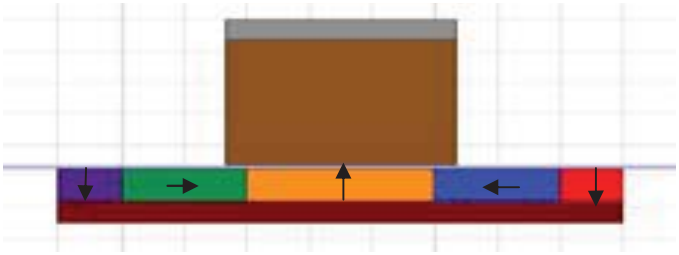


Fig. 1. 2D design for the 1st proposed generator

The second proposed design as shown in the Fig. 2 is called T-halbach layout since the shape of the permanent magnet resembles the letter T. The advantage of this design is low cost because of only half of rectangular shape is permanent magnet, as well as the design is easy to construct.

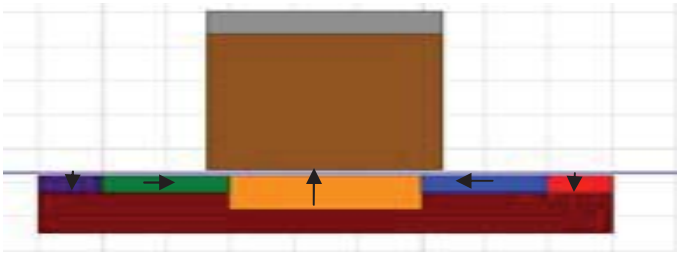


Fig. 2. 2D design for the 2nd proposed generator

The third proposed design as shown in the Fig. 3 is known as trapezoid-halbach layout. The advantage of this design is trapezoid shape of permanent magnet that will create the flux and focus more on the direction of the halbach using full permanent magnet.

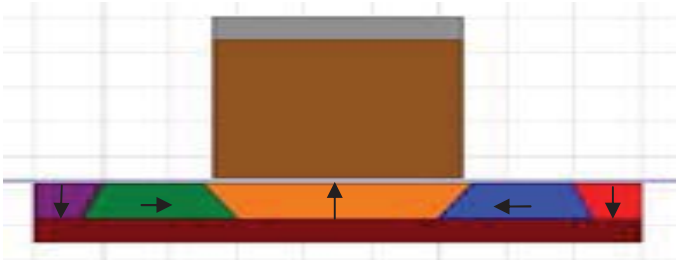


Fig. 3. 2D design for the 3rd proposed generator

Fig. 4 shows the schematic design used for the overall proposed designs with only different in the permanent magnet shape. The entire dimension is fixed to ensure the comparison based on the finite element method can be obtained. Table II tabulated the specific parameter or dimension used in the design.

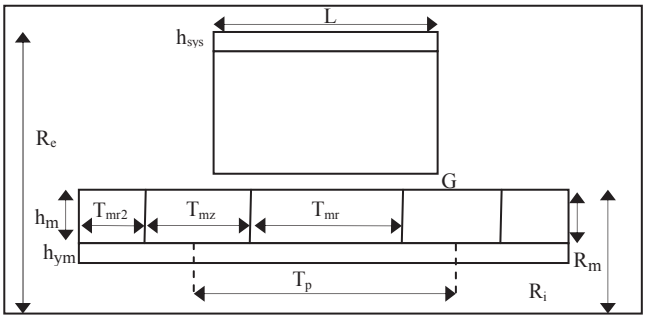


Fig. 4. Schematic design of the proposed design

TABLE II. Specific parameter used in the design

Parameter	Dimension (mm)
Outer radius of stator core, R_e	45.1
Yoke thickness of stator core, h_{sys}	3.3
Length of stator, L	18.4
Outer radius of magnet, R_m	21
Radial thickness of magnets, h_m	5
Radial thickness of supporting tube, h_{ym}	3.5
Inner radius of supporting tube, R_i	12.5
Pole pitch, T_p	25
Axial length of radically magnetized magnet at the centre, T_{mr}	15
Axial length of axially magnetized magnet at two end, T_{mr2}	5
Axial length of axially magnetized magnets, T_{mz}	10
Air gap length, G	0.8

IV. RESULTS AND DISCUSSION

A. Mesh, Flux Line and Flux Density in the Generator

The three proposed designs are compared based on three different results namely the mesh, flux line and flux density.

Mesh process is important for any design to be analyzed in the finite element method (FEM). The process is necessary to get the accurate result in the air gap distribution. For the good mesh plotted, the particles have to be as fine as possible. Fig. 5 to 7 show the mesh plotted for the three proposed designs.

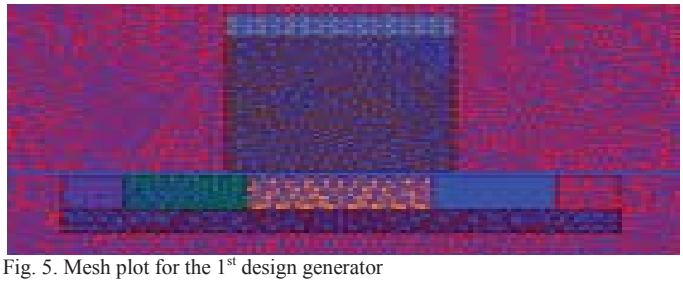


Fig. 5. Mesh plot for the 1st design generator

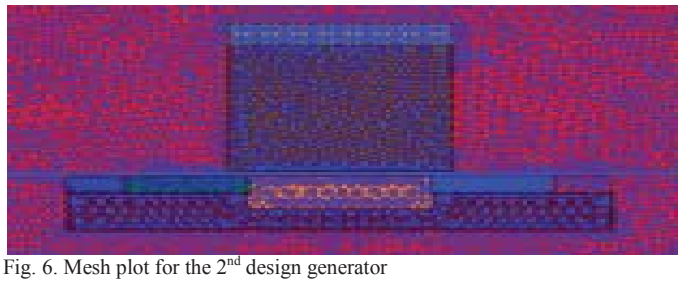


Fig. 6. Mesh plot for the 2nd design generator

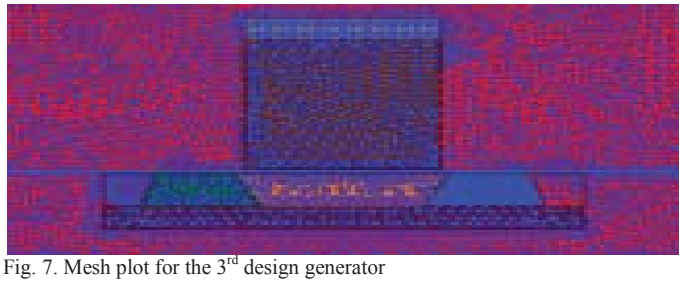


Fig. 7. Mesh plot for the 3rd design generator

Fig. 8 to 10 show the result of the flux line distribution in generator for the three designs. Halbach arrangement will create the flux line in round or closed loop pattern. Third proposed design is the smooth shape of the flux line meanwhile second proposed design has some distortion shape of flux line.

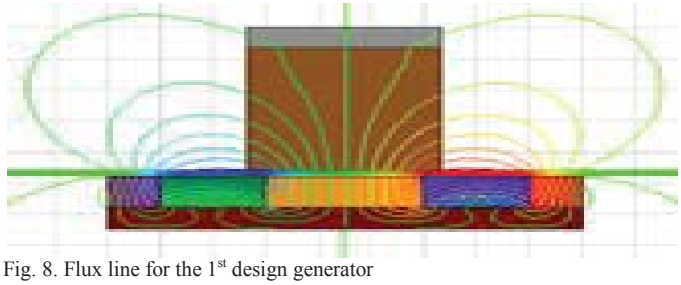


Fig. 8. Flux line for the 1st design generator

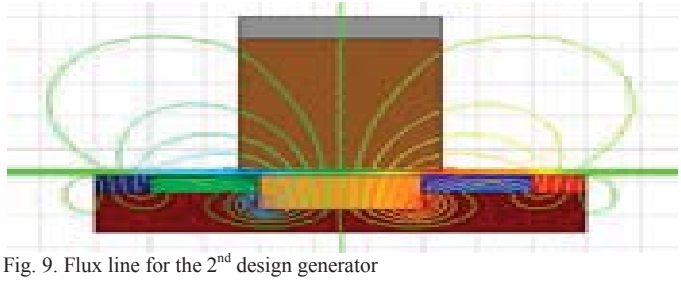


Fig. 9. Flux line for the 2nd design generator

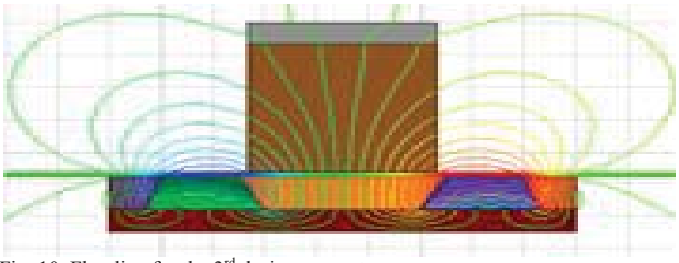


Fig. 10. Flux line for the 3rd design generator

On the other hand, the flux density is also an important parameter to ensure the performance of the generator. Based on the result shown in Fig. 11 to 13, first and third design have similar pattern of the flux density while the second design have different pattern. This is because the quantity of permanent magnet in first and third design is the same while only half in the second design.

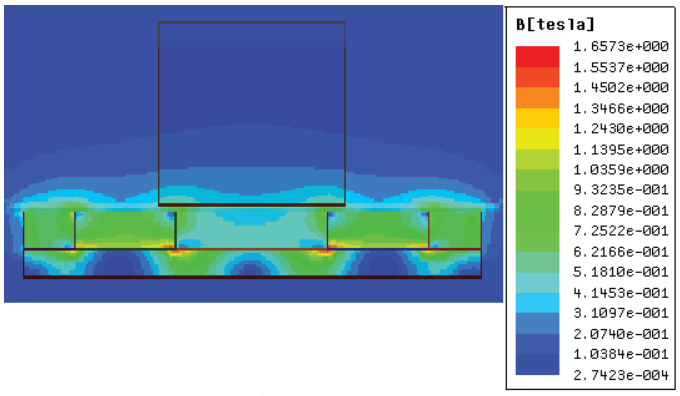


Fig. 11. Flux density for the 1st design generator

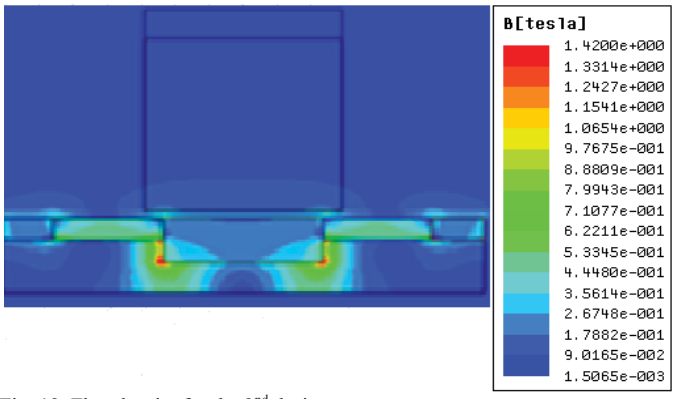


Fig. 12. Flux density for the 2nd design generator

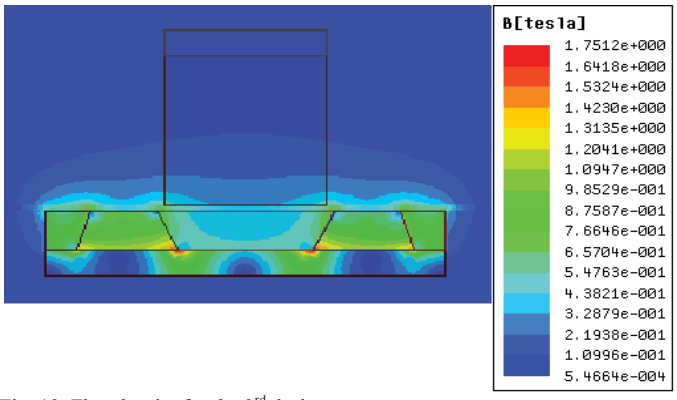


Fig. 13. Flux density for the 3rd design generator

B. Air Gap Distribution

The three proposed designs are compared based on two different results namely the flux line and flux density.

Fig. 14 shows the comparison of the three proposed designs of flux line in the air gap distribution. The result is symmetrical sinusoidal waveform because it is one of the characteristic of halbach. As depicted, the second design produced lesser flux line compared to the other two designs. Table III tabulated the value of flux line for the maximum and average of the three proposed designs. The third design has highest maximum value and second design produced lowest maximum value of flux line distributed in the air gap. Hence, the average value of the first design is the highest and the third design is the lowest.

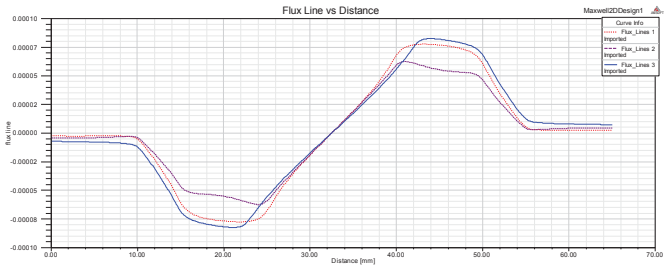


Fig. 14. Flux line in the air gap for the proposed design generator

TABLE III. Maximum and average value of flux line

Flux Line	Design 1	Design 2	Design 3
Maximum	7.75×10^{-5}	6.22×10^{-5}	8.22×10^{-5}
Average	1.61×10^{-8}	2.66×10^{-9}	3.80×10^{-9}

Fig. 15 shows the comparison of the three proposed designs in term of flux density in the air gap distribution. The graph is symmetric on both sides because halbach produced two closed loop flux line with different direction as shown in Fig. 8 to 10. Table IV tabulated the value of flux density for the maximum and average of the three proposed designs. The first design has the highest maximum value and the second design produced lowest maximum value of the flux density distributed in the air gap. Hence, the average value of the first design is the highest and the second design is the lowest.

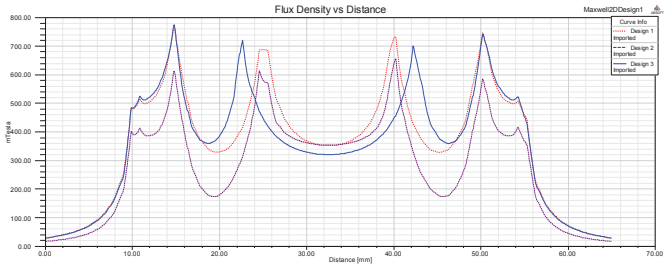


Fig. 15. Flux density in the air gap for proposed design generator

TABLE IV. Maximum and average value of flux density

Flux Density	Design 1	Design 2	Design 3
Maximum (T)	769.14	580.55	757.53
Average (T)	362.19	250.95	354.45

C. Open Circuit Test

The three proposed designs are compared based on four different results namely the induced voltage, flux linkage, thrust force and cogging force, with the specific parameter in the Table V.

TABLE V. Parameter for the open circuit test

Parameter	Value
Velocity	1.0m/s
Moving distance	10mm
Stop time	0.05s
Time step	0.001s
Step size	0.00005s

Fig. 16 shows the comparison of the three proposed designs in term of induced voltage. Table VI tabulated the value of induced voltage for the maximum and the average of the three proposed designs. The second design has produced lowest induced voltage because of the magnet quantity used in the shape which is approximately half from other design. The third design produced highest voltage due to the middle permanent magnet have longer open surface facing the coils.

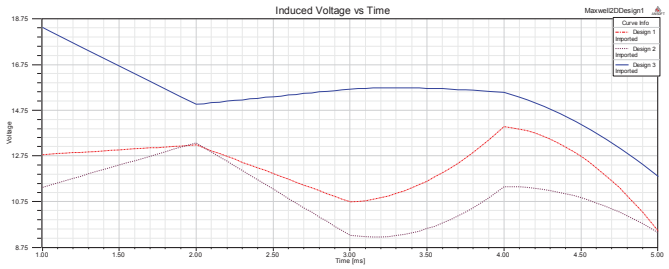


Fig. 16. Induced voltage in open circuit test for proposed design generator

TABLE VI. Maximum and average value of induced voltage

Induced Voltage	Design 1	Design 2	Design 3
Maximum (V)	14.02	13.18	18.51
Average (V)	12.29	10.52	15.66

Fig. 17 shows the comparison of the three proposed designs of flux linkage. The result is approximately increased in straight line where time is proportional with the flux. Table VII tabulated the value of flux linkage for the maximum and average of the three proposed designs. The third design has produced the highest maximum value and the second design has produced the lowest maximum value of flux line in the open circuit test. Flux linkage result is related the induced voltage where amount of flux linkage is proportional the quantity of induced voltage.

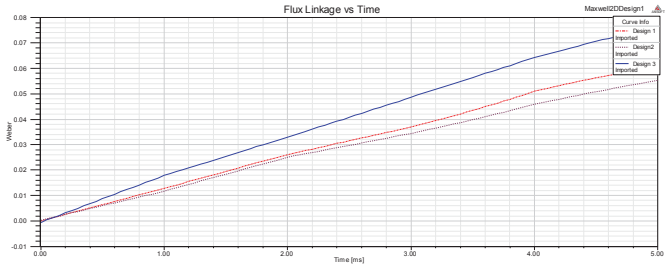


Fig. 17. Flux linkage in open circuit test for proposed design generator

TABLE VII. Maximum and average value of flux linkage

Flux Linkage	Design 1	Design 2	Design 3
Maximum (Wb)	0.06	0.05	0.08
Average (Wb)	0.03	0.03	0.04

Fig. 18 shows the comparison of the three proposed designs of thrust force in the generator. The result is approximately decreased in straight line where time is proportional with the thrust force. As depicted, the second design produced lesser force compared to the other two designs. Table VIII tabulated the value of thrust force for the maximum and the average of the three proposed designs. The first design has the highest maximum and the average value of force compare to the other designs. This result shows that first design has stronger bond between coil and permanent magnet.

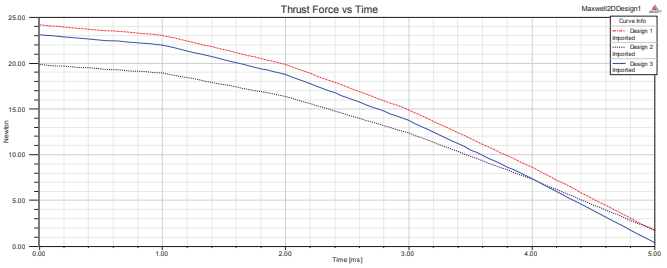


Fig. 18. Thrust force in the open circuit test for proposed design generator

TABLE VIII. Maximum and average value of thrust force

Thrust Force	Design 1	Design 2	Design 3
Maximum (N)	24.12	19.80	23.07
Average (N)	16.09	13.33	14.94

Fig. 19 shows the comparison of the three proposed designs in term of cogging force between stator and translator. The result is approximately sinusoidal wave between positive and negative value of force. Table IX tabulated the value of force for the maximum and the average of the three proposed designs. The second design has produced the highest maximum value and the third design has produced the lowest maximum value of cogging force in the open circuit test. Meanwhile, all the design in average show approximately 0N. Cogging force is an interaction between permanent magnet and tooth of stator. Even though all design is slotless, cogging force is appeared due to the longitudinal end issue.

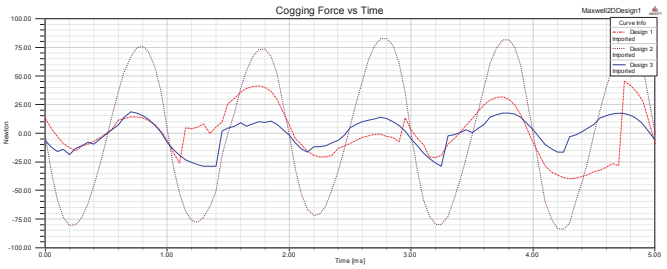


Fig. 19. Cogging force in the open circuit test for proposed design generator

TABLE IX. Maximum and average value of cogging force

Cogging Force	Design 1	Design 2	Design 3
Maximum (N)	45.23	82.70	29.01
Average (N)	0.19	1.02	0.97

Based on the overall findings, we can conclude that the quantity and shape of permanent magnet in the design affects the performance of the generator.

V. CONCLUSION

Trapezoid permanent magnet layout has produced impressive result based on the FEM as compared to the rectangular and T-halbach. With the same quantity of permanent magnet and inspiring result were obtained and compared to the rectangular, however higher cost need to be considered when to cut the permanent magnet into trapezoid shape. To obtain higher amount of induced voltage, higher flux need to be produced. This may effect to the performance of wave energy conversion system which may produce more induced voltage. In addition, the wave energy conversion is also associated to the velocity of wave.

ACKNOWLEDGMENT

The authors would like to thank Universiti Teknologi PETRONAS and Ministry of Higher Education for the awarded of ERGS grant.

REFERENCES

[1] Ivanova, I.A.; Agren, O.; Bernhoff, H.; Leijon, M.; , "Simulation of wave-energy converter with octagonal linear generator," *Oceanic Engineering, IEEE Journal of* , vol.30, no.3, pp.619-629, July 2005

[2] Cecconi, V.; Trapanese, M.; , "An optimum design of the magnetic circuit of a PM linear electrical generator for the exploitation of sea waves.," *Power Electronics, Electrical Drives, Automation and Motion, 2006. SPEEDAM 2006. International Symposium on* , vol., no., pp.116-119, 23-26 May 2006

[3] Faiz, J.; Ebrahimi-Salari, M.; , "Design and simulation of a 250 kW linear permanent magnet generator for wave energy to electric energy conversion in caspian sea," *Sustainable Power Generation and Supply, 2009. SUPERGEN '09. International Conference on* , vol., no., pp.1-6, 6-7 April 2009

[4] Prudell, J.; Stoddard, M.; Amon, E.; Brekken, T.K.A.; von Jouanne, A.; , "A permanent-magnet tubular linear generator for ocean wave energy conversion," *Industry Applications, IEEE Transactions on* , vol.46, no.6, pp.2392-2400, Nov.-Dec. 2010

[5] Vermaak, R.; Kamper, M.J.; , "Design of a novel air-cored permanent magnet linear generator for wave energy conversion," *Electrical Machines (ICEM), 2010 XIX International Conference on* , vol., no., pp.1-6, 6-8 Sept. 2010

[6] H. Polinder, M.A. Mueller, M. Scuotto and M. Goden de Sousa Prado, "Linear generator systems for wave energy conversion", Proceedings of the 7th European Wave and Tidal Energy Conference, Porto, Portugal.,2007

[7] Drew, B.; Plummer, A.R.; Sahinkaya, M.N.; , "A review of wave energy converter technology," *Proc.IMecheE, Part A:J.Power and energy* , vol.223, no., pp.887-901, June 2009

[8] Hodgins, N.; Keysan, O.; McDonald, A.S.; Mueller, M.A.; , "Design and testing of a linear generator for wave-energy applications," *Industrial Electronics, IEEE Transactions on* , vol.59, no.5, pp.2094-2103, May 2012

[9] Bracco, G.; Giorcelli, E.; Mattiazzo, G.; Marignetti, F.; Carbone, S.; Attaianese, C.; , "Design and experiments of linear tubular generators for the inertial sea wave energy converter," *Energy Conversion Congress and Exposition (ECCE), 2011 IEEE* , vol., no., pp.3864-3871, 17-22 Sept. 2011

[10] Oprea, C.A.; Martis, C.S.; Biro, K.A.; Jurca, F.N.; , "Design and testing of a four-sided permanent magnet linear generator prototype," *Electrical Machines (ICEM), 2010 XIX International Conference on* , vol., no., pp.1-6, 6-8 Sept. 2010

[11] Cheng-Tsung Liu; Hsin-Nan Lin; , "Development of a systematic scheme for direct driven slotless tubular linear generator design," *Power Electronics and Drive Systems, 2009. PEDS 2009. International Conference on* , vol., no., pp.843-847, 2-5 Nov. 2009

[12] Prado, M.; Polinder, H.; , "Direct drive in wave energy conversion — AWS full scale prototype case study," *Power and Energy Society General Meeting, 2011 IEEE* , vol., no., pp.1-7, 24-29 July 2011

[13] Pirisi, A.; Gruosso, G.; Zich, R.E.; , "Novel modeling design of three phase tubular permanent magnet linear generator for marine applications," *Power Engineering, Energy and Electrical Drives, 2009. POWERENG '09. International Conference on* , vol., no., pp.78-83, 18-20 March 2009

[14] Jiabin Wang; Weiya Wang; Atallah, K.; Howe, D., "Design of a linear permanent magnet motor for active vehicle suspension," *Electric Machines and Drives Conference, 2009. IEMDC '09. IEEE International* , vol., no., pp.585,591, 3-6 May 2009

Design of Matching Impedance for Ultra Wideband Partial Discharge Detection

Primas Emeraldi, Umar Khayam
School of Electrical Engineering and Informatics
Bandung Institute of Technology
Bandung, Indonesia
primas.emeraldi@gmail.com, umar@hv.ee.itb.ac.id

***Abstract*—Partial Discharge (PD) detection in the ultra wideband (UWB) at frequency from 100 kHz up to and above 1 GHz give some benefits especially for nature observation of PD pulse shape and frequency spectrum. One of the methods to measure the UWB PD signal is a method of impedance matched. Impedance of 50 ohm corresponding to the characteristic impedance of the coaxial cable and the internal impedance of oscilloscope is used as a coupling device to maximize the power transfer of PD current so that it can obtain real PD signals. This paper designs a matching impedance which composed from attenuator as detecting impedance and UWB amplifier as signal amplifier and evaluates the performance using S-parameter value. The matching impedance has good reflection loss below -10 dB over low frequencies up to 3 GHz frequency bandwidth, with a minimum value of S11 is -28.8 dB at a frequency of 322 MHz and -18.3 dB at 3 GHz. The input impedance and output impedance values are close to 50 ohm over the frequency bandwidth. The gain (S21) of the impedance matching circuit is 14 dB at a frequency close to DC and down to a 6 dB at a frequency of 100 MHz and have flat value of 6 dB up to 3 GHz frequency. From the simulation results, the designed matching impedance has frequency bandwidth from DC up to 3 GHz that can be implemented as a coupling device for UWB PD detection.**

Keywords— partial discharge; ultra wideband; coupling device; matching impedance

I. INTRODUCTION

Isolation is an important component of high voltage electrical power system equipment but often also regarded as the weakest part of the high voltage equipment which in most cases the main cause of failure is due to insulation degradation. Insulation degradation commonly associated with the activity of partial discharge (PD). The presence of PD can be considered as an indication of the insulation condition and also as a mechanism for further insulation damage, so among various insulation diagnostic method, PD method occupies an important position to evaluate and diagnose the isolation problem[1-3].

PD detection can be performed either conventionally by direct measurement of apparent charge according to the IEC60270 standard or unconventionally by using other indirect indicators of PD occurrence, which includes electromagnetic, acoustic, optical, and, chemical emission measurements. PD detection methods can also be categorized in terms of the ability of the measurement frequency bandwidth where the frequency of PD can range from a few Hz to GHz order. One such method is PD detection in UWB range which has a very wide measurement frequency bandwidth (from 100 kHz up to and beyond 1 GHz.). UWB PD detection method may provide some benefits, especially for nature observation PD pulse shape and frequency spectrum to be associated with the underlying nature of the PD process, and thus the results obtained may help in the identification of PD sources [4-7].

UWB PD detection method requires a different PD detector from conventional PD detector which has been commonly used according to IEC 60270, because it must operate at a very wide frequency range. One of method to measure the UWB PD signal is a method of impedance matched. Impedance of 50 Ω corresponding to the characteristic impedance of the coaxial cable and the internal impedance oscilloscope is used as a detection impedance to the maximize power transfer of PD current so that it can obtain real PD signals [8,9,10,11]. This paper designs a matching impedance as a coupling device for UWB PD detection that have frequency bandwidth of low frequency near DC to above 1 GHz.

II. UWB PD DETECTION

PD detection methods can also be categorized in terms of frequency bandwidth capabilities where the frequency of PD can range from a few Hz to GHz order. Fig. 1 shows the illustration of the frequency band option in detection of PDs relative to the full frequency spectrum of the PD original signal. Based on the detection frequency band option, PD detection methods are categorized as [4-5]: narrow-band: bandwidth of 9 kHz to 30 kHz with mid band frequency between 50 kHz and 1000 kHz, wide-band: signal frequency range bounded by lower limit of between 30 kHz and 100 kHz, upper limit of 500 kHz and detection bandwidth of between 100 kHz and 400 kHz and Ultra wideband: frequency range from 100 kHz up to and beyond 1 GHz.

Test circuit commonly used for PD measurement system according to the standard IEC 60270 is shown in Fig. 2, consists of a high voltage supply (U), noise filter (Z), Test object capacitance (Ca), Coupling Capacitance (Ck), Coupling Device (CD) , connecting cable (CC) and the Measuring Instrument (MI). High frequency current associated with PD

pulses flowing through the coupling capacitor and a coupling device.

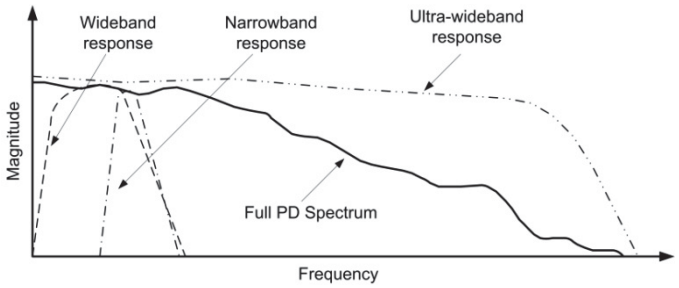


Fig. 1. Illustration of PD measurement systems frequency response [4-5].

Coupling device of conventional PD measurements according to IEC 60270 consists of detecting impedance as the main component that provides voltage pulses on the output that converted from the PD current on the input. Impedance detection is usually an RLC (resistor, inductor and capacitor) circuit that has a narrow measurement bandwidth.

Using measuring instrument such as oscilloscope that allows direct observation of pulses with a low repetition rate with a duration of less than 1 ns and a proper coupling device allows detection of PD in UWB range in a simple way with same test circuit configuration of IEC 60270. In addition, the amplifiers are available with a very wide bandwidth so that helps in the detection of PD UWB signal. Coupling device which is used for UWB PD detection is usually a 50 Ω resistance value without induction directly connected in high voltage test circuit and can be equipped with UWB amplifiers [9-11]. UWB PD measurement circuit configuration is shown in Fig. 3.

Detecting impedance circuit which has a very wide bandwidth can be designed in the form of matching impedance with input and output impedance of 50 Ω constant on a very wide frequency range. Actual PD signal can be obtained by the method of impedance matched, because the value 50 Ω impedance corresponding to the characteristic impedance of the coaxial cable and the internal impedance of oscilloscope.

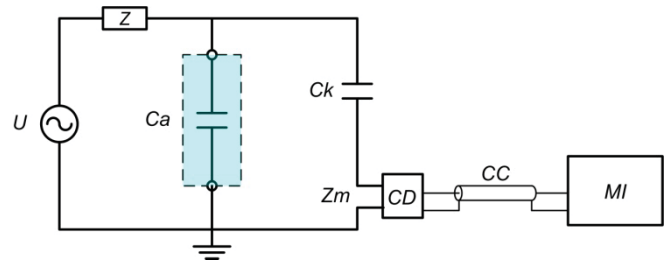


Fig 2. PD measurement circuit [4].

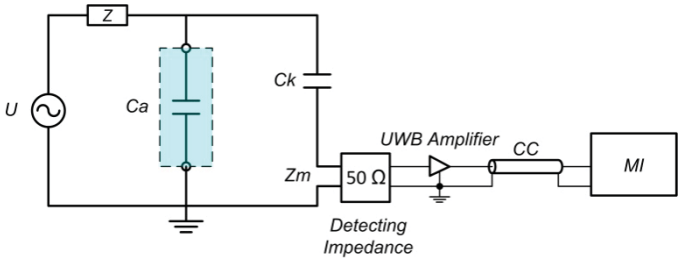


Fig. 3.UWB PD measurement circuit.

III. Design of Matching Impedance

Matching impedance for UWB PD detection composed of the detecting impedance and UWB amplifier as its main components. Detecting impedance is attenuator circuit instead of 50 Ω resistive component, because the attenuator circuit also serves as the impedance matching of the UWB amplifier. Attenuator circuit is placed on the input and output impedance amplifier circuit to get impedance value close to 50 Ω. The use of amplifiers so that small magnitude PD signal can still be detected obtains more PD signal. Signal from the input port will be detected by the impedance of attenuator circuit then amplified by UWB amplifier on a very wide frequency bandwidth so that it can detects large numbers of PD pulses. Scheme of matching impedance as coupling device for UWB PD detection is shown in Fig. 4.

A. Design Aspect

1. Attenuator Circuit

The main components of the designed matching impedance is attenuator circuit. Attenuator circuit is designed to provide input and output impedance of 50 Ω of matching impedance. One of attenuator circuit type is using π topology as shown in Fig. 5. Attenuator circuit is formed by the arrangement of three resistors that provide input and output impedance purely resistive with a certain attenuation value.

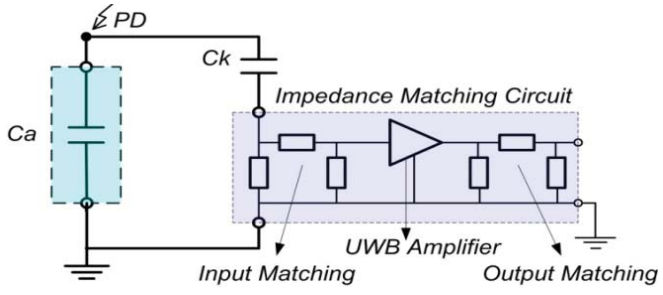


Fig. 4. Scheme of PD measurement coupling device using impedance matching circuit.

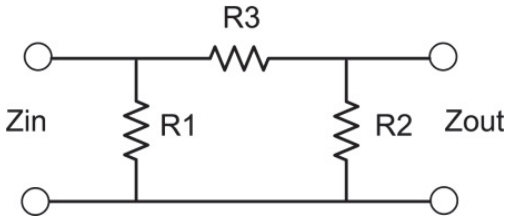


Fig. 5. Attenuator circuit using π topology [12]

In designing the attenuator circuit, there are two things that concern are the attenuation and impedance of inputs and outputs. Attenuation values is:

$$ATT = 10\log N \quad (1)$$

where ATT is the attenuation (in dB) and N is the ratio between the power source that absorbed by the circuit with power delivered to the load. Attenuator design begins with determining the value of the input impedance (Zin) and output impedance (Zout) desired, and given attenuation value then calculated resistor values to form the attenuator circuit. The equation for determining the resistance value of attenuator using π topology is as follows:

$$R_2 = \frac{1}{2(N-1)}\left(\frac{Z_{in}Z_{out}}{N}\right)^{1/2} \quad (2)$$

$$\frac{1}{R_1} = \frac{1}{Z_{in}}\frac{N+1}{N-1} - \frac{1}{R_2} \quad (3)$$

$$\frac{1}{R_3} = \frac{1}{Z_{out}}\frac{N+1}{N-1} - \frac{1}{R_2} \quad (4)$$

2. UWB Amplifier

UWB amplifiers used is MMIC (Monolithic Microwave Integrated Circuit). MMIC is a monolithic integrated circuit amplifier that operates from DC up to microwave frequency. MMIC devices which have the characteristic impedance of the input and output are constant close to 50 Ω at a wide frequency range (DC - several GHz) so suitable for use as an amplifier for impedance matching circuit. Internal structure of the MMIC is a Darlington transistor pair connected to the resistive feedback and a simple resistive biasing scheme. Biasing circuit of the MMIC amplifier is shown in Fig. 6 consists of biasing resistors (Rbias), blocking capacitors (Cblock1 and Cblock2) and radio frequency choke (RFC). Rbias is used to set the bias currents that determine the performance of the MMIC. The selection of Rbias based on the voltage (Vd) and current (Id) rating of MMIC components and amplifier bias voltage (Vcc) according the following equation [12-13]:

$$R_{bias} = \frac{V_{cc} - V_d}{I_d} \quad (5)$$

Cblock used to protect the input and output terminal of amplifier that blocking DC signal and determine the low frequency response of the amplifier. RFC used to increase the separation of the output RF signal from the amplifier DC supply as well as a "peaking coil" to increase the gain at high frequencies compensate the gain decrease at high frequency.

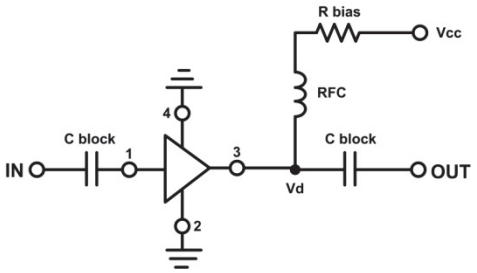


Fig. 6. Biasing circuit of MMIC amplifier [12-13].

3. S-Parameter

The matching impedance is a 2-port network that can be described as in Fig. 7, where a1 is the incident signal that transmits into the network from port 1 (the input port), b1 is the reflected signal from the network at port 1, a2 is the incident signal that transmits into the network at port 2 (the output port), and b2 is the reflected signal from the network at port 2.

If the incident signal coming through the port 1 (a1), the signal may be exiting from either port 1 itself (b1) or port 2 (b2). If port 2 is connected to a load with impedance identical to the system impedance (Z0) then, according to the maximum power transfer theorem, b2 will be totally absorbed making a2 equal to zero. S-parameters can be seen as a relationship of the case where:

$$S_{11} = \frac{b_1}{a_1}, S_{12} = \frac{b_1}{a_2}, S_{21} = \frac{b_2}{a_1}, S_{22} = \frac{b_2}{a_2} \quad (5)$$

Generally, the S-parameters of a 2-port network is defined as follows: S11 is the input reflection coefficient, S12 is the reverse transmission coefficient, S21 is the forward transmission coefficient, and S22 is the output reflection coefficient. Analysis of the matching impedance that can be considered as 2-port network can be performed using the S-parameter. Reflection losses (S11) design to have low value, which means there is a good impedance matching between the input and output impedance, so all the power is transmitted without any reflection. The minimum value of S11 is generally accepted is below -10 dB, which means that only 10% of the incident signal that reflected.

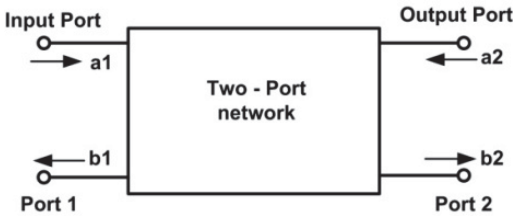


Fig. 7. Incident and reflected waves in a two-port netwok [14].

B. Modeling

The matching impedance consists of input matching circuit, MMIC amplifiers with biasing circuit and output matching circuits. Input and output matching circuits is a π topology attenuator with a value of R1 = R2 = 910 Ω and R3 = 5.6 Ω provide 1 dB attenuation that sufficiently low so does not affect the gain of the amplifier. The Model of MMIC amplifier components have a voltage rating Vd = 7.8 V and a rated current Id = 35 mA and amplifier supply voltage is Vcc = +12 V DC, so the bias resistor used is Rbias = 120 Ω.

The value of blocking capacitor (Cblock1 and Cblock2), and RFC component of amplifier biasing circuit will determine the response of the overall circuit of matching impedance [11-12]. Selection of the amplifier biasing circuit is component is based on S-parameter simulation results of the circuit model. Cblock and RFC component values were selected that provide input and output impedance that comes closest to 50 Ω, minimum input reflection loss (S11) and the maximum gain (S21) over desired bandwidth that suitable for PD measurement UWB. The desire bandwidth is from DC to 3 GHz that can be used to detect the signal PD from very low frequency (VLF) to ultra high frequency (UHF). Modeling of matching impedance is shown in Fig. 8.

C. Simulation Result

Performance of the matching impedance evaluates using S-parameter to see if it can work properly in certain frequency bandwidth. Reflection loss of the matching impedance refers to the value of S11 while the gain refers to the value of S21. Value of S11, S21 and input impedance (Zin) and output impedance (Zout) of simulation result determine the proper value Cblock and RFC components of the matching impedance to be used.

Fig. 9 shows the value of reflection loss of the matching impedance with variations blocking capacitors value. Small value of Cblock 1 pF give the value of S11 below -10 dB at frequencies above 2 GHz while bigger value of Cblock 1 mF give S11 above -10 dB that mean the matching impedance have poor reflection loss. Cblock 1 nF give the value of S11 below -10 dB at low frequency so that the matching impedance can work at low frequencies. Cblock 100 uF gives slightly better low frequency response than the value of Cblock 1 nF. So the proper Cblock component used in the matching impedance is 100 uF that give the optimum frequency response.

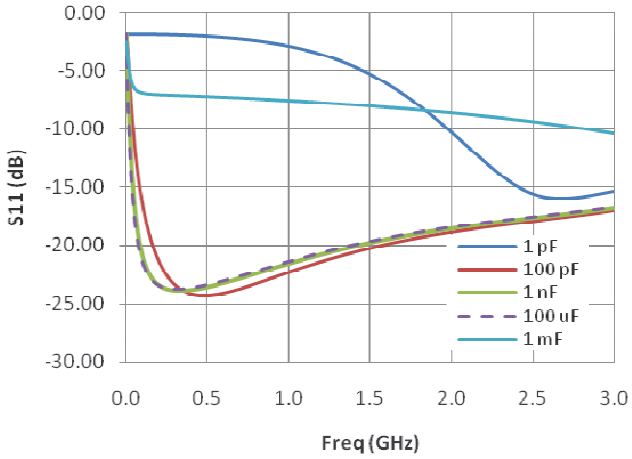


Fig. 9. Reflection Loss (S_{11}) with variation of Cblock value

Gain (S21), reflection loss (S11) and input impedance (Zin) and output impedance (Zout) values of the matching impedance simulation result with the variations of RFC value is shown in Fig. 10. From the simulation results it can be seen that the value of the RFC components affect the, gain (S21), reflection loss (S11) and input impedance (Zin) and output impedance (Zout) of matching impedance. The value of RFC 10 nH provide gain of 6 dB at frequencies above 1.5 GHz, that mean the circuit can not be operate in low frequency. RFC 100 nH giving 6 dB gain at frequencies above 100 MHz, but below that gain is smaller than 6 dB. While RFC 10 uH provides 6 dB gain at frequencies above 100 MHz and have flat gain up to 3 GHz and also have gain over 6 dB at frequencies below 100 MHz so that the matching impedance can operate at low frequency region. The value of RFC of 10 uH also gives the input impedance (Zin) and output impedance (Zout) that closest to 50 Ω where there is maximum power transfer to provide minimum value of reflection loss (S11) is -28.8 dB at 322 MHz and -18.3 dB at 3 GHz.

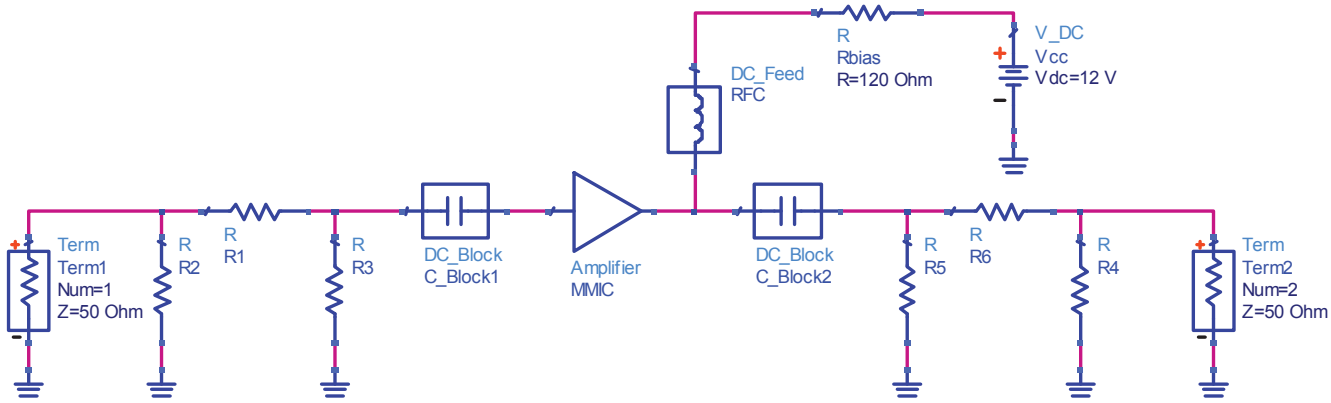
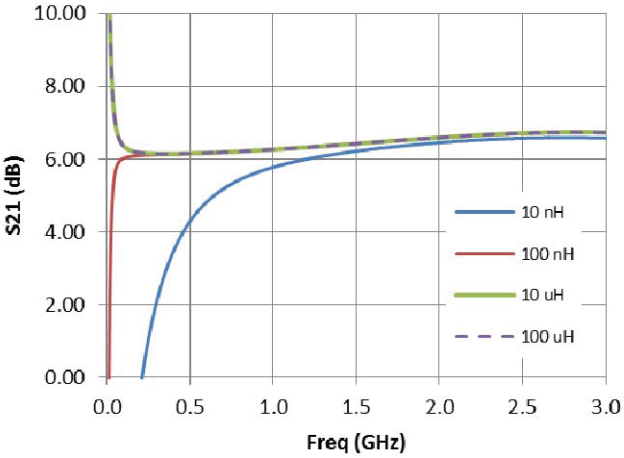
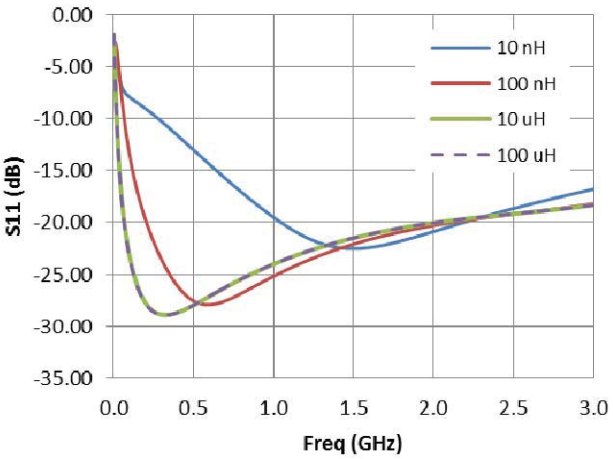


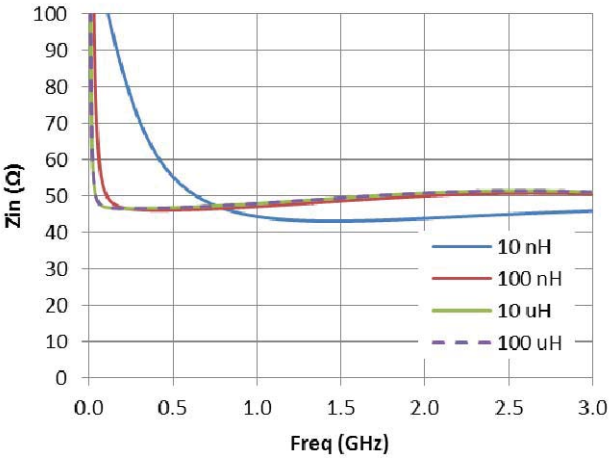
Fig. 8. Schematic of the designed matching impedance .



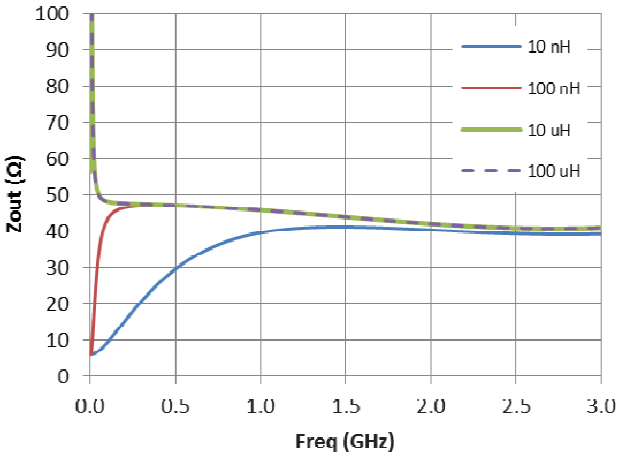
(a) Gain (S_{21})



(b) Reflection Loss (S_{11})



(c) Impedansi input (Z_{in})



(d) Impedansi input (Z_{out})

Fig. 10. Gain, Reflection Loss, Input and output impedance of the matching impedance with variation of RFC value

Bigger RFC value of 100 uH did not gives any significant improvement however gives bigger impedance value in high frequency that affect the total impedance of the circuit. So the optimum value of the RFC which gives the desired frequency response is 10 uH. The values of reflection loss, gain and input impedance and output impedance from simulation results show that the designed matching impedance has frequency bandwidth from very low frequencies near DC up to 3 GHz.

IV. CONCLUSION

This paper designs a matching impedance to obtain the frequency response appropriate for PD detection in UWB (100 kHz up to and beyond 1 GHz). The performance of the matching impedance evaluates using S-parameter value. From the reflection loss (S11), gain (S21), input impedance (Zin) and output impedance (Zout) values of the simulation result, the most desirable frequency response of the impedance matching circuit is given by the value of component Cblock of 1 nF and RFC of 10 uH.

The designed matching impedance has good reflection loss below -10 dB over low frequencies up to 3 GHz frequency bandwidth, with a minimum value of S11 is -28.8 dB at a frequency of 322 MHz and -18.3 dB at 3 GHz. The good reflection loss at wide bandwidth is also seen in the input impedance (Zin) and output impedance (Zout) values which is close to 50 Ω over frequency bandwidth. The gain (S21) of matching impedance is 14 dB at a frequency close to DC and down to a 6 dB at a frequency of 100 MHz and have flat value of 6 dB up to 3 GHz frequency. From the simulation results, designed matching impedance has frequency bandwidth from DC up to 3 GHz that can implemented as a coupling device for UWB PD detection. Once implemented, the designed matching impedance needs to be tested using a network analyzer is used on PD detection in the laboratory and in the field.

REFERENCES

[1] Han, Y., Song, Y. H, "Condition monitoring techniques for electrical equipment: a literature survey, Power Engineering Review", IEEE Volume: 22, Issue: 9, pp. 59, 2002.

[2] Montanari, G. C, "Partial discharge measurements: becoming a fundamental tool for quality control and risk assessment of electrical equipment?", Conference Record of the 2006 IEEE International Symposium on Electrical Insulation", pp. 281-284, 2006.

[3] Bartnikas. R, " A comment concerning the rise times of partial discharge pulses", IEEE Transactions on Dielectrics and Electriacal Insulation" , Vol.12 (No.2), pp. 197-202. 2004.

[4] IEC-60270, High voltage test techniques - Partial discharge measurements, 2000.

[5] IEEE std 400.3, "IEEE Guide for Partial Discharge Testing of Shielded Power Cable Systems in a Field Environment", New York: IEEE Power Engineering Society, 2007.

[6] Muhr, M., Schwarz, R., Pack, S., Koerbler, B., "Unconventional partial discharge measurement [electrical insulation evaluation]", Electrical Insulation and Dielectric Phenomena, 2004. CEIDP '04. 2004 Annual Report Conference on, pp. 430-433, 2004.

[7] Czaszejko, T., Chen., S, "Ultra Wide Band Response of Partial Discharge Test Circuit", International Conference on High Voltage Engineering and Application (ICHVE), pp. 61-64, 2010

[8] Lemke, E., Strehl, T., & Markalous, S, "Ultra-wide-band PD Diagnostics of Power Cable Terminations in Service. IEEE Transactions on Dielctrics and Electrical Insulation", Vol. 15 (No. 6), pp. 1570-1575. 2008.

[9] C. Yonghong, C. Xiaolin, Y. Lidong, J. Yan, X. Hengkun, "Comparative study of two kinds of ultra-wideband partial discharge detection techniques "Electrical Insulating Materials, 2001. (ISEIM 2001). Proceedings of 2001 International Symposium on Digital Object Identifier, pp. 258-261, 2001.

[10] G. C. Stone, H. G. Sedding, N. Fujimoto, and J. M Braun, "Practical Implementation of Ultrawideband Partial Discharge Detectors", IEEE Transactions on Electrical Insulation Vol. 27 No. 1, February 1992.

[11] Alain F. S. Levy, Orsino B, Oliveira Filho, Eric 0. Forster, "A Study of Partial Discharges in Air with An Ultra-wideband Signal Detector", Proceeding of the 4th International Conference on Properties and Applications of Dielectric Materials July 3-8,1994.

[12] Carr, J, "Secret of RF Circuit Design", The McGraw-Hill, 2004.

[13] Aplication Note AN-S003, "Biasing MSA Series RF Integrated Circuits", Agilent Technologies, Inc, 1999.

[14] Gonzalez, G, " Microwave Transistor Amplifier Analysis and Design",Prentice Hall 1996.

Design of New Shape Printed Bowtie Antena for Ultra High Frequency Partial Discharge Sensor in Gas-Insulated Substations

Hanalde Andre, Umar Khayam
School of Electrical Engineering and Informatic
Bandung Institute of Technology
Bandung, Indonesia
Hanalde.andre@gmail.com, umar@hv.ee.itb.ac.id

***Abstract*—Detection of partial discharge (PD) with an antenna as ultra high frequency (UHF) sensor at a frequency of 300 MHz - 3 GHz has been widely used and proven to be an effectife away for diagnosis of insulation in gas insulation switchgear (GIS). Various types of antennas have been developed to obtain better sensitivity and accuracy in the UHF bandwidth range. Bowtie antenna can be used as a sensor because it has the characteristics of ultra wide band (UWB). Sensitivity antenna greatly affects the ability of the sensor to detect the electromagnetic signals (EM) from source PD. Return loss (RL) used as a parameter to see the sensitivity of the antenna because it has a better resolution to see the reflected signal. This study new shape bowtie antenna to get better sensitivity. Influence edge modification and sliced modification will be discussed and simulated in antenna design. Simulation results show modification edge and sliced bowtie antenna able to produce optimal sensitivity antenna with a radius of 18 mm and 16 mm. The technique led to the reduction of the antenna surface significantly.**

Keywords*—*Bowtie Antena; Ultra High Frequency (UHF); Partial Discharge (PD); Gas Insulation Switchgear (GIS)

I. INTRODUCTION

Partial discharge (PD) measurement can be an effective diagnostic tool to determine the condition of insulation of high voltage equipment. Propagation of electromagnetic (EM) waves generated by partial discharge (PD) activity can be detected by using sensors. This technique is widely used in gas insulation switchgear (GIS) using an antenna sensor and has been demonstrated by measuring both in the laboratory and in the field [1-3]. PD characteristics that occur in GIS produces EM wave signal in the range of ultra high frequency (UHF) which occurred at a frequency of 300 MHz to 3 GHz. UHF printed antenna can be used as a sensor in the PD detection by this method.

The advantage of the use of an antenna is that it can be done online PD measurements. Things to consider in the use of an antenna as a sensor is a kind of signal that it is detected, in which case relation with bandwidth. One of important parameter in the design of the antenna is S_{11} or RL[1]. Optimization of the antenna can be done by taking into account the parameters required for the PD sensor [2]. At GIS, the use of UHF sensor allow high sensitivity and fast response compared to the use of gas sensors. [3] Another advantage of the measurement using UHF method is its resistance to interference [8].

Bowtie antenna has a characteristic Ultra wide band (UWB) [4, 7,9, 10]. This characteristic is very advantageous for use in UHF sensor. Several studies regarding the bowtie antenna design provides parameters that can be used reference to obtain the optimal sensor performance [9, 10]. Antenna geometry shape affect the frequency response characteristics. The use of simulation is very helpful for designing the antenna in accordance with the required characteristics.

In this paper, new shape bowtie antenna will be studied in accordance with the characteristics of UHF sensor using the method of finite element method (FEM). Section 2 discusses the bowtie antenna parameter and simulated to see the effect on frequency response. Modifications antenna will study and simulated to obtain the optimal shape is discussed in section 3. Current distribution on the surface of the antenna will be considered in the design.

II. PARAMETER BOWTIE ANTENNA

Bowtie antenna using FR4 epoxy substrate ($\varepsilon_r = 4.4$) with a thickness of 1.6 mm. Type of substrate is widely used as a Print Circuit Board (PCB). It is choosen to get a sensor with the low budget. Both ends of the antenna associated with a transmission line that has a 50 ohm impedance. Adjustable impedance to the impedance of the coaxial cable is connected to the oscilloscope.

Using FEM simulation method in three-dimensional (3-D. This technique use small elements to represent complex geometric shapes. The number of elements determine to obtain accurate simulation results. Antenna is placed on the XY plane in the form of 2-dimensional (2-D) perfect electric boundary. The thickness of the antenna material is not taken into account. Antenna placed in the radiation boundary using air material ($\varepsilon_r = 1$) accordance with the conditions in the field at a distance of 250 mm in accordance with quarter length 300 MHz frequency which is the lowest frequency in the UHF range.

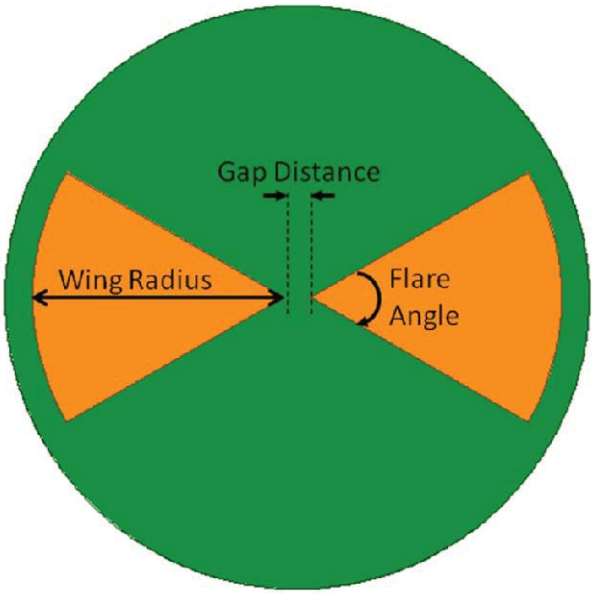


Fig. 1. Bowtie Antena

Fig. 1 can be seen the bowtie antenna which has several parameters. Wing radius, flare angle and gap distance will be further discussed and simulated.

A. Wing Radius

Wing radius parameter is a component that affects the length of the Antena. Changes to the wing radius causes changes to the resonance frequency of the antenna. In Fig. 2 can be seen in the graph of RL on the effect of changes in wing length bowtie antenna radius. Resonance frequency for wing length radius 12mm, 24 mm, 36 mm and 48 mm are 2.97 GHz, 1.71 GHz, 1.25 GHz, 0.95 GHz.

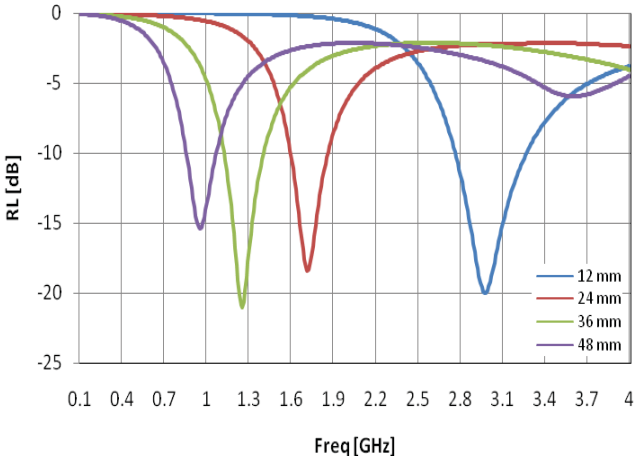


Fig. 2. Simulation Results Return Loss $|S_{11}|$ Vs. Frequency [Wing Radius Comparison]

B. Flare Angle

Flare angle is the angle of the tip of the arm lies bowtie antenna. The greater flare angle causes increased return loss values. Simulation results obtained with an angle of 60 degrees produces a minimum value of -21.02 dB. RL is slightly higher than a 30 degree angle, -21.73 dB. Below -6 dB bandwidth to be obtained for the flare angle of 60 degrees has an 511 MHz bandwidth in the range of 1.04 - 1.65 GHz . RL -6 dB shows the received power of the antenna half power comes. Bandwidth for the flare angle with an angle of 30, 90 and 120 degrees are 291, 406 and 379 MHz. Based on the simulation results obtained flare angle 60 degree angle has to be the optimum value.

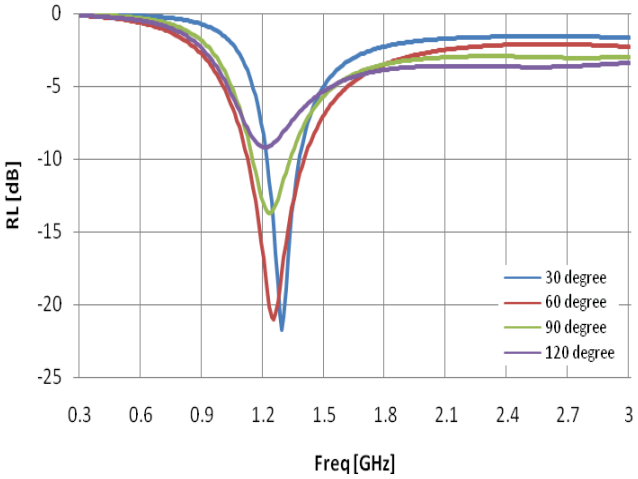


Fig. 3. Simulation Results Return Loss $|S_{11}|$ Vs. Frequency [Flare Angle Comparison]

C. Gap Distance

Antenna length be affected by gap distance. Increasing the value gap distance can affect the resonant frequency of the antenna. Simulations performed with gap dictance variation distance.. simulation results can be seen in Fig. 4. Resonance frequency for gap distance 2 mm, 4 mm, 8 mm, and 16 mm are 1.29 GHz, 1.25 GHz, 1.15 GHz, and 1.03 GHz.

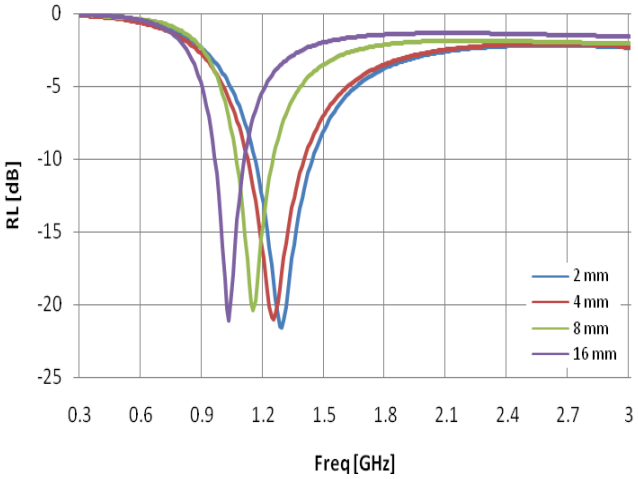


Fig. 4. Simulation Results Return Loss $|S_{11}|$ Vs. Frequency [Gap Distance Comparison]

Effective length of the antenna is a combination of wing radius with distance gap that has ties to the wavelength of an electromagnetic signal.

$$\lambda = c/f \quad (1)$$

Wavelength to frequency relationship can be seen in Equation 1. λ is the wavelength, c is the speed of light (300,000,000 m/s) and f is the frequency of the electromagnetic wave signal.

III. MODIFICATION BOWTIE ANTENNA

Antenna pattern affects the response characteristics of the frequency response. This section will discuss the modification of the bowtie antenna. antenna which has a modified wing radius of 36 mm, flare angle of 60 degrees, 4 mm gap distance. surface substrate circular with a diameter of 80 mm.

A. Edge Modification Bowtie Antena

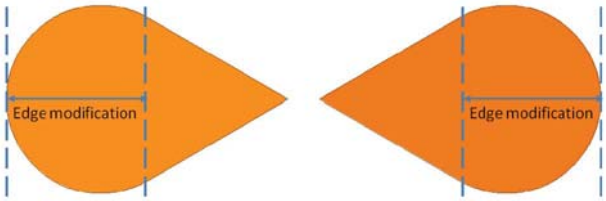
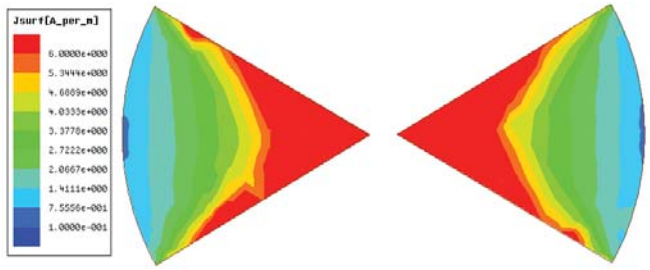
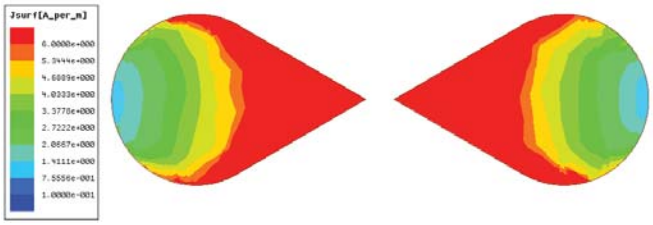


Fig. 5. Edge Modification Bowtie Antena

Tests have been done to change the tip shape of the bowtie antenna [9,11]. The result shows the changes in the form of a curved end of the bowtie antenna can increase the signal can be received by an antenna. In designing the simulation is performed to obtain the optimal shape for the antenna to be made. In Fig. 5 can be seen form of edge modification bowtie antenna.



(a)



(b)

Fig. 6. Current Density Distribution at Freq 1.3 GHz, phase 0 deg
(a) Bowtie Antenna, (b) Edge Modification Bowtie Antenna

Modified edge bowtie antenna surface area leads to a lack antenna. for the bowtie antenna has a surface area of 1.357 mm^2, while the antenna has been modified with a length of 18 mm has a surface area of 1.102 mm^2. With this reduction in RL value obtained is also reduced. In Fig. 6 it can be seen the current distribution of the antenna. Curved shape at the end of the current distribution of the antenna causes a meeting in the area. Modified antenna pattern can be seen in Fig. 7.

From the simulation results obtained in the form of the response RL graph as shown in Fig. 8. Wing to the antenna with a radius of 36 mm and a gap distance of 4 mm edge length modification obtained a maximum value of 18 mm with a minimum RL value of -28.28 dB at frequencies 1.33 GHz. values are decreased by 6 dB of the bowtie antenna.

Resonance frequency shift occurs with increasing edge length modification. Resonance frequency shift for 10 and 18 mm long by 40 MHz. Changes the capacitance of the antenna can affect the resonance frequency. Modified edge bowtie antenna surface area leads to a lack antenna. for the bowtie antenna has a surface area of 1.357 mm^2, while the antenna has been modified with a length of 18 mm has a surface area of 1.102 mm^2. With this reduction in RL value obtained is also reduced. In Fig. 6, it can be seen the current distribution of the antenna. Curved shape at the end of the current distribution of the antenna causes a meeting in the area. Modified form simulated antenna can be seen in Fig. 7.

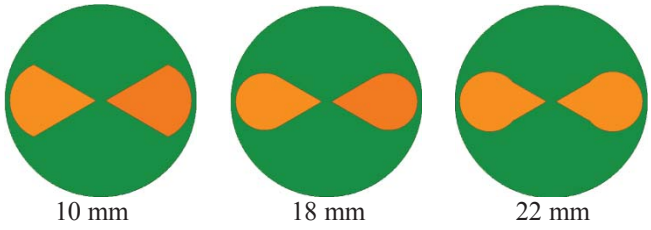


Fig. 7. Edge Modification Bowtie Antena

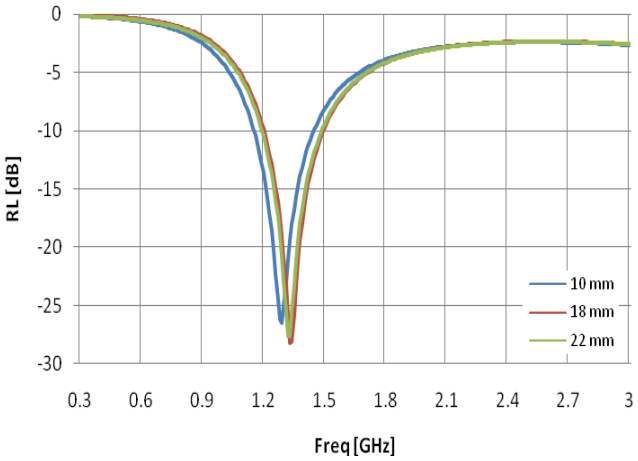


Fig. 8. Simulation Results Return Loss $|S_{11}|$ Vs. Frequency
[edge modification antena bowtie comparison]

From the simulation results obtained in the form of the frequency response RL graph as shown in Fig. 8. Wing radius

antenna with a radius of 36 mm and a gap distance of 4 mm edge length modification obtained a maximum value of 18 mm with a minimum RL value of -28.28 dB at frequencies 1.33 GHz. values are decreased by 6 dB of the bowtie antenna.

Resonance frequency shift occurs with increasing edge length modification. Resonance frequency shift for 10 and 18 mm long by 40 MHz. Changes the capacitance of the antenna can affect the resonance frequency.

B. Sliced Edge Modification Bowtie Antena

Modified antenna is an advanced modification of the edge length of 18 mm. Shape the current distribution on the surface is uneven bowtie antenna. The end of the antenna is connected to transmission line has a higher current value compared with other parts of the antenna.

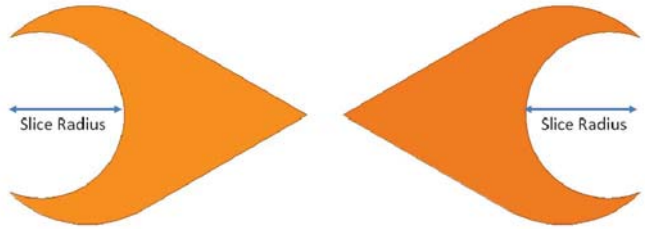


Fig. 9. Sliced Edge Modification Bowtie Antena

Current distribution is not uniform on the surface of the antenna was simulated to discard parts that have a low current density. This is done with consideration to be able to lower the los s return value is generated. The antenna by cutting part of the antenna modification with low currents is shown in Fig. 9. Area of surface modification with sliced antenna with a radius of 16 mm to 638 mm^2.

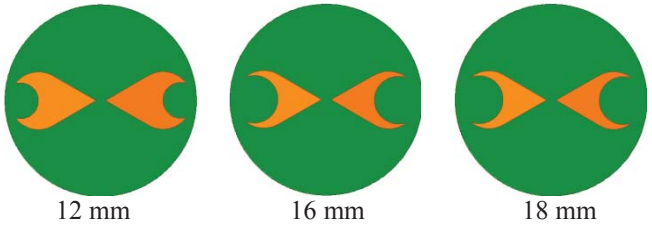
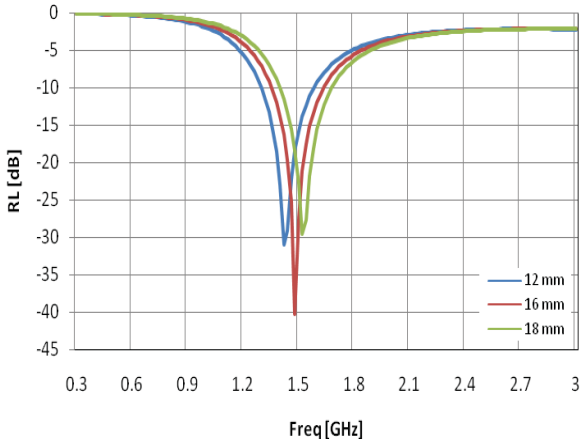


Fig. 10. Sliced Bowtie Antena Variation

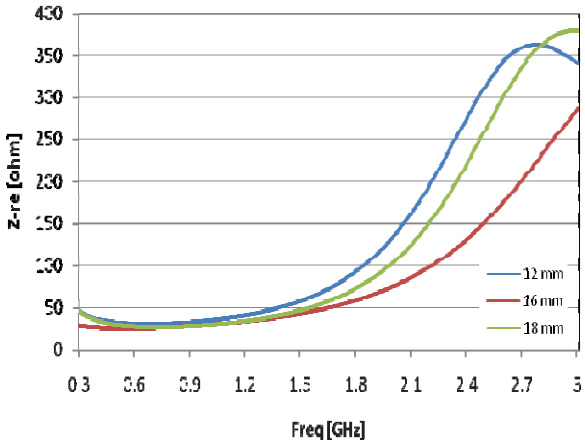
Effect of cutting the outer side of the antenna was simulated using three different antenna shapes. In fig. 10 can be seen form of sliced edge Modification bowtie antenna. Variations performed by differentiating radius cut of 12 mm, 16 mm and 18 mm. by reducing the surface area of the antenna can reduce losses that occur. Compatibility of the line impedance can also reduce the value of return loss.

Sliced edge modification bowtie antena have been simulated. The minimum value is obtained at a distance of 16 mm was -40.27 dB at 1.48 fekuensi GHz. The distance of modification 12 and 18 mm has a minimum RL at 1.42 and 1.52 GHz frequency with RL values of -31.04 and -29.5 dB.

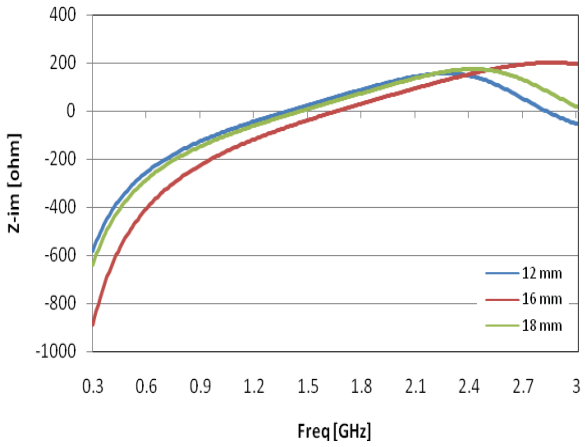
Antenna impedance values change according to the shape of the antenna geometry. Imaginary part of the impedance input seen dominant at low frequencies while the real part dominant at high frequencies. In the fig. 11 (a) and Fig. 11 (b) can be seen graphs the impedance of the antenna that has modified.



(a) Return Loss



(b) Real part of Impedance Input



(c) Imaginary part of Impedance Input

Fig. 11. Influence Modification Sliced Bowtie Antenna

C. Fabrication Final Antenna

The antenna has been printed using specifications that have been discussed in the beginning. Bowtie antenna design has been modified by edge modification by 18 mm and 16 mm sliced modification. Tests using a network analyzer can be done to get the frequency response characteristics of the antenna. Furthermore antenna can be tested for PD measurements in the laboratory and in the field.

Fig. 12. Antena Fabrication

IV. CONCLUSION

New shape bowtie antenna was discussed to get better sensor sensitifity. Simulation results using FEM method shows increased sensitivity antenna from original bowtie with minimum RL value -21.02 dB to -40.27 dB at the edge modification distance 18 mm and sliced bowtie antenna with a radius of 16 mm. Resonance frequency shift from the initial resonance 1.25 to 1.48 GHz GHz. Antenna surface area is reduced to 638 mm^2 of the initial shape with an area of 1357 mm^2. Shape The optimum geometry of the bowtie antenna obtained with considering current density distribution at the surface of the antenna. Antenna fabrication has been completed. In the further, antenna measurements were taken using a network analyzer and PD measurements performed in the laboratory and in the field.

REFERENCES

[1] G. Robles, M. Sánchez-Fernández, R. Albarracín Sánchez, M.V Rojas-Moreno, E. Rajo-Iglesias, J.M Martínez-Tarifa, "Antena parametrization for the detection of partial discharges", IEEE Transactions on instrumentation and measurement, vol. 62, no. 5, pp. 932-940, may 2013

[2] Tianhui Li, Mingzhe Rong, Chen Zheng, Xiaohua Wang, "Development Simulation and Experiment Study on UHF Partial Discharge Sensor in GIS", IEEE Transactions on Dielectrics and Electrical Insulation vol. 19, no. 4, pp.1421-1430, August 2012

[3] T. Hoshino, S. Maruyama, S. Ohtsuka, M. Hikita, J. Wada, S. Okabe, "Sensitivity Comparison of Disc- and Loop-type Sensors using the UHF Method to Detect Partial Discharges in GIS", IEEE Transactions on Dielectrics and Electrical Insulation vol. 19, issue. 3, pp.1544-1552, June 2012

[4] Y. Su, L. Xing, J. Ding, C.J Guo, "Design of planar antena with bowtie shape for ultra-wideband (UWB) applications", China-Japan Joint Microwave Conference Proceedings (CJMW), pp. 1-3, April 2011

[5] Yoshikazu Shibuya, Satoshi Matsumoto, Masayoshi Tanaka, Hirotaka Muto and Yoshiharu Kaneda, "Electromagnetic waves from partial discharges and their detection using patch antena", IEEE Transactions on Dielectrics and Electrical Insulation vol. 17, issue. 3, pp.862-871, June 2010

[6] S. Kaneko, S. Okabe, M. Yoshimura, H. Muto, C. Nishida, M. Kamei, "Detecting Characteristics of Various Type Antenas on Partial Discharge Electromagnetic Wave Radiating through Insulating Spacer in Gas Insulated Switchgear", IEEE Transactions on Dielectrics and Electrical Insulation vol. 16, issue. 5, pp.1462-1472, October 2009

[7] S. Raut, A. Petosa, "A Compact Printed Bowtie Antena for Ultra-Wideband Applications", European Microwave Conference(EuMC), pp. 81-84, October 2009

[8] S. Tenbohlen, D. Denissov, S. M. Hoek, S. M. Markalous, "Partial Discharge Measurement in the Ultra High Frequency (UHF) Range", IEEE Transactions on Dielectrics and Electrical Insulation vol. 15, issue. 6, pp.1544-1552, December 2008

[9] S.W. Qu, C.L. Ruan, "Effect Of Round Corners On Bowtie Antenas", Progress In Electromagnetics Research, PIER 57, pp. 179–195, 2006

[10] S. Qu, C. Ruan, "Quadrate Bowtie Antena With Round Corners", IEEE International Conference Ultra-Wideband (ICU), September 2005

Dynamic DC Optimal Power Flow Using Quadratic Programming

Rony Seto Wibowo, Nursidi, IGN Satriyadi H, DF Uman P, Adi Soeprijanto, Ontoseno Penangsang
Institut Teknologi Sepuluh Nopember, Indonesia
E-mail : ronyseto@ee.its.ac.id, satriadi@ee.its.ac.id

Abstract –This paper proposes a quadratic programming for solving the dynamic direct current optimal power flow (DDCOPF). The DDCOPF solves OPF with multi load levels in which ramp rate of committed units become coupling between two series load levels. To overcome this problem, a very large matrix may be required. The more number of load levels are considered, the larger matrix will be used. Consequently, it may take long computation time to solve. Therefore, the DC load flow is preferable than AC load flow. To show the effectiveness of the proposed approach, IEEE 14 bus test system is used. In addition, application of the proposed approach to real system Jawa Bali 500 kV 25-bus is presented.

Key words – DC load flow, dynamic optimal power flow, ramp rate, quadratic programming.

I. INTRODUCTION

Economic dispatch is a crucial part in operating power system with minimum operation cost. However, the result of economic dispatch may not be applied in the real operation as ramp rate violations may exist. Consequently, ramp rate should be considered in the formulation of economic dispatch

In 2001, X.S. Han, H.B. Gooi dan Daniel S. Kirschen proposed two methods to overcome dynamic economic dispatch (DED) [1]. The first method uses "Look A Head" approach to ensure that the result of ED does not violate ramp rate. The second method is quadratic programming (QP). Then, in 2012 "Look A Head" approach to determine feasible region was improved by Naoto Yorino and Mohd. Hafiz Habibbuddin in [2]. The improvement is conducted mainly in speeding up the computation time. Fast computation is very important to deal with real time load demand fluctuation. In addition, penetration of renewable energy resources may cause fluctuative power supply depending on real time wheather condition and this resources are generally modeled as negatif load.

All DED method mentioned above neglect voltage and transmission line capacity constraints. To incorporate these constraints, Dynamic Optimal Power Flow (DOPF) method was proposed. This method is similar to DED with network constraints. In 2001, K. Xie dan Y.H. Yong proposed interior point method to solve DOPF [3]. In 2011, paper [4] was improved by C.Y. Chung, Wei yan dan Fang Liu by applying decomposed predictor-corrector interior point to [3]. By using decomposition method, calculation process can be done parallelly. Thus, it speed up the computation time. Paper [5] presents the utilization of Modified Honey Bee Mating Optimisation (MHBMO) to solve DOPF. The result then compares to other metaheuristic approaches such as Honey Bee Mating Optimisation (HBMO), Genetic Algorithm (GA) dan Particle Swarm Optimization (PSO). Moreover, the proposed DOPF considers non-smooth generation cost function.

This paper presents quadratic programming for dynamic DC optimal power flow. Thus, this paper improves [1] by considering transmission line capacity constraint. To show the effectiveness of the proposed approach, IEEE 14 bus system is used. Moreover, Jawa Bali 500 kV 25 bus system will be utilized to show the applicability of the proposed approach in the real system.

II. DYNAMIC DC OPTIMAL POWER FLOW

The objective function of DDCOPF is to minimize generator cost along specified period of time consiting of several time intervals while satisfying transmission line capacity and ramp rate of each generator unit. The formulation of DDCOPF can be written as follow :

Objective function :

minimize $C = \sum_{i=1}^{T} \sum_{i=1}^{N} F_i(P_{it})$ (1)

$$F_i(P_i) = a_i P_i^2 + b_i P_i + c_i \quad (2)$$

Where T is number of time intervals, N is number of generators, P_i is active power output of generator i and a_i, b_i and c_i are cost coefficient of i-th generator.

Subject to :

(1) Active power balance constraint ,

$$P_{L\,k}^{t} - P_{G\,k}^{t} + P_{IN\,k}^{t} = 0 \quad (3)$$

$$P_{IN\,k}^{t} = \sum F_{km}^{t} = \sum \frac{1}{x_{km}} [\delta_k - \delta_m] \quad (4)$$

Where, $P_{L\,k}^{t}$ is power demands at bus k at times t, $P_{G\,k}^{t}$ is power generation at bus k at times t, and $P_{IN\,k}^{t}$ is the summation of power flows at transmission lines connected to bus k at time t.

(2) Transmission line capacity constraint

$$|F_{km}^{t}| \leq F_{km}^{U} \quad (5)$$

Where F_{km}^{t} and F_{km}^{U} is the power flow via branch km at time t and capacity of branch km, respectively.

(3) Generation constraint

$$P_{Gi}^{L} \le P_{Gi} \le P_{Gi}^{U} \quad (6)$$

P_{Gi}^{L} dan P_{Gi}^{U} is minimum and maximum power generation of the *i*-th generator, respectively.

(4) Ramp rate constraint

$$-\delta_i \le P_i^{t+1} - P_i^{t} \le \delta_i \quad (7)$$

$\delta_i$is ramp rateof the *i*-th generator. In MW/h

III. Application of Quadratic Programming to Dynamic DC Optimal Power Flow

Quadratic programming (QP) [6] is used to optimize the dynamic DC optimal power flow by expressing objective function as follow:

$$F_i(P_i) = \frac{1}{2}x^T Hx + G^T x \quad (8)$$

subject to

$$lb \le Ax \le ub \quad (9)$$
$$x_{min} \le x \le x_{max} \quad (10)$$

Where x is a vector consisting of several control variables i.e. voltage angles δ and active power generation P_G. This vector also covers variables of all load levels. For simplicity, the application of QP will be illustrated using 3 bus system with two load levels.

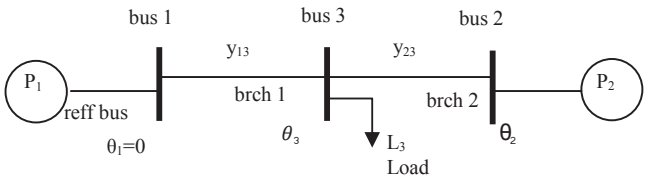


Figure 1. Simple 3 bus system

Matrix H is a diagonal matrix developed using quadratic coefficient of generation cost function a_i of all generators and can be written as:

$$a = [a_1 \quad \cdots \quad a_{n_g}] \quad (11)$$

Matrix a is then manipulated by inserting [1 x nb] zero matrix, where nb is number of buses. It becomes matrix aa.

$$aa = [0 \quad 0 \quad 0 \quad a_1 \quad a_2] \quad (12)$$

Then build H square matrix from aa. Size of H is [(nxt)x(nxt)], where n = ng + nb. ng is number generators while t is number of time intervals.

$$H = 2 * \mathrm{diag}(aa) \quad (13)$$

$$H = 2 * \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{21} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{12} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{22} \end{bmatrix} \quad (14)$$

Matrix G consists of linear coefficient of cost function $b_i$of all generators. Then, it is manipulated into [1 x (nxt)] matrix.

$$G = [0 \quad 0 \quad 0 \quad b_{11} \quad b_{21} \quad 0 \quad 0 \quad 0 \quad b_{12} \quad b_{22}]^T \quad (15)$$

Matrix A comprises of equality constraints A_{eq} and inequality constraints A_{ineq}. A_{eq} is for power balance while A_{ineq} is for transmission line capacity constraint. A_{eq} is [nb x n] matrix and A_{ineq} is [2*nbr x n] matrix.

$$A = \begin{bmatrix} A_{eq} \\ A_{ineq} \end{bmatrix} \quad (16)$$

For two load levels, there will be two matrixes A and constituted diagonally as in equation (19). In addition, matrix representing ramp rate limit in equations (17) and (18) should be inserted.

$$-\delta_1 \le P_{12} - P_{11} \le \delta_1 \quad (17)$$
$$-\delta_2 \le P_{22} - P_{21} \le \delta_2 \quad (18)$$

$$A * x = \begin{bmatrix} y_{11} & -y_{11} & -y_{13} & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -y_{21} & y_{22} & y_{23} & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -y_{31} & -y_{32} & -y_{33} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ y_{13} & 0 & -y_{13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & y_{23} & -y_{23} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -y_{13} & 0 & y_{13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -y_{23} & -y_{23} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & y_{11} & -y_{11} & -y_{13} & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -y_{21} & y_{22} & y_{23} & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & -y_{31} & -y_{32} & -y_{33} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & y_{13} & 0 & -y_{13} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & y_{23} & -y_{23} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -y_{13} & 0 & y_{13} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -y_{23} & -y_{23} & 0 & 0 \end{bmatrix} * \begin{bmatrix} \theta_{11} \\ \theta_{21} \\ \theta_{31} \\ P_{11} \\ P_{21} \\ \theta_{11} \\ \theta_{21} \\ \theta_{31} \\ P_{11} \\ P_{21} \end{bmatrix} \quad (19)$$

Vector of lower bound and upper bound of variables are denoted by lb and ub, respectively. Size of lb and ub are [nb+2*nbr+ng : 1] containing power demands, branch capacity and ramp rate limits. For 3 bus system, lb and ub are written as (20)

$$lb = \begin{bmatrix} P_{L1\,1} \\ P_{L2\,1} \\ P_{L3\,1} \\ -inf \\ -inf \\ -inf \\ -inf \\ -\delta_1 \\ -\delta_2 \\ P_{L1\,2} \\ P_{L2\,2} \\ P_{L3\,2} \\ -inf \\ -inf \\ -inf \\ -inf \end{bmatrix} \quad ub = \begin{bmatrix} P_{L1\,1} \\ P_{L2\,1} \\ P_{L3\,1} \\ F_{13}^{max} \\ F_{23}^{max} \\ F_{31}^{max} \\ F_{23}^{max} \\ \delta_1 \\ \delta_2 \\ P_{L1\,2} \\ P_{L2\,2} \\ P_{L3\,2} \\ F_{13}^{max} \\ F_{23}^{max} \\ F_{31}^{max} \\ F_{23}^{max} \end{bmatrix} \quad \begin{matrix} \} \text{ Power demand at } 1^{st} \text{ level} \\ \\ \\ \} \text{ Capacity of branches} \\ \\ \\ \\ \} \text{ Ramp rate} \\ \\ \} \text{ Power demand at } 2^{nd} \text{ level} \\ \\ \\ \} \text{ Capacity of branches} \\ \\ \\ \end{matrix} \quad (20)$$

IV. Simulation Results and Analysis

A. IEEE 14 bus System

Load profiles for this simulation are given in figure 2

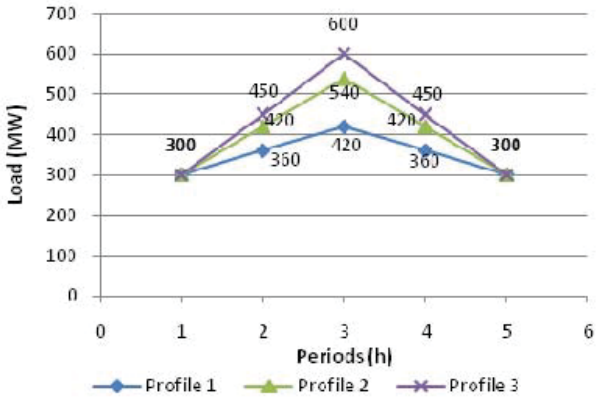


Figure 2. Load profiles used for IEEE 14 bus

Table 1
Generators data of IEEE 14 bus test system

Unit	Generation Cost function	Min output (MW)	Max output (MW)	Ramp (MW/ hour)
Unit 1	$0.043P_1^2 + 20P_1$	0	332.4	45
Unit 2	$0.250P_1^2 + 20P_1$	0	141	40
Unit 3	$0.010P_1^2 + 40P_1$	0	100	36
Unit 4	$0.010P_1^2 + 40P_1$	0	100	36
Unit 5	$0.010P_1^2 + 40P_1$	0	100	36

Table 2
Active power output by DCOPF for load profile 1

Generator	Power output (MW) at time t				
	1	2	3	4	5
Unit 1	225.2174	227.2412	229.265	227.2412	225.2174
Unit 2	40.8521	41.7114	42.5708	41.7114	40.8521
Unit 3	15.601	35.69	55.7789	35.69	15.601
Unit 4	8.2892	26.5888	44.8885	26.5888	8.2892
Unit 5	10.0404	28.7686	47.4968	28.7686	10.0404
Total	300	360	420	360	450

a) Load profile 1

For simulation using load profile 1, there is no difference in generation cost between DCOPF and DDCOPF since no ramp rates are violated when DCOPF is applied. The result of simulation for load profile 1 can be seen in table 2.

b) Load profile 2

For load profile 2, the generation cost of DCOPF is cheaper than DDCOPF. However, the result of DCOPF contains some violations against ramp rate limit at unit 3, 4, and 5. In case of unit 4, the change from load level 2 to load level 3 is 36.6 MW while the ramp rate of unit 4 is only 30 MW/h. On the contrary, there is no ramp rate violation in the result of DDCOPF.

c) Load profile 3

For load profile 3, generation cost of DCOPF is also cheaper than that of DDCOPF. Nevertheless, there are some violations against ramp rate limit at unit 3, 4, and 5. From time 1 to time 2, the different of generation of unit 3 is 50.22 but the ramp rate of unit 3 is only 36 MW/h. On the other hand, DDCOPF result in no ramp rate violations. The generation cost of DDCOPF is 463.97 more expensive than that of DCOPF.

Table 3
Active power output by DCOPF in load profile 2

Generator	Power output (MW) at time t				
	1	2	3	4	5
Unit 1	225.22	229.27	233.31	229.27	225.22
Unit 2	40.85	42.57	44.29	42.57	40.85
Unit 3	15.60	55.78	95.96	55.78	15.60
Unit 4	8.29	44.89	81.49	44.89	8.29
Unit 5	10.04	47.50	84.95	47.50	10.04
Total	300	420	540	420	300

Table 4
Active power output by DDCOPF in load profile 2

Generator	Power output (MW) at time t				
	1	2	3	4	5
Unit 1	221.82	229.31	234.28	229.31	221.82
Unit 2	38.18	42.69	49.72	42.69	38.18
Unit 3	19.84	55.84	91.84	55.84	19.84
Unit 4	08.75	44.75	80.75	44.75	08.75
Unit 5	11.41	47.41	83.41	47.41	11.41
Total	300	420	540	420	300

Table 5
Ramp rate violations in load profile 2

Method	ramp rate violation			
	1=>2	2=>3	3=>4	4=>5
DCOPF	Unit 3,4,5	Unit 3,4,5	Unit 3,4,5	Unit 3,4,5
DDCOPF	~	~	~	~

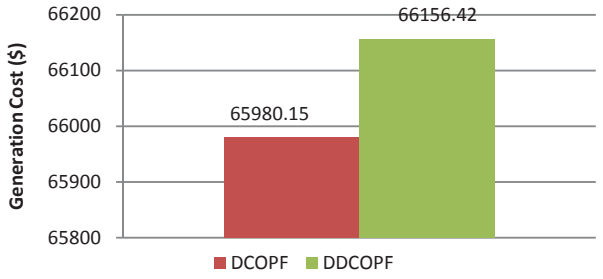


Figure 3. Comparison of generation cost between DCOPF and DDCOPF for load profile 2

B. Jawa Bali 500 kV System

To show the effectiveness of the proposed DDCOPF in the real system, the approach is applied in Jawa Bali 500 kV System, Indonesia. This system consists of 25 bus and 55 generators.

We use 24 hours load profile as displayed in figure 7. The trajectory of power generations of some generators are shown in figure 5 to figure 9 while ramp evaluation can be seen in the table 11. Those figures also display the trajectory of power generation and ramp evaluation during contingency circumstance. This circumstance is useful to make sure that the committed units are also able to follow the fluctuation of load not only under normal circumstance but also under contingency circumstance.

There are some violations against ramp rate limit in DCOPF calculation. It means that the result of DCOPF is not feasible to be implemented. On the contrary, DDCOPF result in no ramp rate violations. In the generation cost point of view, DDCOPF takes 0.013 % higher than DCOPF.

Table 6
Active power output by DCOPF in load profile 3

Generator	Power output (MW) at time t				
	1	2	3	4	5
Unit 1	225.22	230.28	226.51	230.28	225.22
Unit 2	40.85	43.00	95.63	43.00	40.85
Unit 3	15.60	65.82	100.00	65.82	15.60
Unit 4	8.29	54.04	100.00	54.04	8.29
Unit 5	10.04	56.86	77.86	56.86	10.04
Total	300	450	600	450	300

Table 7
Active power output by DDCOPF in load profile 3

Generator	Power output (MW) at time t				
	1	2	3	4	5
Unit 1	203.17	224.52	226.52	224.52	203.17
Unit 2	34.97	55.63	95.63	55.63	34.97
Unit 3	28.00	64.00	100.00	64.00	28.00
Unit 4	28.00	64.00	100.00	64.00	28.00
Unit 5	58.60	41.86	77.86	41.86	58.60
Total	300	450	600	450	300

Table 8
Ramp rate violations in load profile 3

Methods	ramp rate violation			
	1=>2	2=>3	3=>4	4=>5
DCOPF	Unit 3,4,5	Unit 3,4	Unit 3,4	Unit 3,4,5
DDCOPF	~	~	~	~

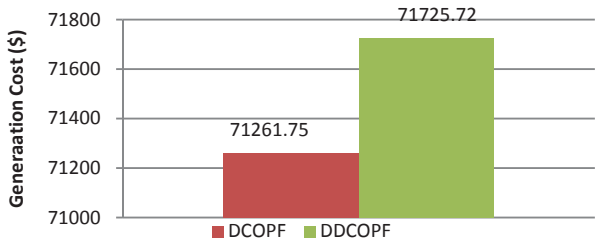


Figure 4. Comparison of generation cost between DCOPF and DDCOPF for load profile 3

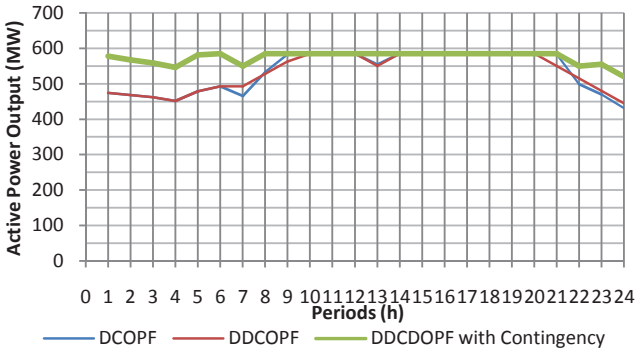


Figure 5. Power generation at PLTU Suralaya unit 5, unit 6, dan unit 7.

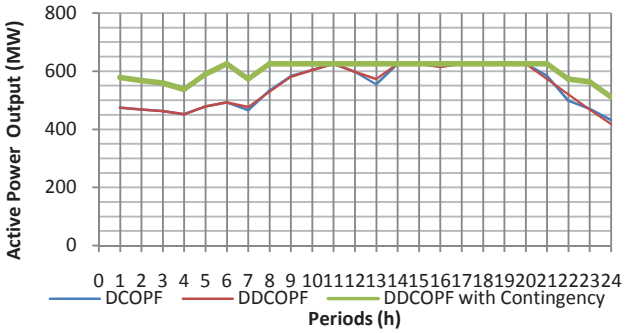


Figure 6. Power generation at PLTU Paiton Unit 8

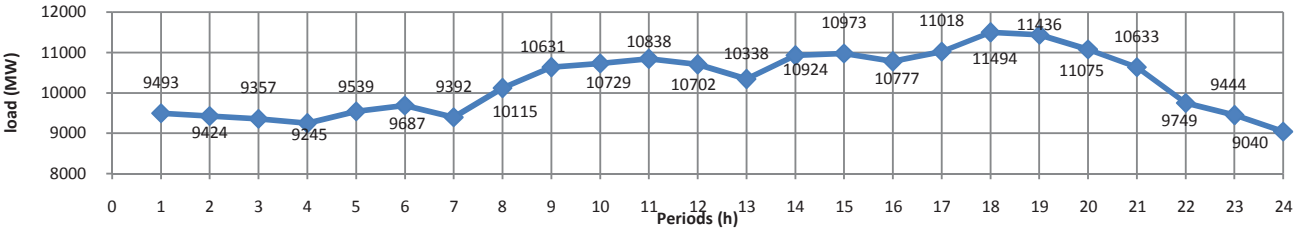


Figure 7. Load profile of Jawa Bali 500 kV System

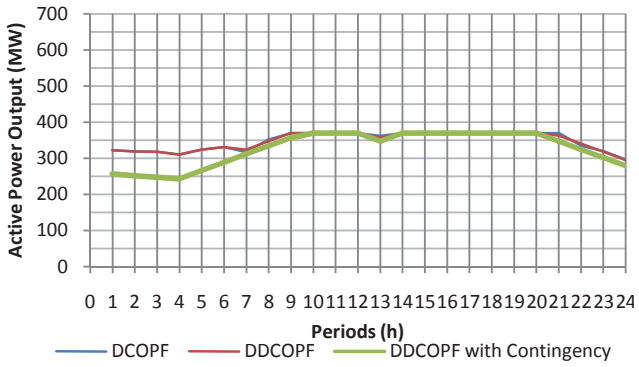


Figure 8. Power generation at PLTU Paiton Unit 1 dan 2.

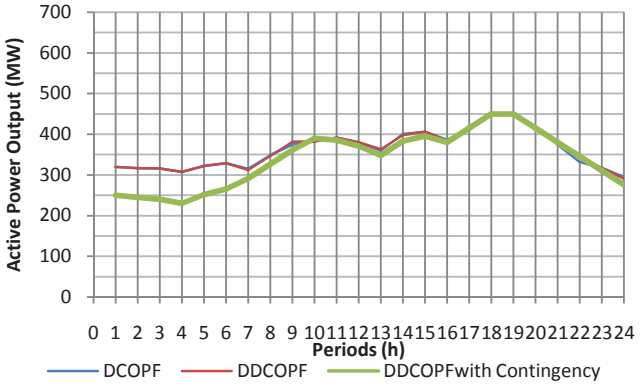


Figure 9. Power generation at PLTU Grati

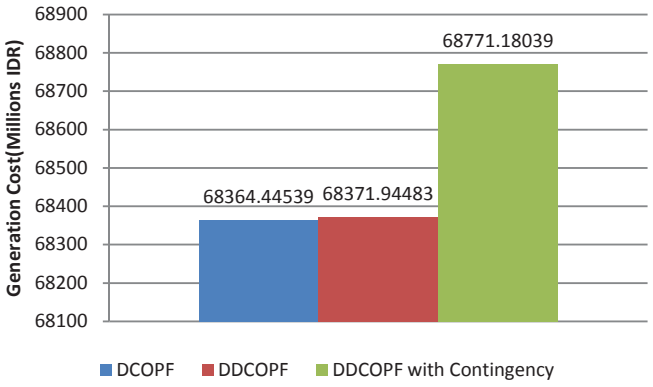


Figure 10. Comparison of total generation cost at Jawa Bali 500 KV System

Table 9
Comparison of DCOPF and DDCOPF against to constraint ramp rate and constraint branch limit

Methods	condition	Constraints evaluation		Conclusion
		Branch limit	Ramp rate limit	
DCOPF	Normal	√	x	Not Feasible
DDCOPF	Normal	√	√	Feasible
	Contingency	√	√	

V. CONCLUSION

From simulations and analysis above, it is observed that quadratic programming has successfully been employed to solve dc dynamic optimal power flow (DDCOPF) problem. By using this approach, ramp rate and transmission line violations can be avoided. Although DCOPF result in cheaper generation cost, it yields ramp rate or transmission line capacity violations. In contrast, DDCOPF result in slightly more expensive generation cost with no violations in both ramp rate and transmission line capacity.

REFERENCES

[1] XS Han, HB Gooi, Daniel S Kirschen, Dynamic Economic Dispatch : Feasible and Optimal Solutions, IEEE Transaction on Power Systems, Vol.16, No.1, February 2001

[2] N Yorino, H M Hafiz, Y Sasaki, Y Zoka, High-Speed Real Time Dynamic Economic Load Dispatch, IEEE Transaction on Power Systems, Vol. 27, No. 2, May 2012,

[3] K Xie, YH Song, Dynamic Optimal Power Flow by Interior Point Methods, IEE Generation Transmission and Distribution, Vol. 148, No. 1, pp.76 – 84, January 2001

[4] CY Chung, Wei Yan, Fang Liu, Decomposed Predictor-Corrector Interior Point Method for Dynamic Optimal Power Flow, IEEE Transaction on Power Systems, Vol.26, No.3, pp.1030-1039, August 2011

[5] T Nikman, MR Narimani, J Ahgaei, S Tabatabaei, M Nayeripour, Modified Honey Bee Mating Optimization to Solve Dynamic Optimal Power Flow Considering Generator Constraints, IET Generation, Transmission and Distribution, Vol.5, No. 10, pp.989-1002, January 2011

[6] Ray D. Zimmerman, Carlos E. Murillo-Sánchez, "*MATPOWER* 3.2. A MATLAB™ Power System Simulation Package", User's Manual, 2007.

Dynamic Response Analysis of Permanent Magnet Synchronous Motor Drives for City Electric Car

Abdullah Assegaf
Department of Electrical Engineering
Bandung state Polytechnic
Bandung, Indonesia
abdulahali@yahoo.com

Agus Purwadi, Yanuarsyah Haroen
School of Electrical Engineering & Informatics
Bandung Institute of Technology
Bandung, Indonesia
agusp@konversi.ee.itb.ac.id

Abstract— In the next few years, electric cars will be one of the main vehicles in transportation system and are designed to meet the requirement of high efficiency and clean energy vehicles. Since the force to propel electric vehicles comes only from electric motor, the profile of tractive effort -speed of the vehicles is determined by the torque – speed profile of the motor. The torque – speed of the electric motor consists of constant torque and constant power region, but the driving pattern in urban area has often forced the electric motor into the constant torque operation area. In this paper, the dynamic response of the PMSM under constant torque operation of city electric car is presented. The mathematical and Matlab/Simulink model of PMSM are derived and constructed, the dynamic response of the motor for city electric car under start-go mode will be obtained from simulation results. The dynamic response of torque and speed are then analyzed.

Keywords —electric car; PMSM; torque-speed profile; dynamic response; Matlab/Simulink;modeling

I. INTRODUCTION

The global issues of clean energy and the rapid depletion of oil resources have changed research and development program to pay more attention in electric vehicles rather than internal combustion engines (ICE). The research objective is mainly a high efficiency with low pollutant emissions vehicles. Electric Vehicles (EVs) or Hybrid Electric Vehicles (HEVs) are expected to replace conventional vehicles in a very near future. The electric motor is taking over the combustion engine to propel the vehicle. As a fundamental components, an electric motor, controllers and energy storage components, holds a crucial role to achieve the objective [1].

The brushless motor is the most widely used in electric vehicle, and the brushed motor is not favorable nowadays. The brushless AC motor, include the induction motor and the permanent magnet motor. The permanent magnet can be divided into two types, i.e. the sine wave distributed stator winding as permanent magnet synchronous motor (PMSM) and trapezoidal distributed stator winding which is famous with permanent magnet brushless dc motor (PMBLDCM) type. These two permanent magnet configuration, recently are most dominant types for small to medium car. The induction motor is more applicable for large vehicle including for mass transportations.

The vehicle performance is completely determined by the profile of tractive effort , which means the torque on the motor versus vehicle speed on the driven wheels [1]. Vehicle profiles with a given power rating, is basically operating at constant power region in the speed range. The ICE (internal combustion engine) powered car uses multi-gear to produce the constant power profile. However, these torque changes resulted in low efficiency, especially when working on start-go driving pattern, as in urban operations. So the multi-gear electric car is undesirable option.

Electric motors on the other hand, has the identical torque-speed characteristics with a required profile of tractive effort - speed of the vehicles as shown in figure 1. Since an electric vehicle is driven by electric motor, it is possible the system to use single gear to get the desired vehicle profile. This will much simplifies the power train structure and will also enhance the efficiency of the vehicle.

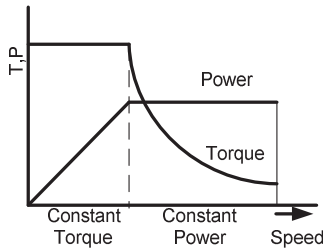


Fig. 1. Torque - speed characteristics electric motor

PMSM with its special features are beginning widely used in electric vehicles. They are robust, high power density, capable of operating at high motor and inverter efficiencies over wide speed ranges. The permanent magnet causes its constant torque area of operation wider than synchronous motor with rotor winding. The constant power operation is achieved through an addition of a control circuit which causes the flux generated by stator current to oppose the flux of permanent magnet.

The characteristic of the electric motor indicates area of the constant torque and constant power as shown in figure 1. The torque sensing can be done in two modes, namely basic rpm mode (ωb) and the order of the power limits. Motor power limit value is defined as:

$$P_{max} = T_{max} \times \omega b \qquad (1)$$

- RPM based mode (ω_b)

This mode of operation uses base speed of the rotor as the boundary to traverse from torque constant operation to the constant power operation, i.e.

$$\omega < \omega_b \qquad T=T_{T\text{-constant}}$$

$$T_{T\text{-constant}}=command\ x\ Tmax \tag{2}$$

$$\omega \geq \omega_b \qquad T = T_{P\text{-constant}},$$

$$T_{P\text{-constant}} = \frac{command\ x\ Pmax}{\omega} \tag{3}$$

- Power limit mode (P_{max})

Motor controllers this mode works on constant torque region if the power absorbed by the motor is less than Pmax motors and switch to the constant power operation when the motor output power P is greater than the maximum power P_{max}, i.e.

$$P<P_{max}. \qquad T=T_{T\text{-constant}}$$

$$T_{T\text{-constant}}= \text{command x } T_{max} \tag{4}$$

$$P \geq P_{max} \qquad T = T_{P\text{-constant}}$$

$$T_{P\text{-constant}}=\frac{P_{max}}{\omega} \tag{5}$$

An electric motor used in electric car is required to have the characteristics of high torque at low speed to obtain high initial torque. This mean the electric motor operated within the constant torque mode at most of the time in traffic jam. On the other hand, adequate acceleration and wide speed bounds for operation with constant power.

The magnetic field of PMSM is excited by high-energy permanent magnets, the overall weight and volume can be significantly reduced for given output torque, resulting in higher torque density. Because of the absence of rotor winding and rotor copper losses, their efficiency is inherently higher than that of induction motors. However, this motor has a short constant power range due to its constant PM flux, which can only be weakened through production of stator field component of stator current [1].

If a PMSM is started up from zero to a some speed, the response is slightly under damped. The torque is held constant at the maximum capability of the machine while the motor runs up to reference speed [2].

This paper presents the modeling of PMSM drives for an urban or city electric car where start – go mode is almost unavoidable. Mathematical modeling of the machines, are described in very famous dqo theory. The dynamic responses of motors torque and speed are then analyzed.

II. MATHEMATICAL AND SIMULINK MODELING.

Nomenclature

B	damping constant (Nm second/ radian)
i_d, i_q	direct and quadrature (dq) axis stator currents (Amperes)
J	moment of inertia (kg –m^2)
L_d , L_q	d, q axis inductances (Henry)
p	derivative operator
P	number of pole pairs
R_a	stator resistance (Ohm)
T_e	electric torque (Newton meter)
T_L	load torque, (Newton meter)
v_d, v_q	d, q axis voltages (Volt)
λ_{af}	mutual flux due to magnets (Webber).
λ_d, λ_q	d, q axis flux linkages (Webber)
M_{sr}	Mutual inductance of stator and rotor winding
ω_r	rotor speed (radian per second)
ω_e	synchronous speed, (radian per second)
θ	rotor position

2.1 Mathematical modeling of PMSM

The general voltage equation for a cylindrical electrical machines, were shown below [3]:

$$[v] = [R][i] + \frac{d}{dt}[L(\theta)][i] \tag{6.a}$$

Or

$$[v] = \underbrace{[R][i] + [L(\theta)]\frac{\partial}{\partial t}[i]}_{transformation\ emf} + \underbrace{\frac{\partial \theta}{\partial t}\{\frac{\partial}{\partial \theta}[L(\theta)]\}[i]}_{rotation\ emf} \tag{6.b}$$

where :

$$\left[\frac{\partial}{\partial \theta}(L(\theta)\right] = \begin{bmatrix} 0 & \frac{\partial}{\partial \theta}L_{sr}(\theta) \\ \frac{\partial}{\partial \theta}L_{sr}(\theta) & 0 \end{bmatrix} \tag{6.c}$$

and

$$\left[\frac{\partial}{\partial \theta}L_{sr}(\theta)\right] = \begin{bmatrix} -M_{sr}\sin(\theta) & -M_{sr}\sin(\theta+2\pi/3) & -M_{sr}\sin(\theta+4\pi/3) \\ -M_{sr}\sin(\theta+4\pi/3) & -M_{sr}\sin(\theta) & -M_{sr}\sin(\theta+2\pi/3) \\ -M_{sr}\sin(\theta+2\pi/3) & -M_{sr}\sin(\theta+4\pi/3) & -M_{sr}\sin(\theta) \end{bmatrix} \tag{6.d}$$

Stator of the PMSM and conventional synchronous machine rotor windings are having the same construction, the difference is the rotor PMSM using permanent magnets. In addition, there is no difference between the back emf generated by permanent magnet and generated by the excitation coil. Therefore, the mathematical model of the PMSM is the same as the SM rotor windings. The model of the PMSM is based on the following assumption.

1 The saturation effect is neglected
2 The back-emf generation is purely sinusoidal
3 Hysteresis and eddy currents losses are neglected
4 No damper winding on rotor.
5 The current in electric motor is symmetrical three phase sinusoidal current.

The most common method in the analysis of electric motor control of PMSM type is in d-q axes mathematical models for both steady state and transient analysis. The two-axis variables are obtained from stator variable through the famous Park

transform. The vector diagram of stator voltages in d-q frames is shown in figure 2. [4]

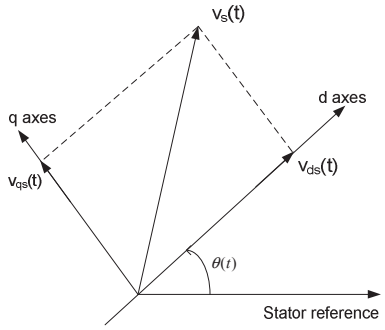


Fig. 2. Stator voltage phasor and its d-q axes components

v_d,v_q and v_o are obtained from v_{as}, v_{bs} and v_{cs} through the Park transformation defined below.

$$\begin{bmatrix} v_d \\ v_q \\ v_o \end{bmatrix} = \frac{2}{3} \begin{bmatrix} \cos(\theta) & \cos\left(\theta-\frac{2\pi}{3}\right) & \cos\left(\theta+\frac{2\pi}{3}\right) \\ -\sin(\theta) & -\sin\left(\theta-\frac{2\pi}{3}\right) & -\sin\left(\theta+\frac{2\pi}{3}\right) \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} v_{as} \\ v_{bs} \\ v_{cs} \end{bmatrix} \quad (7)$$

The abc variables are obtained from d, q variables through the inverse of the Park transform defined below.

$$\begin{bmatrix} v_{as} \\ v_{bs} \\ v_{cs} \end{bmatrix} \begin{bmatrix} \cos(\theta) & \sin(\theta) & 1 \\ \cos\left(\theta-\frac{2\pi}{3}\right) & \sin\left(\theta-\frac{2\pi}{3}\right) & 1 \\ \cos\left(\theta+\frac{2\pi}{3}\right) & \sin\left(\theta+\frac{2\pi}{3}\right) & 1 \end{bmatrix} \begin{bmatrix} v_d \\ v_q \\ v_o \end{bmatrix} \quad (8)$$

Note that since the Park transformed defined is not power invariant, the factor 3/2 is used in power formulae.

The total input power to the machine in terms of abc variables is :

$$s=v_a i_a + v_b i_b + v_c i_c \quad (9)$$

while in terms of d, q variables

$$s=v_d i_d + v_q i_q \quad (10)$$

Mathematical models typically consist of voltage equations, stator flux equations, electromagnetic torque equation and dynamic equation machine. By deriving equation (1.b), the d-q axis voltage equations can be obtained in the following form:

$$v_d=Ri_d+p\lambda_d-\omega_e\lambda_q \quad (11)$$

$$v_q=Ri_q+p\lambda_q+\omega_e\lambda_d \quad (12)$$

From equation (6) & (7), the equivalent circuit of PMSM can be obtained as shown in figure 3.

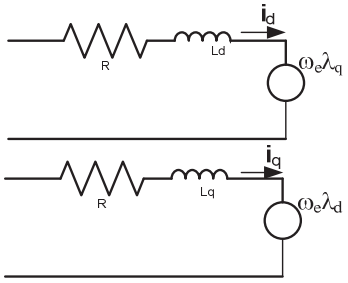


Fig. 3. PMSM equivalent circuit from dynamic equations.

The flux equations are:

$$\lambda_q=L_q i_q \quad (13)$$

$$\lambda_d=L_d i_d+\lambda_{af} \quad (14)$$

λ_{af}, is mutual flux linkage of the magnet. And the electromagnetic torque equation is:

$$T_e = 3P/2[\lambda_{af} i_q + (L_d - L_q) i_d i_q] \quad (15)$$

A dynamic equation system with moment of inertia J and friction coefficient B with the load torque T_L is:

$$T_e = B\omega_r + Jd\omega_r/dt + T_L \quad (16)$$

And the relationship between the position rotor (θ_r) with rotor angular velocity is:

$$\omega_e = d\theta_e/dt = (P/2)\,\omega_r \quad (17)$$

PMSM dynamic modeling is obtained by changing the equation (7) to (12) into the following equations:

$$pi_d= (v_d- Ri_d- \omega_e L_q i_q)/L_d \quad (18)$$

$$pi_q = (v_q - Ri_q- \omega_e L_d i_d - \omega_e\lambda_{af})/L_q \quad (19)$$

$$p\omega_r = (T_e - T_L - B\omega_r)/J. \quad (20)$$

2.2 Modeling PMSM in simulink

Park transformation is then used to obtain variable in dqo coordinates from phase variables a, b and c. And variables a, b and c can be obtained using the inverse Park transformation of variable in coordinates dqo. Figure 3 and 4 show the model in simulink of the Park transform and inverse transform respectively.

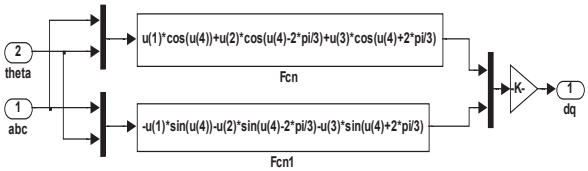


Fig. 4. The abc to dq transform in simulink

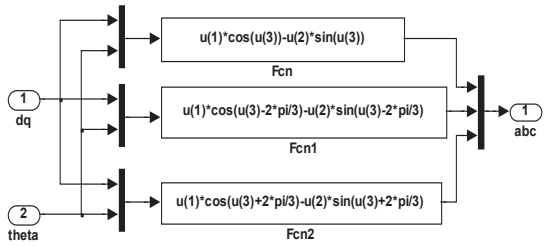


Fig. 5. The dq to abc transform in simulink

The whole model of PMSM itself in simulink is shown in Figure 5.

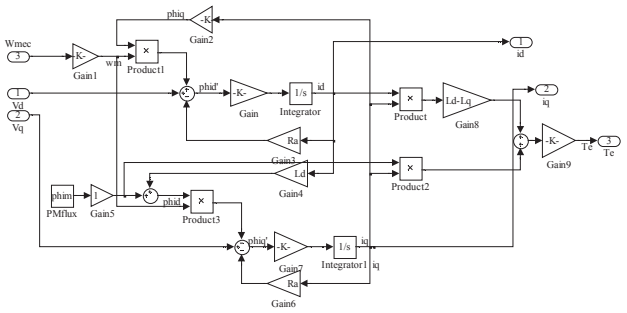


Fig. 6. Model of PMSM in simulink

An electric motor used in electric car is required to have the characteristics of high torque at low speed to obtain high initial torque. The motor is operated within the constant torque mode at most of the time in traffic jam. On the other hand, adequate acceleration and wide speed bounds for operation with constant power.

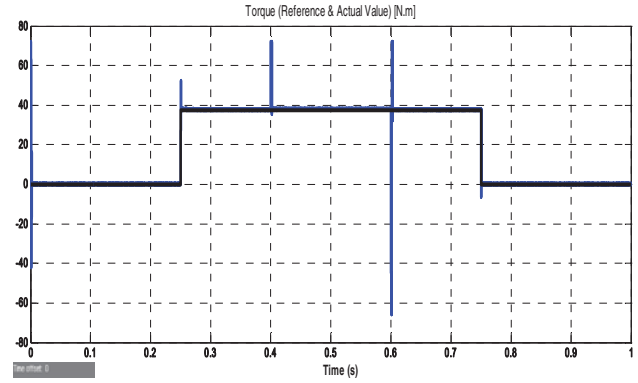
This paper discusses the PMSM responses to abrupt changes of load torque under constant torque operation. The boundary of the two mode of operation of motor, i.e. constant torque and the constant power mode, is the power limits of the motor. Motor is operated in constant torque if the power absorbed by the motor is less than its power limit.

III. THE SIMULATION RESULTS

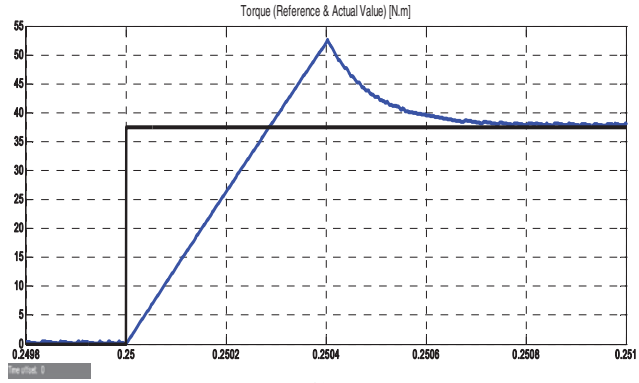
Simulations was performed on motor with power rating 10 kW, 72 V and 1500 RPM speed and rotor flux 0.3 Weber, pole number 4 and the stator resistance Ra = 0,0575 Ohm. The Simulink of PMSM has been tested with sudden insertion and detachment of load torque and varying speed reference as given below:

- T_b=50 Nm – base load
- t_{mech_time} = [0 0.25 0.25 0.75 0.75 tstop] time of load torque variations
- t_{mech_value} = [0.0 0.0 0.75 0.75 0.0 0.0]*Tb Torque variations
- w_{mech_time} = [0 0.4 0.4 0.6 0.6 tstop] time of RPM variations
- w_{mech_value} = [500 500 1000 1000 500 500] RPM variations

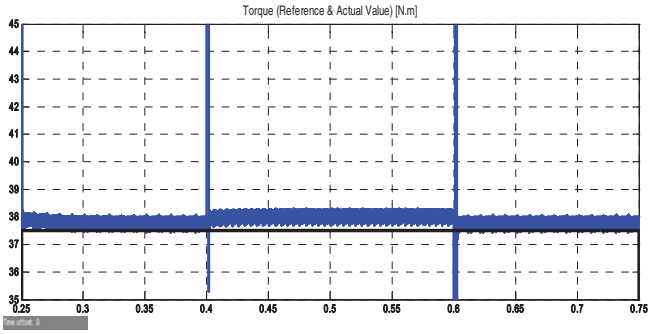
Simulation results of the dynamic responses of torque and speed of PMSM are shown below



a)



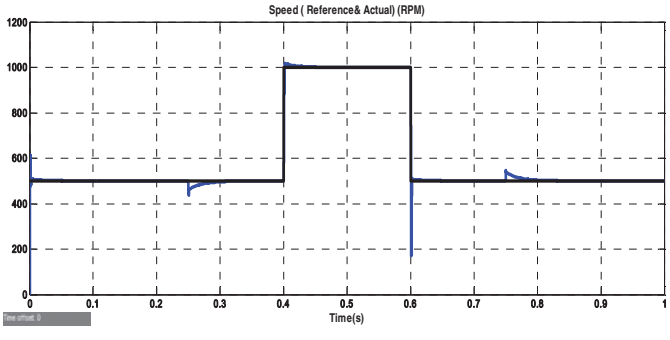
b)



c)

Fig. 7. Torque Response :

a) full period of test 0 -1 second b) at insertion of load 0.249 to 0.251 second c) constant torque period 0.25 to 0.75 Second



a)

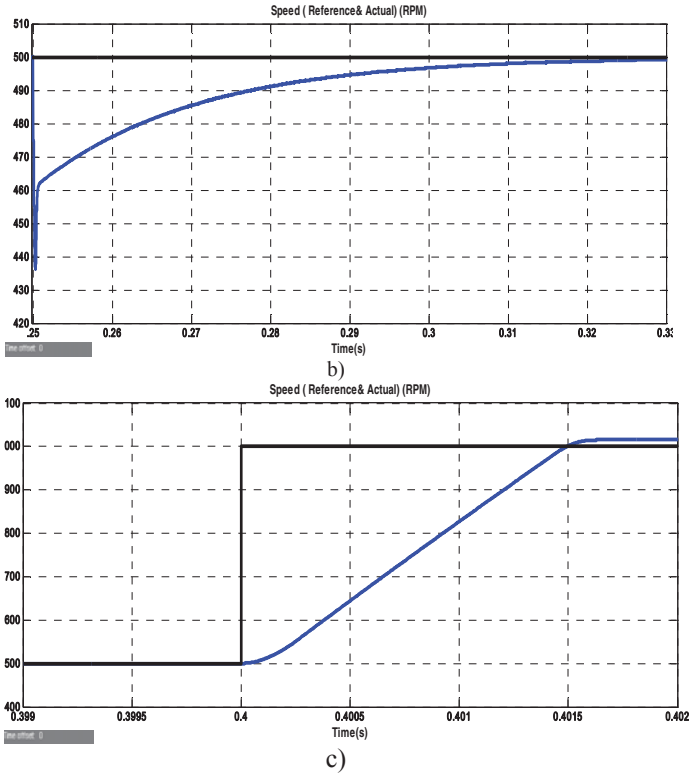


Fig. 8. Speed Response :

a) full period of test 0 -1 second b) speed response of disturbance of torque variation c) at reference speed change from 500 to 1,000 RPM

IV. CONCLUSION:

The torque of PMSM has a very fast response with rise time less than 0.0004 second and its ripple is about 1.33%. The speed response due to its reference value variations are very good with settling time less than 0.0025 second, but the speed response due to disturbance of torque variation has a much longer settling time(0.075 s). The dqo mode of the motor control is also widely used for PMSM, it will makes it possible to control torque and speed individually just like a conventional dc motor. With all of these reasons, the application of the PMSM as a part of power train system for city electric car is become favorable.

V. REFERENCE:

[1] M. Ehsani, Yimin Gao and Sebastian Gay," Characterization of Electric Motor Drives for Traction Applications" Advanced Vehicle systems Research Program Department of electrical engineering Texas A&M University, IEEE, 2003

[2] Pragasen Pillay and Ramu Krishnan "Modeling, Simulation, and Analysis of Permanent-Magnet Motor Drives, Part I: The Permanent-Magnet Synchronous Motor Drive", IEEE Transactions On Industry Applications, Vol. 25, No. 2, March - April 1989

[3] Haroen, Yanuarsyah "Dinamika Mesin Elektrik, Pemodelan, Analisis dan Aplikasinya", Penerbit ITB,Edisi Pertama, 2010

[4] Pragasen Pillay and Ramu Krishnan "Modeling of Permanent Magnet Motor Drives" IEEE Transactions On Industrial Electronics, Vol. 35, No. 4, November 1988

[5] PMSM control, Texas Instrument, 2007

[6] Cetin Elmas and Oguz Ustun," A hybrid controller for the speed control of a permanent magnet synchronous motor drive", Control Engineering Practice 16 (2008) 260–270, Science Direct,2007

[7] C. C ,Chan,. , et. all." Novel Permanent Magnet Motor Drives for Electric Vehicles" IEEE Transactions On Industrial Electronics, Vol. 43, No. 2, April 1996

[8] C. C. Chan, "The State of the Art of Electric, Hybrid, and Fuel Cell Vehicles", Proceedings of the IEEE | Vol. 95, No. 4, April 2007

INPUT CURRENT RIPPLE ANALYSIS OF DOUBLE STATOR AC DRIVE SYSTEMS

Raymond Samuel Parlindungan[1] and Pekik Argo Dahono[2]
School of Electrical Engineering and Informatics, Institute of Technology Bandung
[1]rsp.htp@gmail.com
[2]pekik@konversi.ee.itb.ac.id

Abstract – An input current ripple analysis of PWM inverter fed double-stator AC drive system is proposed in this paper. The expression of rms input current ripple of PWM inverter as a function of phase difference between stator windings is derived. It is found that the optimum phase difference that results in minimum input current ripple is 60°. Several experimental results are included to show the validity of the proposed analysis method.

Key words: double stator, PWM inverter, current ripple, AC drive

I. PENDAHULUAN

High-power electric AC drive systems are widely used in propulsion systems, petrochemical industries, and steel industries. Various converter topologies were proposed and applied for these applications [1]-[3]. In these applications, high power-density and high reliability are desirable.

Multiphasing is an enabler to achieve high power-density and reliability AC drive systems[4]-[10]. By using high phase number AC drive systems, the following advantages can be achieved:

- Reduced rotor harmonic losses
- Reduced torque ripple
- Increased power density
- Increased reliability
- Reduced inverter input current ripple

Reducing inverter input current ripple is very important as the current ripple will directly determine the size of the DC link capacitor of the inverter. It has been reported that the DC link capacitor is the weakest component in an inverter [11].

The most commonly used multiphase AC drive system is the one using double stator AC motor. The AC motor can be induction or synchronous motor. Various control techniques for double stator AC drive system were proposed[7]-[10]. Though a lot of work have been published, only a few of them working on inverter input current ripple[12]-[18]. In the case of square-wave inverters, it has been shown that two stator windings with phase difference of 30° results in minimum input current ripple and minimum torque pulsation. As far as authors are aware, the optimum phase difference for PWM mode of operation has not been reported.

This paper presents an input current ripple analysis of PWM inverter fed double stator AC drive systems. At first, the expression of rms input current ripple of PWM inverter as a function of stator phase difference is derived. Based on the derived expression, the optimum phase difference that results in minimum input current ripple under carrier based PWM technique is determined. Different to square wave mode of operation, it is found that the optimum phase difference under PWM operation is 60°. Experimental results are included to show the validity of the proposed analysis method. The derived expression is useful in designing the required DC filter capacitor.

II. DOUBLESTATOR AC DRIVE SYSTEM

The scheme of double stator AC drive system can be seen in Fig. 1. The system is constructed by a six-leg inverter anda double stator AC motor. The DC voltage source is usually obtained by rectifying a three-phase AC voltage supply. A large electrolytic DC capacitor is connected directly to the input of inverter to absorb the current ripple that is generated by the inverter. The motor can be synchronous or asynchronous motor. Though various PWM techniques have been proposed to control this type of inverter, carrier based PWM technique is still the most popular due to its simplicity. In carrier based PWM technique, the ON-OFF signals for inverter switching devices are determined by comparing reference signals to a high-frequency triangular carrier signal.

The AC motor that is used in this system has two three-phase stator winding sets as shown in Fig. 2. The phase angle difference between these two sets is α. Both sets of windings have separated neutral points. In the case of square wave mode of operation, the common phase difference between two sets of windings is 30°. By using this phase difference, the inverter input current ripple and motor torque pulsation will be minimum.

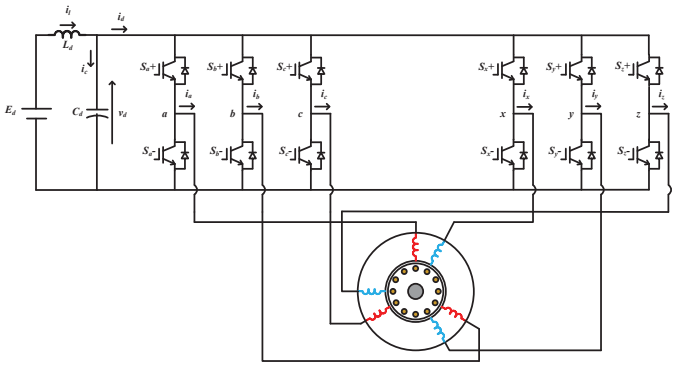


Fig. 1.Double stator AC drive system.

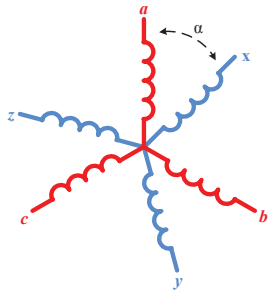


Fig.2.Double stator winding configuration.

III. INVERTER CURRENT RIPPLE ANALYSIS

With the advance of power semiconductor devices, several kilohertz of switching frequencies can be achieved easily without sacrificing the efficiency of inverter. Under such high switching frequencies, the inverter output current can be assumed almost sinusoidal as described by the following expression:

$$\begin{aligned} i_a &= \sqrt{2}I_l \sin[\theta - \phi] \\ i_b &= \sqrt{2}I_l \sin\left[\theta + \frac{2\pi}{3} - \phi\right] \\ i_c &= \sqrt{2}I_l \sin\left[\theta - \frac{2\pi}{3} - \phi\right] \\ i_x &= \sqrt{2}I_l \sin[\theta + \alpha - \phi] \\ i_y &= \sqrt{2}I_l \sin\left[\theta + \frac{2\pi}{3} + \alpha - \phi\right] \\ i_z &= \sqrt{2}I_l \sin\left[\theta - \frac{2\pi}{3} + \alpha - \phi\right] \end{aligned} \quad (1)$$

where I_l is rms outputphase current, α is phase angle difference between two stator winding sets, andϕ is the load power factor.

Based on Fig.1, the inverter input current as a function of the output currents can be expressed as follow:

$$i_d = S_a i_a + S_b i_b + S_c i_c + S_x i_x + S_y i_y + S_z i_z \quad (2)$$

Where S_n are inverter switching states with n = a, b, .. , z.The inverter switching state is unity (zero) when the upper switching device of the associated phase receives an ON (OFF) signal.

In carrier based PWM technique for double stator AC drive systems, two three-phase reference signals are compared to a high-frequency triangular signal to determine the ON-OFF signals for inverter switching devices. The phase angle difference between these two three-phase signals is α. An example of two three-phase reference signals is shown in Fig. 3. In this analysis, it is assumed that the frequency of carrier signal is much higher than the reference ones.

For the purpose of analysis, six intervals are identified as shown in Fig. 3. Each interval will result different switching pattern. The detailed inverter waveforms during one carrier period that is valid in the interval AB is shown in Fig. 4. As the carrier frequency is much higher than the reference signal frequency, the reference signals can be assumed constant during one carrier period.

The time intervals in Fig. 4 can be calculated as follows:

$$\frac{T_0}{T_S} = \frac{1 - v_a^r}{4} \quad (3)$$

$$\frac{T_1}{T_S} = \frac{v_a^r - v_z^r}{4} \quad (4)$$

$$\frac{T_2}{T_S} = \frac{v_z^r - v_x^r}{4} \quad (5)$$

$$\frac{T_3}{T_S} = \frac{v_x^r - v_c^r}{4} \quad (6)$$

$$\frac{T_4}{T_S} = \frac{v_c^r - v_b^r}{4} \quad (7)$$

$$\frac{T_5}{T_S} = \frac{v_b^r - v_y^r}{4} \quad (8)$$

$$\frac{T_6}{T_S} = \frac{v_y^r + 1}{4} \quad (9)$$

Based on (2) and inverter switching states as shown in Fig. 4, the inverter input current in one carrier period can be written as follow:

$$i_{d1} \begin{cases} 0 & t_0 < t < t_1 \\ i_a & t_1 < t < t_2 \\ i_a + i_z & t_2 < t < t_3 \\ i_a + i_x + i_z & t_3 < t < t_4 \\ -(i_b + i_y) & t_4 < t < t_5 \\ -i_y & t_5 < t < t_6 \\ 0 & t_6 < t < t_7 \\ -i_y & t_7 < t < t_8 \\ -(i_b + i_y) & t_8 < t < t_9 \\ i_a + i_x + i_z & t_9 < t < t_{10} \\ i_a + i_z & t_{10} < t < t_{11} \\ i_a & t_{11} < t < t_{12} \\ 0 & t_{12} < t < t_{13} \end{cases} \quad (10)$$

The mean square value of the inverter input current in one carrier period can be determined as follow:

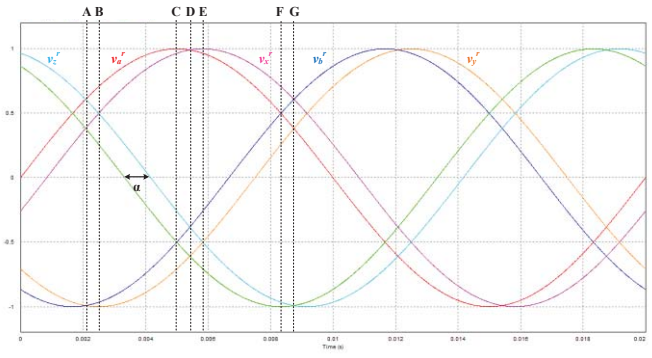


Fig. 3. Reference signals of double stator AC drive system

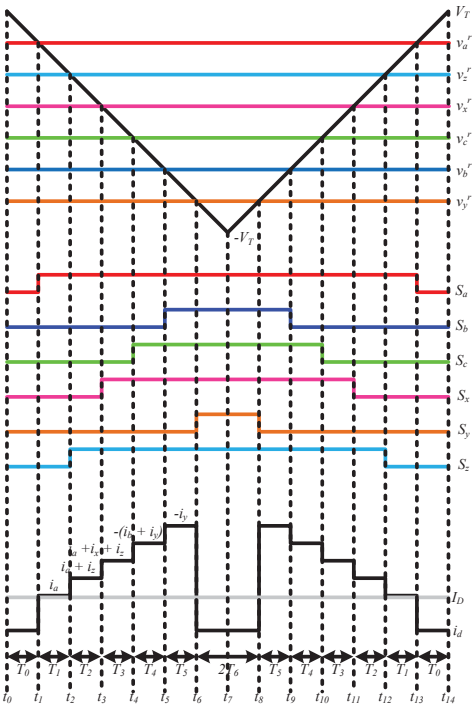


Fig. 4. Detailed inverter waveforms.

$$I_d^2 = \frac{2}{T_S}\left[\int_0^{T_1}(i_a)^2\,dt + \int_0^{T_2}(i_a+i_z)^2\,dt + \int_0^{T_3}(i_a+i_x+i_z)^2\,dt + \int_0^{T_4}(i_b+i_y)^2\,dt + \int_0^{T_5}(i_y)^2\,dt\right] \quad (11)$$

By using a similar method, the mean square values of inverter input current in the other intervals can be determined. The average value over one fundamental period of inverter output voltage of the mean square value of inverter input current can be determined as follow:

$$I_{d,avg}^2 = \frac{1}{2\pi}\int_0^{2\pi} I_d^2 d\theta \quad (12)$$

As the waveform is symmetrical the above integral can be done only over 120° interval. The rms value of the ripple component of the inverter input current can be determined by using the following expression

$$\tilde{I_d} = \sqrt{{I_{d,avg}}^2 - I_{dc}^2} \quad (13)$$

Where I_{dc} is the dc component of inverter input current. The dc component of the inverter input current can be obtained easily by assuming the inverter losses are small and can be neglected:

$$I_{dc} = \frac{3}{\sqrt{2}} k I_l \cos\phi \quad (14)$$

where$k$ is the modulation index.

Substituting (12) and (14) into (13) will result in the expression of inverter input current ripple. The obtained expression is shown in Table 1.

The optimum phase angle difference between the set of stator windings that results in minimum input current ripple can be determined as follow:

$$\frac{d\tilde{I}_d}{d\alpha} = 0 \quad (18)$$

Substituting (13) into (15) results in an optimum phase angle of α=60°. Thus, the optimum configuration of double stator AC drive system that results in minimum input current ripple is the one that equal to symmetrical six phase AC system

Table 1 shows that the inverter input current ripple is not influenced by the switching or carrier frequency.Thus, we cannot reduce the inverter input current ripple by increasing the switching frequency. The inverter input current ripple under the optimum phase angle is

$$\tilde{I}_{d\ optimum} = \frac{I_l}{2}\left[\frac{k}{\pi}\begin{pmatrix}12+12\sqrt{3}-9k\pi\\+(16+8\sqrt{3}-9k\pi)\cos 2\phi\end{pmatrix}\right]^{\frac{1}{2}} \quad (19)$$

Table 1.Input current ripple for double stator AC drive system

$0° \le \alpha \le 60°$.
$\frac{I_l}{2}\left[\frac{k}{\pi}\begin{pmatrix}-9k\pi+4\sqrt{3}\cos\frac{\alpha}{2}(3+\cos 2\phi)\\+\cos 2\phi\left(-9k\pi+8\left(\sqrt{3}+\sin\frac{\alpha}{2}\right)\right)\\+12\left(\sin\frac{\alpha}{2}-\sin\frac{3\alpha}{2}\right)\end{pmatrix}\right]^{\frac{1}{2}}$ (15)
$60° \le \alpha \le 120°$.
$\frac{I_l}{2}\left[\frac{k}{\pi}\begin{pmatrix}-9k\pi+4\sqrt{3}\cos\frac{\alpha}{2}(3+\cos 2\phi)\\+\cos 2\phi\left(-9k\pi+8\left(\sqrt{3}+\sin\frac{\alpha}{2}\right)\right)\\+12\left(\sqrt{3}+\sin\frac{\alpha}{2}-\sin\frac{3\alpha}{2}\right)\end{pmatrix}\right]^{\frac{1}{2}}$ (16)
$120° \le \alpha \le 180°$
$\frac{I_l}{2}\left[\frac{k}{\pi}\begin{pmatrix}-9k\pi+\cos 2\phi\left(8\sqrt{3}-9k\pi+16\sin\frac{\alpha}{2}\right)\\+12\left(\sqrt{3}+2\sin\frac{\alpha}{2}-\sin\frac{3\alpha}{2}\right)\end{pmatrix}\right]^{\frac{1}{2}}$ (17)

IV. Experimental ResultS

In order to verify the proposed analysis method, some measurements on a small double stator AC drive system were conducted. A carrier frequency of 1000 Hz was used in the experiment. The DC input voltage for the inverter was fixed at 50 V_{DC} during the experiment. Double stator three-phase induction motors were used in the experimental system. The rotor of induction motor was locked during the measurements. Based on locked rotor tests, it is found that the equivalent leakage inductance and resistance of the induction motor are 0.9 Ω and 21 mH, respectively. The inverter input current waveform is recorded by using a digital oscilloscope and the result is processed by a digital computer to determine the inverter input current ripple component.

Figs. 5-7 show comparison between calculated and experimental results of inverter input current ripple under three different stator phase angles. Agreement between calculated and experimental results can be appreciated from these figures.

V. Conclusion

A method for analysis the inverter input current ripple of a double stator ac drive system has been proposed and the results have been validated through experiments. The value of the RMS input ripple current is influenced by the phase angle between the two sets of three-phase windings, phase current, modulation index, and power factor of the load. In addition, it has been revealed that the phase angle of the the minimum input current ripple is produced when the phase angle is 60° between each phase. Laboratory experiments have also been done to validate the proposed analysis method.

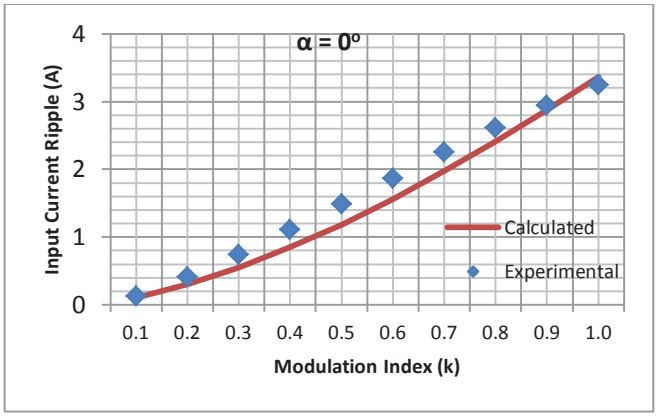


Fig. 5.Input current ripple for$\alpha = 0°$.

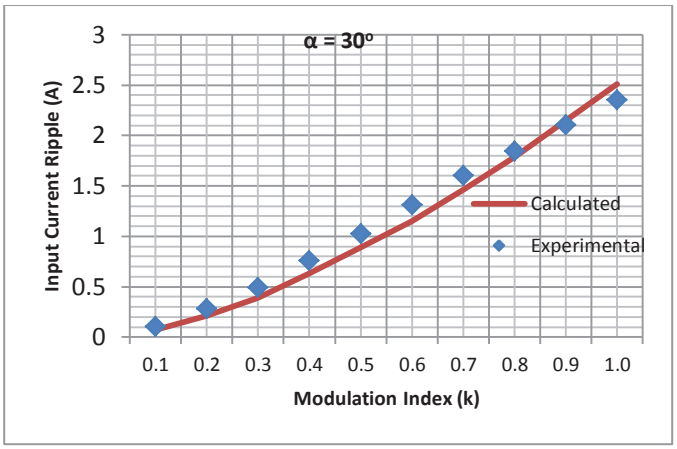


Fig. 6. Input current ripple for$\alpha = 30°$.

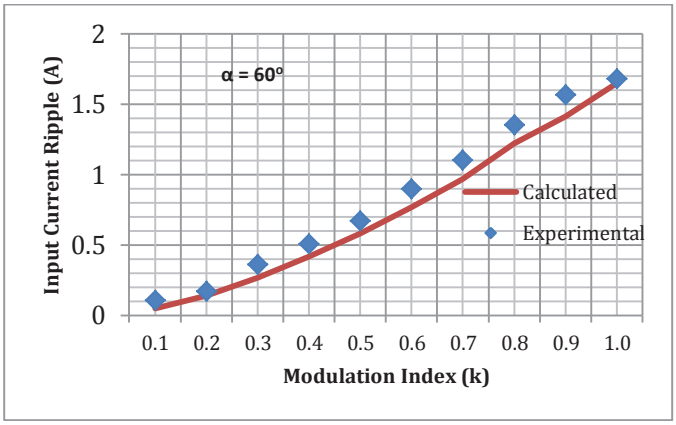


Fig. 7. Input current ripple for$\alpha = 60°$.

II. Referennce

[1] H. Kielgas and R., Nill, "Converter Propulsion Systems with Three-Phase Induction Motors for Electric Traction Vehicles," *IEEE Trans. Ind. Appl., Vo. 1A-16, No.2, March/April 1980*.

[2] T. M. Jahns and V. Blasko, "Recent Advances in Power Electronics Technology for Industrial and Traction Machine Drives," *Proc. IEEE, Vol. 89, No. 6, June 2001*.

[3] Z. J. Shen and I. Omura, "Power Semiconductor Devices for Hybrid, Electric, and Fuel Cell Vehicles," *Proc. IEEE, Vol. 95, No. 4, April 2007*.

[4] R. H. Nelson and P. C. Krause, "Induction Machine Analysis for Arbitrary Displacement Between Multiple Winding Sets," IEEE Tran. Power App. Sys., Vol. PAS-93 , No. 3, 1974.

[5] E. A. Klingshirn, "High Phase Order Induction Motors – Part 1 – Description and Theoretical Considerations," *IEEE Trans. Power App. Sys., Vol. PAS-102, No. 1, January 1983*.

[6] H. A. Toliyat, T. A. Lipo, and J. C. White, "Analysis of a concentrated winding induction machine for adjustabel speed drive applications. I. Motor analysis" *IEEE Trans. Energy Con., Vol. 6, No. 4, Dec. 1991*.

[7] L. Parsa, "On Advantages of Multi-Phase Machines," 31st Annual Conf. of IEEE Industrial Electronics Society, 2005.

[8] J. M. Apsley, S. Williamson, A. C. Smith, and M. Barnes, "Induction Motor Performance as A Function of Phase Number," *IEE Proc. – Electr. Power Appl., Vol. 153, No. 6, November 2006*.

[9] E. Levi, R. Bojoi, F. Profumo, H. A. Toliyat, and S. Williamson, "Multiphase Induction Motor Drives – A Technology Status Review," *IET Power Electr., vol. 1, No. 4, pp. 489 – 516, July 2007*.

[10] E. Levi, "Multiphase Electric Machines for Variable-Speed Applications," *IEEE Trans. Ind. Electr. , vol. 55, No. 5, May 2008.*

[11] M. Gaspari, "Life Prediction Model for Aluminum Electrolytic Capacitors", IEEE Ind. Appl. Soc. Ann. Meeting, 1996.

[12] Cross, A.M., Evans, P.D., Forsyth, A.J.,"DC Link Current in PWM Inverters with Unbalanced and Non-Linear Loads," *IEE Proc.-Electr. Power Appl., Vol. 146, No. 6, November 1999*.

[13] Ziogas, P., Photiadis, P. N. D., "An Exact Input Current Analysis of Ideal Static PWM Inverters," *IEEE. Trans. on Industry Applications, Vol. 1A-19, No. 2, March/April 1983*.

[14] Bojoi, R., Caponet, M.C., Grieco, G., Lazzari, M., Tenconi, A., Profumo, F., "Computation and measurements of the DC link current in six-phase voltage source PWM inverters for AC motor drives",*Proceeding of the Power Conversion Conference, 2002. PCC - Osaka 2002.*

[15] B. P. McGrath and D. G. Holmes,, "A General Analytical Method for Calculating Inverter DC-Link Current Harmonics," *IEEEIndustry Appl. Soc. Ann. Meeting, 2008*.

[16] P. A. Dahono, Y. Sato, and T. Kataoka, "Analysis and Minimization of Ripple Components of Input Current and Voltage of PWM Inverters," *IEEE Trans. Ind. Appl., Vol. 32, No. 4, July/August 1996.*

[17] P. A. Dahono, Deni, C. P. Akbarifutra, and A. Rizqiawan, "Input Ripple Analysis of Five-Phase Pulse Width Modulated Inverters," *IET Power Electron, vol. 3, Iss. 5, 2010*.

[18] P. A. Dahono, A. Satria, A., and D. Nurafiat, "Analysis of DC Current Ripple in Six-Legs Twelve-Devices Inverters," *IEEE Conf. on Power Eng. and Renew. Energy, 2012.*

Inter-Area Power Oscillation Identification Using Synchronized Ambient and Ringdown Data

Husni Rois Ali
Department of Electrical Engineering and Information Technology
Gadjah Mada University
Jalan Grafika No 2, Kampus UGM Yogyakarta Indonesia 55281
Email: husni.rois.ali@gmail.com

***Abstract*—This paper presents inter-area power oscillation identification based on the synchronized PMU data. The identification is carried out during both ringdown and ambient condition. Prony analysis is used to identify the mode during ringdown condition. While, modified Yule-Walker (MYW) is deployed during ambient condition. To verify the result of identification, it is compared to modal analysis result. The benchmark two-area test system is used to demonstrated the identification process. The result shows that it is close to result from modal analysis. Thus, it indicates that both algorithms are promising for online monitoring of inter-area power oscillation.**

***Keywords*—System identification, Inter-area power oscillation, synchronized data, Prony analysis, modified Yule-Walker.**

I. INTRODUCTION

The problem of inter-area power oscillation has been long understood. Inter-area power oscillation occurs when an area in power system oscillates against another area with the frequency ranging from 0.1 to 0.8 Hz [1]. In case the oscillation is poorly damped, it can cause major power system blackout as being reported in [2].

In order to study the problem of inter-area power oscillation, the modal analysis is normally used [3]. Modal analysis requires detailed model of power system and then linearize the model around particular operating point. The system is stable if all eigenvalues lie on left half plane of complex coordinate.

However inter-area power oscillation is varying phenomenon. It depends on the power system operating point, system structure, etc. Thus, power system monitoring is important to keep tracking the oscillation activity in real-time. Once the low damping is detected, countermeasure can be taken to prevent blackout. Unfortunately, modal analysis cannot be used to monitor the oscillation in real-time due to, at least, two reasons. First, Obtaining the detailed model of power system is very hard. Second, modal analysis can be considered very slow for real-time monitoring due to numerical reason. These two problems are more noticeable as the size of the system increases, i.e. the case of practical power system.

To cope this problem, measurement-based methods has been widely used for monitoring inter-area power oscillation [4]–[7]. Basically, It is based on system identification techniques. By using power system measurement such as current, bus voltage, current, etc., measurement-based methods estimate mode of oscillations. In addition, the advent of Phasor Measurement Unit (PMU) provides more accurate, faster-sampling, and more reliable measurement data. These favorable factors make measurement-based methods more promising for oscillation monitoring.

Zhou., N., et al [8] categorizes the oscillation identification algorithm according to the power system data. In general, power system measurement can be divided into typical and non-typical data. Typical data provides information about power system oscillation therefore it can be used for identification purposes. While, non-typical data does not contain any information regarding the oscillation. As a result, it cannot be used for mode identification. Typical data is further categorized into ringdown and ambient data. Ringdown data is characterized by strong oscillation due to large disturbance such as short-circuit or loss of big load. Meanwhile, ambient data is produced by small random excitation of load. Therefore, algorithm classification is based on the data type i.e. ringdown algorithm and ambient algorithm.

To properly identify the oscillation modes, a particular data type must use proper algorithm. If it is not fulfilled, the results may be inaccurate. This paper demonstrates how to identify oscillation modes of power system during ringdown and ambient condition. Prony analysis [9], a well-known algorithm belong to ringdown algorithm, is utilized to identify oscillation mode during ringdown condition. While, during ambient condition, Modified Yule-Walker (MYW) [10] is used. The data length or window length and noise presence are studied to see the effect to both algorithm's result.

This paper is organized as follow. Section II presents the detail of Prony analysis and MYW algorithm. Performance verification is given in Section III. While, result and discussion are provided in Section IV. Section V ends by underlining some important results of this paper.

II. ALGORITHM FOR OSCILLATION IDENTIFICATION

A. Prony Analysis

power system is naturally highly nonlinear. However, in normal operation, it can be considered as linear system [9]. As a result, the response can be expressed as

$$y(t) = \sum_{i=1}^{n} c_i \exp(\lambda_i t) \quad (1)$$

where $y(t)$ denotes measurement at time t, λ_i and c_i are the i^th eigenvalues and its amplitude, respectively. Since we measure in discrete, (1) can be expressed as

$$y[k] = y(t) \mid_{t=k\Delta t} = \sum_{i=1}^{n} c_i z_i^k, \text{ where } z_i = \exp(\lambda_i \Delta t) \quad (2)$$

Δt is sampling interval. After N data measurement, (2) can be formulated as

$$\begin{bmatrix} y[n] & y[n-1] & \cdots & y[0] \\ y[n+1] & y[n] & \cdots & y[1] \\ y[n+2] & y[n+1] & \cdots & y[2] \\ \vdots & \vdots & \ddots & \vdots \\ y[N-1] & y[N-2] & \cdots & y[N-n-1] \end{bmatrix} = \begin{bmatrix} c_i z_1^0 & c_2 z_2^0 & \cdots & c_n z_n^0 \\ c_1 z_1^1 & c_2 z_2^1 & \cdots & c_n z_n^1 \\ c_1 z_1^2 & c_2 z_2^2 & \cdots & c_n z_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ c_1 z_1^{N-n-1} & c_2 z_2^{N-n-1} & \cdots & c_n z_n^{N-n-1} \end{bmatrix}^* \begin{bmatrix} z_1^n & z_1^{n-1} & \cdots & z_1^0 \\ z_2^n & z_2^{n-1} & \cdots & z_2^0 \\ \vdots & \vdots & \ddots & \vdots \\ z_n^n & z_n^{n-1} & \cdots & z_n^0 \end{bmatrix} \quad (3)$$

where n is estimated system order. Note that (3) is Vandermonde matrix. Thus, it has property

$$z^n + a_1 z^{n-1} + a_2 z^{n-2} + \cdots + a_n z^0, \text{ for } i = 1, 2, \cdots, n \quad (4)$$

From (3) and (4) , it can be obtained

$$\begin{bmatrix} y[n-1] & y[n-2] & \cdots & y[0] \\ y[n] & y[n-1] & \cdots & y[1] \\ y[n+1] & y[n] & \cdots & y[2] \\ \vdots & \vdots & \ddots & \vdots \\ y[N-2] & y[N-3] & \cdots & y[N-n-1] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix} = - \begin{bmatrix} y[n] \\ y[n+1] \\ y[n+2] \\ \vdots \\ y[N-1] \end{bmatrix} \quad (5)$$

The coefficient a_i can be obtained by using overdetermined equation with N is normally set to be greater than 2*n [9]. In practice, the measurement $y(k)$ will be corrupted by noise. Let $\hat{y}[k]$ be measurement after adding the noise, (5) turns to be

$$\begin{bmatrix} \hat{y}[n-1] & \hat{y}[n-2] & \cdots & \hat{y}[0] \\ \hat{y}[n] & \hat{y}[n-1] & \cdots & \hat{y}[1] \\ \hat{y}[n+1] & \hat{y}[n] & \cdots & \hat{y}[2] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}[N-2] & \hat{y}[N-3] & \cdots & \hat{y}[N-n-1] \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \vdots \\ \hat{a}_n \end{bmatrix} = \begin{bmatrix} \hat{y}[n] \\ \hat{y}[n+1] \\ \hat{y}[n+2] \\ \vdots \\ \hat{y}[N-1] \end{bmatrix} + \begin{bmatrix} e[n] \\ e[n+1] \\ e[n+2] \\ \vdots \\ e[N-1] \end{bmatrix} \quad (6)$$

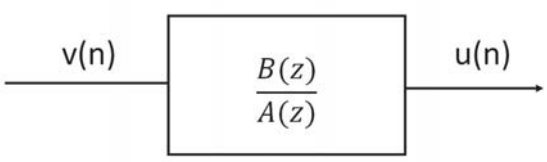


Fig. 1. ARMA model with white-noise input

Consequently, (3) becomes

$$\hat{z}^n + \hat{a}_1 \hat{z}^{n-1} + \hat{a}_2 \hat{z}^{n-2} + \cdots + \hat{a}_n \hat{z}^0, \text{ for } i = 1, 2, \cdots, n \quad (7)$$

Let $\hat{z}_i$ be the roots of polynomial equation in (7). Thus the estimated mode can be obtain by

$$\hat{\lambda}_i = \frac{1}{\Delta t} \ln \hat{z}_i \quad (8)$$

B. Modified Yule-Walker

In modified Yule-Walker (MYM) method [10], the system is represented as ARMA model shown in Fig.1. The input $v(n)$ is assumed as random signal. The input-output can be related by

$$\begin{aligned} &u(n) + a_1 u(n-1) + \cdots + a_N u(n-N) \\ &= b_0 v(n) + b_1 v(n-1) + \cdots + b_M v(n-M) \end{aligned} \quad (9)$$

where $a_1, \cdots, a_N$ and $b_1, \cdots, b_M$ denote the coefficient for characteristic equation of poles and zeros, respectively.

The poles coefficient $a_1, \cdots, a_N$ can be derived from covariance of the data, $r(k) = Eu(n)u(n-k)$. Let multiply both sides of difference equations by $u(n-k)$. It gives

$$r(k) = \sum_{i=1}^{N} a_i r(k-i) = \sum_{j=0}^{M} b_j E v(n-j) u(n-k) \quad (10)$$

For $k > m$ the output is uncorrelated with future input. Thus, the expectation will be zero. Equation (10) is then simplified to

$$\begin{aligned} &r(k) + \sum_{i=1}^{N} a_i r(k-i) = 0 \\ &\text{for } k > M \end{aligned} \quad (11)$$

Equation (11) can be expanded for $k = M+1, M+2, \cdots M+P$ in matrix equation

$$-\underline{r} = R\underline{a} \quad (12)$$

P is the number of equations being used to estimate AR coefficient.

Since the real values of autocorrelation are unknown, it can be approximated from sample data

$$\hat{r}(k) = \frac{1}{N} \sum_{n=k}^{N-1} u(n)u(n-k) \quad (13)$$

The matrix equation in (12) is usually over determined which is well-known as Modified Yule-Walker (MYW) method. In this case, the order of P must be strictly greater than N. Then, the estimated poles can be calculated by finding the root of characteristic equation similar to [7], and it can be transformed in s-domain using [8]

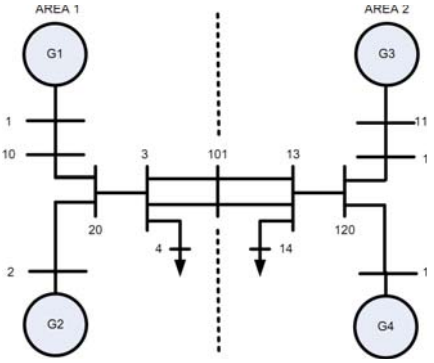


Fig. 2. Two-area test system

III. Performance Verification

This section elaborates test system used for simulation and also the test procedure. The detail is as follow

A. Test System

The two-area test system is used to test the algorithm [1]. It consists two areas with two generators in each area. All generators are modeled as sub-transient model equipped by exciter, PSS, and speed governor. Although the system is quite small, the two-area system behaves similar to large system. Using modal analysis tool in Power System Toolbox (PST) [11], it can be found out that the system has one inter-area mode given in Table I. This mode is the low frequency oscillation between area I and II. The complete system is depicted in Fig. 2.

B. Test Procedure

The two-area test system is simulated for ringdown and ambient condition. In order to produce ringdown data, the system is perturbed using line-to-ground fault at t= 2 s with duration 0.2 s. The fault is applied at bus 101. While, to produce ambient condition, the load is characterized by random Gaussian white noise. It is hypothesized that the representation is sufficient to resemble random load changing of practical power system [12].

Prony analysis is utilized to extract the inter-area mode during the ringdown condition. Meanwhile, the MYW is used during the ambient condition. The result of the identification of two algorithms is then compared to modal analysis's result in Table I to check the accuracy of the result. The complete step of this research is illustrated in the Fig. 3.

Since all identification algorithms rely on the data, it can be understood that the factors affecting the data will also affect the output. Therefore, the research considers the effect of window length and noise presence on the data to identification results of Prony analysis and MYW.

TABLE I. Inter-area mode from modal analysis

Eigenvalue	Frequency (Hz)	Damping (%)	Interaction
-0.5249±3.849*i*	0.61249	13.513	area 1 vs area 2

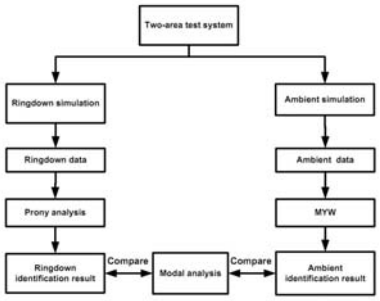


Fig. 3. Test procedure

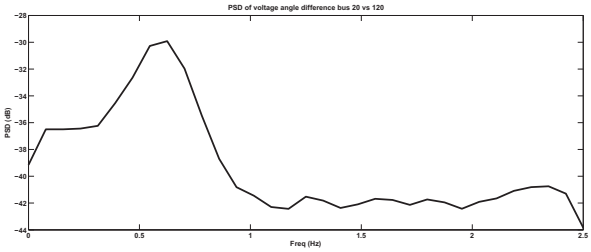


Fig. 4. PSD of voltage angle different bus 20 vs 120

IV. Test Results

According to the scenario presented in Section III, the system is simulated in time-domain analysis using PST. The data is then captured from the simulation and it is set to have sampling rate 5 Hz to mimic PMU data [8]. There is no need to use very high sampling rate since the inter-area mode has always low frequency. It is also necessary to preprocess the data to improve the results of identification. In this research, the preprocessing is performed by removed the offside and trend of the data.

In addition, the input data for identification should be selected properly. Normally, the global measurement is more preferable than local measurement. To study the quality of the measured signal, Power Spectral Density (PSD) of the corresponding signal may be used. Good signal for identification will have significant peak around the inter-area frequency. It indicates that the signal has significant energy about the inter-area frequency. In this research, voltage angle different bus 20 vs bus 120 is chosen. The PSD of this signal is presented in the Fig. 4, and it can be observed that the graph has significant peak around 0.6 Hz. It is corresponding to the frequency of inter-area mode in Table I. In practical application, voltage angle different among the bus in the system can be obtained easy from PMU measurement. Thus, this signal is suitable as the input of identification. Fig. 5 and 6 present the voltage angle different bus 20 vs 120 during ambient and ringdown condition, respectively. It can be observed that during ambient condition the system response is dominated by small random fluctuation. While, during ringdown condition, the signal is characterized by strong oscillation after the fault is inserted to the system, i.e. after t = 2 s.

Since the true order of the system is assumed unknown, the order of Prony analysis and MYW is estimated. Note that

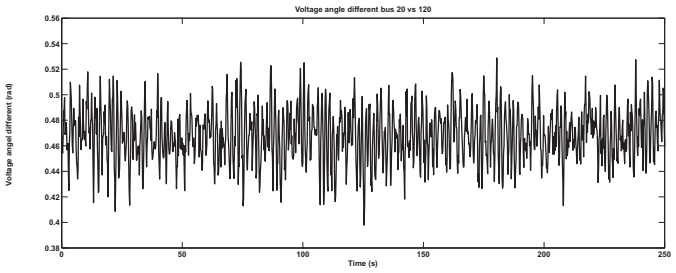


Fig. 5. Ambient data of voltage angle different bus 20 vs 120

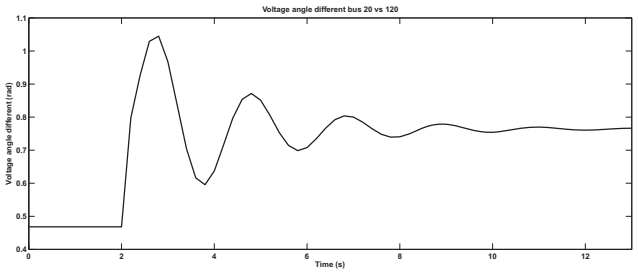


Fig. 6. Ringdown data of voltage angle different bus 20 vs 120

in practical application, the true order of system of interest is also often unknown. According to Section II, Prony analysis requires input of estimated system order n. While, MYW need three estimated parameters, i.e. N, M, and P. In this research, the parameter of Prony analysis n is set to 40. Whereas, the order of N, M, and P are chosen to 20, 5, and 30, respectively.

The following section will present the effect of window length and the presence of noise to identification result. The effect of these factors will be studied during both ambient and ringdown condition.

A. *Effect of window length*

The effect of window length for MYW and Prony analysis is provided in Table II and III, respectively. It can be concluded that MYW, in general, requires much longer window length than that of Prony analysis. The data for MYW is in order of minutes, whereas for Prony is normally only few seconds, i.e. during the strong oscillation.

Table II reveals variation of window length from 450 s to 750 s. It turns out that longer data provides more accurate result. However, very short data, i.e. data length 450 s, will produce inaccurate result. It can be also drawn a conclusion that estimating the damping of oscillation is more challenging that estimating the frequency. The result can be readily compared to the result of modal analysis in Table I to check the accuracy. It can be observed that the result is quite accurate.

For Prony analysis in Table III, it can be seen that longer data does not guarantee accurate result. It is caused by the nature of the data. Prony analysis analyses ringdown data therefore it should be only used ringdown period. From Fig. 6, ringdown data occurs between t= 3 s and t= 12 s. Other than this period, the oscillation is very small and it may be considered as ambient data. Therefore, Prony analysis should

TABLE II. WINDOW LENGTH EFFECT ON MYW'S OUTPUT

Window length (s)	Frequency (Hz)	Damping (%)
450	0.5846	14.51
500	0.5934	14.15
550	0.5931	14.85
600	0.5966	12.59
650	0.5984	12.52
700	0.600	12.21
750	0.5907	11.28

TABLE III. WINDOW LENGTH EFFECT ON PRONY'S OUTPUT

Window length (s)	Frequency (Hz)	Damping (%)	Time
8	0.6402	9.49	t= 3-11 s
9	0.6162	13.20	t= 3-12 s
10	0.7053	6.02	t= 3-13 s

be used only during ringdown period. In case ambient data is used, it will deteriorate the output. According to Table III, the accurate result is obtained only when ringdown data is used, i.e. during t= 3 s and t= 12 s.

It should also be observed that accurate result for Prony analysis is attained by only using 9 s window length. This window length is significantly shorter than MWY's data length. For MYW, accurate result is resulted for the window length around 600 s and 900 s. This occur since rindown data contains more information about the system oscillation that that of ambient data. The rich information is stored in strong data oscillation. Whereas, ambient data has small and random fluctuation. As a result to obtain accurate result, Prony method need shorter window length.

All above mentioned factors make the MYW can be used on the normal data during small and random load changing. While, Prony analysis may only be deployed after large fault on the system such as short circuit. Nevertheless, the prony analysis requires significantly short ringdown data.

B. *Effect of noise presence*

The presence of noise in the power system measurement cannot be completely avoided. It can heavily affect the final result of identification. In practical application, the presence of noise can be treated by filtering in order to minimize the noise. This section discusses the effect of noise presence on identification result of Prony analysis and MYW.

Table IV and V give the result of noise effect on MYW and Prony's result, respectively. The window 700 s window length is set as base case, i.e. noise-free, for MYW. While, the 9 s data length is selected as base case for Prony analysis. To simulate the noise presence, the white Gaussian noise is added to measurement. The noise energy is measured as Signal to Noise Ratio (SNR) in dB. The SNR is varied to 10 dB, 20 dB, 30 dB, and 40 dB. The greater SNR value means the noise power is less significant in the signal.

According to Table IV, as the SNR increases, the identification result becomes more and more accurate. It can also be observed that the noise will make the identification of mode damping more difficult than that of mode damping. The case also remain similar to the case of Prony analysis. However, Prony analysis suffer more erroneous result than the case of MYW. It is because the MYW uses significantly longer data.

TABLE IV. NOISE EFFECT ON MYW'S OUTPUT

SNR (dB)	Frequency (Hz)	Damping (%)
Noise-free	0.600	12.21
10	0.86	13.51
20	0.6879	19.88
30	0.5957	17.8
40	0.599	11.00

TABLE V. NOISE EFFECT ON PRONY'S OUTPUT

SNR (dB)	Frequency (Hz)	Damping (%)
Noise-free	0.6162	13.20
10	0.639	2.58
20	0.6146	6.24
30	0.629	9.82
40	0.600	11.3

V. CONCLUSION

The paper has demonstrated the identification of inter-area power oscillation during both ambient and ringdown condition. The result is compared to modal analysis to examine the accuracy. The MYW is used to identify the mode during ambient condition. While, the Prony analysis is used for ringdown condition. The result of both algorithm is consistent with modal analysis. Thus, it is very promising to be used for real-time application. It can be also seen that window length and noise presence significantly affect the result.

REFERENCES

[1] M. Klein, G. Rogers, and P. Kundur, "A fundamental study of inter-area oscillations in power systems," *IEEE Trans. Power Syst*, vol. 6, no. 3, pp. 914–921, 1991.

[2] B. Pal and B. Chaudhuri, *Robust control in power systems*. Springer Science+ Business Media, 2005.

[3] G. Rogers, *Power Systems Oscillations*. Kluwer Academic, 2000.

[4] Z. Ning, J. W. Pierre, D. J. Trudnowski, and R. T. Guttromson, "Robust RLS methods for online estimation of power system electromechanical modes," *IEEE Trans. Power Syst*, vol. 22, no. 3, pp. 1240–1249, 2007.

[5] J. W. Pierre, D. J. Trudnowski, and M. K. Donnelly, "Initial results in electromechanical mode identification from ambient data," *IEEE Trans. Power Syst*, vol. 12, no. 3, pp. 1245–1251, 1997.

[6] D. Trudnowski, "Estimating electromechanical mode shape from synchrophasor measurements," *IEEE Trans. Power Syst*, vol. 23, no. 3, pp. 1188–1195, 2008.

[7] D. Trudnowski, J. Pierre, N. Zhou, J. Hauer, and M. Parashar, "Performance of three mode-meter block-processing algorithms for automated dynamic stability assessment," *IEEE Trans. Power Syst*, vol. 23, no. 2, pp. 680–690, 2008.

[8] N. Zhou, Z. Huang, F. Tuffner, J. Pierre, and S. Jin, "Automatic implementation of prony analysis for electromechanical mode identification from phasor measurements," in *Power and Energy Society General Meeting, 2010 IEEE*. IEEE, 2010, pp. 1–8.

[9] J. F. Hauer, C. Demeure, and L. Scharf, "Initial results in prony analysis of power system response signals," *Power Systems, IEEE Transactions on*, vol. 5, no. 1, pp. 80–89, 1990.

[10] R. Wies, J. Pierre, and D. Trudnowski, "Use of ARMA block processing for estimating stationary low-frequency electromechanical modes of power systems," *IEEE Trans. Power Syst*, vol. 18, no. 1, pp. 167–173, 2003.

[11] J. Chow and K. Cheung, "A toolbox for power system dynamics and control engineering education and research," *IEEE Trans. Power Syst*, vol. 7, no. 4, pp. 1559–1564, 1992.

[12] A. Hauer, D. Trudnowski, and J. DeSteese, "A perspective on wams analysis tools for tracking of oscillatory dynamics," in *Power Engineering Society General Meeting, 2007. IEEE*. IEEE, 2007, pp. 1–10.

Macro Demand Spatial Approach (MDSA) with Principal Component Analysis (PCA) on Spatial Demand Forecasting for Industrial Area in Transmission Planning

Sudarmono Sasmono
School of Electrical Engineering and Informatics
Bandung Institute of Technology
Bandung, Indonesia
udar@power.ee.itb.ac.id
IEEE Student Member # 92352179

Mukmin W Atmopawiro
School of Electrical Engineering and Informatics
Bandung Institute of Technology
Bandung, Indonesia
mukmin@power.ee.itb.ac.id

Ngapuli Irmea Sinisuka
School of Electrical Engineering and Informatics
Bandung Institute of Technology
Bandung, Indonesia
ngapuli@hv.ee.itb.ac.id

Djoko Darwanto
School of Electrical Engineering and Informatics
Bandung Institute of Technology
Bandung, Indonesia
djoko@hv.ee.itb.ac.id

Abstract— **Macro Demand Spatial Approach (MDSA) is an approach introduced in long time electricity demand forecasting considering location. It will be used at transmission planning and policy decision on electricity infrastructure development in a region. In the model, MDSA combined with principal component analysis (PCA) method to determine the variables that affecting electricity demand in industrial area. The variables are different for each load sector. Hypothesis on unique variables affecting electricity demand on every load sector in the industrial area were analyzed with qualitative methods and references. The variables have no significant effect can be reduced by using PCA. The generated models tested to assess whether it still at the range of confidence level of electricity demand forecasting. At the case study, generated model for South Sumatra Subsystem as a part of Sumatra System is still in the range of confidence level.**

Keywords : macro demand spatial approach; principal component analysis; electricity demand forecasting; industrial area; transmission planning

I. INTRODUCTION

Transmission planning requires long-term electric load forecast in the future. Unlike the electricity demand for generation planning, transmission planning requires electricity demand projections that considering the location of electrical load, their magnitude and time when load occurs. Model of electricity demand projections be in accordance with their characteristics known as spatial load forecasting model. The approach was first introduced by Van Wormer in 1954 [12]. Furthermore the approach has been developed in electricity distribution planning.

In electricity distribution planning, spatial load forecasting conducted for small areas and short time forecasting. On the model, the electricity demand driver is change in land use [5]. The models were developed then, basically find the relationship between changes in land use with demand and electrical loads in a spatial area [1], [3], [6], [7].

Spatial load forecasting approach on transmission planning is different approach with spatial load forecasting approach on distribution planning. Characteristics of spatial load forecasting on transmission planning is broader area to long time forecasting. The driver on long term electricity demand forecasting has been proved to have a strong correlation with it is economic variables [2]. The previous publications had proposed the use of a spatial approach with economic variables as a driver in the spatial load forecasting for transmission planning [11] and decision making on electricity infrastructure development in an area [10]. Due to use of economic variables as a driver, then the spatial load forecasting for the transmission planning must started with energy forecasting. This approach be introduced as Macro Demand Spatial Approach (MDSA).

Due to spatial approach, the region is served by electrical system with the transmission network can be divided into regions with direction towards the development of which is dominated by industrial sectors or non-industrial sectors. Therefore, the economic variables that influence electricity demand in the region developed in the direction of a predominantly industrial area would be distinctive. The paper conducted deterministic process combined with the qualitative process to determine unique variables in industrial areas. Case

study conducted in South Sumatra subsystem. It is a part of Sumatra Interconnection System.

II. METHOD

A. *Selection of the appropriate spatial characteristics*

Determination on the characters of the area that will be served by the transmission system, conducted using a qualitative approach. In qualitative approach, a number of document analysis is done to obtain the characteristics of the region that will be served. This research focused on industrial area. Economic variables which hypothesized affect electricity demand in the industrial area defined base on reference.

B. *Formation of Regression Model*

Regression model was formed using the method of principal component analysis (PCA). In general, the formed regression model has general equation following

$$yt = \alpha\, pt + \beta\, qt + \ldots + \gamma\, rt \quad (1)$$

Where, y = electricity demand in year-*t*; p,q,r = economic variables affect electricity demand in year-*t;* α, β, γ = regression coefficient. The goal of principal component analysis is to identify the most meaningful basis to re-express a data set. The hope is that this new basis will filter out the noise and reveal hidden structure. The explicit goal of PCA is to determine: "the dynamics are along the x-axis." In other words, the goal of PCA is to determine that ˆx, i.e. the unit basis vector along the x-axis, is the important dimension [8]

Assuming the independent variables X, P, Q and so on are a set of data, then equation (1) can be transformed into the form of a matrix X (n x p) which satisfies the equation (2).

$$Y = \delta^T X = \delta_1 X_1 + \delta_2 X_2 + \ldots + \delta_p X_p \quad (2)$$

Where $\delta = (\delta 1, \delta 2, \ldots, \delta p)T$ are a column vector of weights with $\delta 1^2 + \delta 2^2 + \ldots + \delta p^2 = 1$. Find δ after maximize the variance of the projection of the observations on the Y variables as following:

$$\mathrm{Var}(\delta T\, X) = \delta T\, \mathrm{Var}(X)\, \delta \text{ is maximal} \quad (3)$$

The matrix C=Var(X) is the covariance matrix of the Xi variables, where:

$$C = \begin{pmatrix} v(x_1) & c(x_1,x_2) & \ldots\ldots & c(x_1,x_p) \\ c(x_1,x_2) & v(x_2) & \ldots\ldots & c(x_2,x_p) \\ c(x_1,x_p) & c(x_2x_p) & \ldots\ldots & v(x_p) \end{pmatrix} \quad (4)$$

The direction of δ is given by the Eigen vector γl corresponding to the largest Eigen value of matrix C. The second vector that is orthogonal (uncorrelated) to the first is the one that has the second highest variance which comes to be the Eigen vector corresponding to the second Eigen value.

So, the PCA give new variables Y*i*. The new variables are linear combination of the original variables (X*i*) following:

$$Y_i = a_{i1}x_1 + a_{i2}x_2 + \ldots a_{ip}x_p \; ; i = 1..p \quad (5)$$

The new variables Yi are derived in decreasing order of importance. They are defined as 'principal components'.

III. SIMULATION ON SOUTH SUMATERA SUBSYSTEM

Simulation was done on industrial area in the electrical power system that is being developed. Electrical power systems that are being developed in Indonesia is Sumatra System. The Sumatera System consist of 8 subsystems. They are Aceh, North Sumatera, West Sumatera, Riau, Jambi, South Sumatera, Bengkulu and Lampung. They are shown in Fig 1. Subsystem selected in this research is South Sumatra. It serves South Sumatra Province. The center of development growth in the province is in Palembang City and its surrounding. According to the Documents of MP3EI, development direction of South Sumatra Province are: (1) Coal mining center (90% of coal reserves in Sumatra) and (2) Coal mining industry became a driver for economic growth in the region [4]. Generally the condition indicates that the South Sumatra Province is the industrial area province. Electricity demand forecasting in this area should be spatially where variables will affect demand are different with the variables in the non-industrial area.

Figure 1. The Sumatera System with 8 subsystem. (Source : PT PLN (Persero) P3BS)

Variables that affect the electricity demand in each sector are different. Variables thought to affect the electricity demand at each sector in Indonesia is shown in Table 1. These variables are based on the assumptions used in the modeling of electricity demand projections by Stoll [9].

If the total electricity demand = Y, electricity demand on residential sector = Y_1, electricity demand on commercial sector = Y_2, electricity demand on industrial sector = Y_3, and electricity demand on public facility = Y_4, then linear regression equation of electricity demand in spatial unit, *s* are following:

$$\begin{cases} Y_1 = \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_3 + \delta_4 X_4 + \delta_5 X_5 \\ Y_2 = \delta_6 X_6 + \delta_7 X_7 \\ Y_3 = \delta_8 X_8 + \delta_9 X_9 \\ Y_4 = \delta_{10} X_{10} + \delta_{11} X_{11} \\ Y = Y_1 + Y_2 + Y_3 + Y_4 \end{cases} \quad (6)$$

TABLE I. VARIABLES THOUGHT TO AFFECT THE ELECTRICITY DEMAND AT EACH SECTOR IN IDONESIA

SECTOR	VARIABLES	SYMBOL
RESIDENTIAL	Population	X_1
	Household	X_2
	GDP	X_3
	GDP per capita	X_4
	Residential Connected Power	X_5
COMMERCIAL	GDP on Commercial Sector	X_6
	Commercial Connected Power	X_7
INDUSTRY	GDP on Industrial Sector	X_8
	Industrial Connected Power	X_9
PUBLIC FACILITY	Regional Revenue	X_{10}
	Public Facility (Schools and Houses of Worship)	X_{11}

Simulations were performed using SPSS 19.0 software. The data used in the simulation are the data for each variable in South Sumatra Province at the period 2006-2010. While the data in 2011 are used to test whether the linear regression equations obtained are still in the range of tolerance. The source of Data are The Statistics of PLN in South Sumatera, Jambi and Bengkulu (S2JB) and The Statistics of South Sumatera Province at the period of 2006 – 2011.

IV. RESULT

A. Residential

Table 2 shows component matrix of the variables affect Y_1. It shows that X_4 has component matrix <0.02. Moreover, value component matrix of X_4 is different than other variables. Therefore, it can be concluded that X_4 has no effect on Y_1

Component score coefficient matrix of each variables form linear regression equation following:

$$Y_1 = = 0.263.X_1 + 0.268.X_2 + 0.263.X_3 + 0.244.X_5 - 27 \quad (7)$$

Comparison between result of projection with equation (7) and actual data Y_1' shown in Fig 2

TABLE II. COMPONENT MATRIX OF VARIABLES Y_1 IN SOUTH SUMATERA

Variable	Component Matrix
X_1	0.951
X_2	0.981
X_3	0.983
X_4	0.018
X_5	0.950

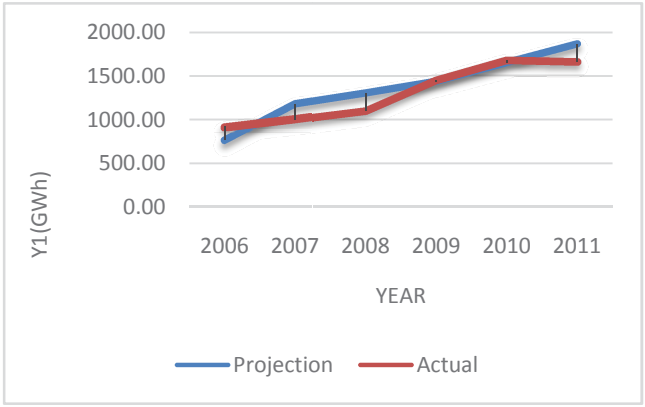


Figure 2. Comparing result of projection (Y_1) and actual data of (Y_1') in South Sumatera

B. Commercial

Table 3 shows component matrix of the variables affect Y_2. As shown in table, all variables has effect on Y_2. Component score coefficient matrix of each variables form linear regression equation following:

$$Y_2 = = 0.337.X_6 + 0.334.X_7 + 1.24 \quad (8)$$

Comparison between result of projection with equation (8) and actual data Y_2' shown in Fig 3

TABLE III. COMPONENT MATRIX OF VARIABLES Y_2 IN SOUTH SUMATERA

Variable	Component Matrix
X_6	0.994
X_7	0.986

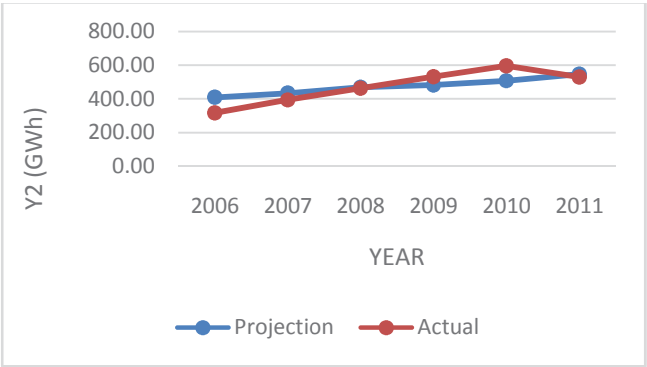


Figure 3. Comparing result of projection (Y_2) and actual data of (Y_2') in South Sumatera

C. Industry

Table 4 shows component matrix of the variables affect Y_3. As shown in table, all variables has effect on Y_3. Component score coefficient matrix of each variables form linear regression equation following:

$$Y_3 = = 0.346.X_8 + 0.338.X_9 + 0.55 \quad (9)$$

Comparison between result of projection with equation (9) and actual data Y_3' shown in Fig 4

TABLE IV. COMPONENT MATRIX OF VARIABLES Y_3 IN SOUTH SUMATERA

Variable	Component Matrix
X_8	0.988
X_9	0.966

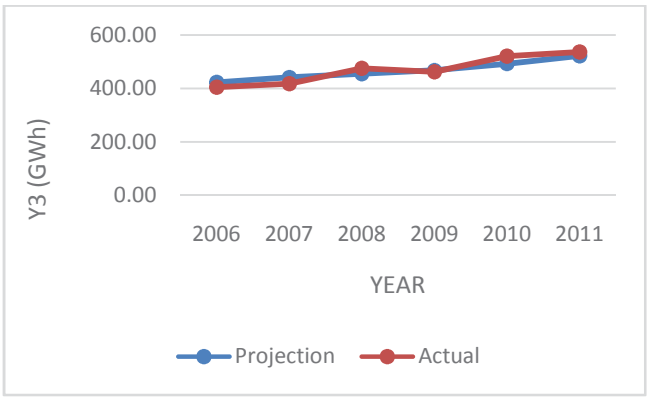


Figure 4. Comparing result of projection (Y_3) and actual data of (Y_3') in South Sumatera

D. Public Facility

Table 5 shows component matrix of the variables affect Y_4. As shown in table, all variables has effect on Y_4. Component score coefficient matrix of each variables form linear regression equation following:

$$Y_4 = = 0.339.X_{10} + 0.343.X_{11} - 78.01 \quad (10)$$

Comparison between result of projection with equation (10) and actual data Y3' shown in Fig 4

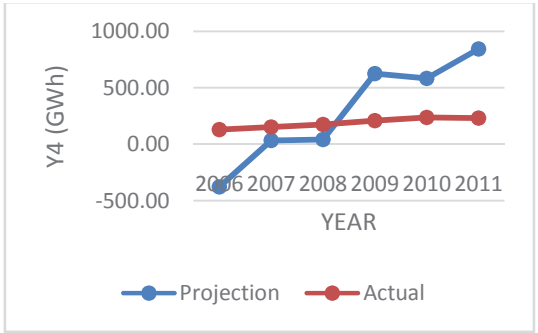


Figure 5. Comparing result of projection (Y_4) and actual data of (Y_4') in South Sumatera

TABLE V. COMPONENT MATRIX OF VARIABLES Y_4

Variable	Component Matrix
X_{10}	0.930
X_{11}	0.941

V. ANALYSIS AND DISCUSSION

Electricity demand forecasting in the industrial sector shows the proximity with the actual data. Thus, in generally X_8 (GDP on industrial sector) and X_9 (connected industrial power) affecting electricity demand forecasting for industrial sector at the spatial unit. On the electricity demand forecasting at the following years, these two variables also need to forecast in advance with certain model.

Electricity demand forecasting in the residential sectors remain at tolerance range. Thus, in generally X_1 (population), X_2 (Household), X_3 (GDP) and X_5 (connected residential power at the spatial unit) affecting electricity demand forecasting for residential sector at the spatial unit. Whilst, X_4 (GDP per capita) has no effect on electricity demand forecasting for residential sector. However, another case study required to be able to generalize the model. Similar with industrial, all variables need to forecast with certain model before carried out electricity demand forecasting for residential sector.

Electricity demand forecasting in the commercial sectors remain at tolerance range. Thus, in generally X_6 (GDP on commercial sector) and X_7 (commercial connected power) affecting electricity demand forecasting for commercial sector at the spatial unit. However, similar with residential sector, another case study required to be able to generalize the model. These two variables also need to forecast in advance with certain model.

Forecasting in the public facility is so different with actual data. Even so, it has not been able conclude that the variable has no effect. Another case study on industrial area required to ensure that.

The simulation shows linear regression approach with principal component analysis can be used to forecast electricity demand in the industrial sector at industrial areas. As for the residential and commercial sectors was still needed testing with non-linear forecasting models. However, the mathematical model obtained from method cannot be used for public facility. On generalizations that apply generally, required comparison with similar methods applied in other industrial area.

VI. CONCLUSION

Spatial electricity demand forecasting for transmission planning can be done using macro demand spatial approach with principal component analysis that applied in the industrial sector at industrial area. Even so, residential and commercial sectors was still required comparison with other methods or other study case at industrial area. Macro demand

spatial approach with principal component analysis model on public facilities at industrial area need improvement models.

VII. REFERENCES

[1] Carreno, E.M, Padilha-Fetrin, Leal, A.G, "Spatial Electric Load Forecasting Using An Evolutionary Heuristic", Revista Controle & Automacao Vol 21 No 4, 2010, pp 379 – 387.

[2] Committee on Electricity in Economic Growth, "Electricity in Economic Growth", National Academy Press, Washington DC, 1986, pp 1 – 13.

[3] Jain, Amit, Jain, M. Babita, "A Novel Hybrid Method for Short Term Load Forecasting using Fuzzy Logic and New Particle Swarm Optimization (NPSO)", 16th National Power System Conference, 2010, pp 132 – 138.

[4] Kementerian Koordinator Bidang Perekonomian Republik Indonesia, "Masterplan Percepatan dan Perluasan Pembangunan Ekonomi Indonesia 2011 - 2025", Jakarta, 2011, pp 51 - 73

[5] Lee Wilis, "Spatial Electric Load Forecasting", 2nd edition, Marcel Dekker Inc, New York, 2002, pp 1-5.

[6] Melo, J.D, Carreno, E.M, Padilha-Fetrin, A, "Multi-Agent Framework for Spatial Load Forecasting", 2011 IEEE Power and Energy Society General Meeting, pp 1-8.

[7] Melo, JD, Carreno, E.M, Padilha-Fetrin, A, Spatial Load Forecasting Using A Demand Propagation Approach, IEEE/PES Transmission and Distribution Conference and Exposition: Latin America (T&D-LA), 2011, pp 1-8

[8] Shlens, Jonathan, "A Tutorial on Principal Component Analysis", Center for Neural Science, New York University, New York City, 2009 pp 1 – 6.

[9] Stoll, Harry G, " Least-Cost Electric Utility Planning", John Wiley & Sons, 1989, pp 193 – 195.

[10] Sudarmono Sasmono, Ngapuli Irmea Sinisuka, Mukmin W Atmopawiro "Applied of Spatial Load Forecasting Model for Regional Infrastructure Electricity Planning in Indonesia", Proceeding of 3rd International Conference on Technology and Operation Management, 2012, pp 340 – 344.

[11] Sudarmono Sasmono, Ngapuli Irmea Sinisuka, Mukmin W Atmopawiro "Alternative Spatial Approach on Spatial Demand Forecasting for Transmission Expansion Planning", Proceeding of 2012 IEEE International Conference on Condition Monitoring and Diagnosis, 2012, pp 577 – 579.

[12] Von Wormer, " Some Aspects of Distribution Load Area Geometry ", Power Apparatus and Systems Vol 73, No 2 1954, pp 1343 – 1349.

Modeling Wind Power Plants in Harmonic Resonance Study – A Case Study in Thailand

Chu Xuan Huan
Department of Electrical Engineering
Faculty of Engineering, Chulalongkorn University
Bangkok 10330, Thailand
Chu.X@student.chula.ac.th

Thavatchai Tayjasanant
Department of Electrical Engineering
Faculty of Engineering, Chulalongkorn University
Bangkok 10330, Thailand
Thavatchai.T@chula.ac.th

***Abstract*—Harmonic resonance phenomenon is a problem paid attention significantly from utilities and system engineers for many years. Nowadays, wind power is the leading renewable source, which is considered as the solution for a clean and sustainable resource; however, this integration of wind power plants into grids brings challenges for harmonic resonance study. This paper compares available methods and their pros and cons for modeling wind power plants in harmonic resonance study. Simulation results are verified with actual measurement data to present the performance among these methods. From the verification, the paper suggests the suitable method for modeling wind power plants in harmonic resonance study.**

Keywords—harmonic resonance; distributed generation; renewable energy; wind power plant; equivalent model

I. INTRODUCTION

Nowadays, an integration of distributed generation into power systems is getting more important in energy policy for each country over the world when the possibility of scarcity of fossil energy resources and pollution problem from power plants running on these resources is forthcoming. However, an integration of distributed generation, especially wind power that is considered as future energy possessing the high speedy growth also brings the complication on power quality, especially in harmonic resonance issues because of applications of electronic devices that are harmonic generating sources. Harmonic resonances in wind power plants (WPPs) mainly result from the capacitive and reactive element of long cables and transformers equipped in the system, especially an installation of capacitor banks.

The purpose of this paper provides general understanding of available methods for equivalent networks of wind power plants and their pros and cons in harmonic resonance study. Verification with actual measurement data is conducted in order to point out the suitable method. The paper is organized as follows. Section II lists available methods for equivalent networks of wind power plants and pros and cons of each method. In section III, the harmonic frequency scan analysis is presented, and harmonic models of network components are introduced as well. The case study from the first wind farm in Thailand is used to demonstrate the suggested method in section IV. The conclusions are summarized in the section V.

II. MODELS OF WIND POWER PLANTS

A. Review on Models of Wind Power Plants

Modeling of wind power plants plays an important role in harmonic assessment while the capacity of wind power plants is getting larger nowadays. A model of wind power plants needs to satisfy imperative requirements that are accuracy and time constraints. Many equivalent models of wind power plants have been proposed; however, some of these methods are still theoretical and need to be verified in practice [1]-[5]. Generally, modeling wind power plants in harmonic assessment can be classified into two main groups as follows [6], [7].

1) Explicit Model

The explicit model demonstrates all the wind turbine generators as complete and detailed models [4], [8], [9]. The method achieves considerable accuracy; nonetheless, the method requires a great amount of time on simulation. A sophisticated model still needs more investigation in order to achieve the objective of simulating the system manner well in harmonic assessment.

2) Aggregate Model

The aggregate model is becoming the noticeable trend in modeling wind power plants. The aggregate model allows simplifying analyses, and minimizing the simulation time. There are some pronounced aggregate methods in following.

The most simple aggregate model is based on IEC 61400-21, and IEC 61000-3-6 standards [10], [11]. The aggregate model represents wind power plants as harmonic current sources with or without the lumped capacitance at the collector [2]. The model is advantageous when considering connection compliance because of the simplicity. However, the model is limited in harmonic assessment, especially harmonic resonance analysis because the impact of medium voltage system is totally neglected [2].

The other model of wind power plants mentioned is based on the reduction method [2]. The method is built on two-port theory and dynamic transformation between two-port model and distributed parameter PI model that is used in modeling transmission lines [12]. The harmonic source is modeled as a harmonic current source, which matches accuracy

significantly in comparison with the explicit model in calculation of individual harmonic voltage distortion. However, the disadvantage of the model is the shift of resonant frequencies, which can result in erroneous emission limit.

The last model is proposed by National Renewable Energy Laboratory (NREL) [13]. The model is based on the loss balance condition to construct an equivalent model. Wind power plants can be modeled including an equivalent transformer connected to the collector bus by the cable systems modeled as a PI model. The model has been validated by verifying with actual measurement data in power flow analysis, stability and electromagnetic transient studies [14]-[16]. Although these studies have done for wind power plants in the range of 50-300 MW, conducting this model in harmonic resonance study is still worth considering. Details of this method will be summarized next.

B. Model of Wind Power Plants

The reduction method has been elaborated well in [2]. Wind power plants modeled in this paper are based on the method proposed by National Renewable Energy Laboratory (NREL) [13]. The procedure of the method can be represented in Fig.1. The equivalent parameters of each feeder are calculated as in

$$Z_{eq} = \frac{\sum_{i=1}^{n} Z_i m_i^2}{\left(\sum_{i=1}^{n} m_i\right)^2}; \quad B_{eq} = \sum_{i=1}^{n} B_i \tag{1}$$

Parameters of an equivalent model of the wind power plants are

$$Z_{eq} = \frac{\sum_{i=1}^{N} Z_{ieq} n_i^2}{\left(\sum_{i=1}^{N} n_i\right)^2}; \quad B_{eq} = \sum_{i=1}^{N} B_{ieq} \tag{2}$$

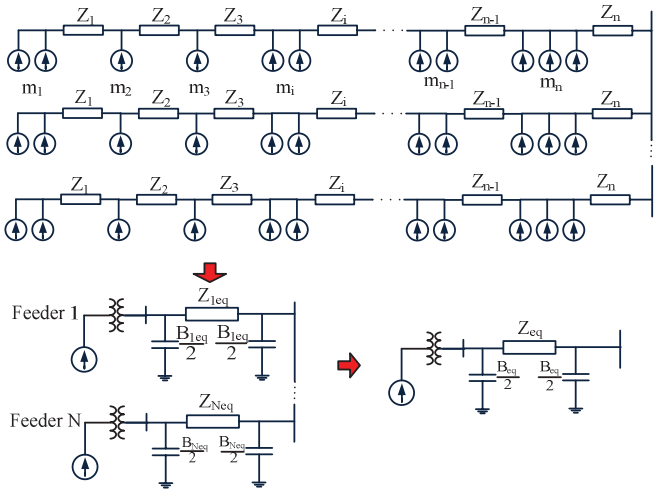


Fig. 1. Equivalent model of a wind power plant

where m_i is the number of wind turbine generators in the i-th group, Z_i & B_i are series impedance and shunt susceptance of each cable of the actual collector system, n_i is the number of wind turbine generators in feeder i-th, and N is the number of feeders.

III. Harmonic Resonance Analysis and Harmonic Modeling

A. Frequency Scan Analysis

The traditional method is conducted to identify harmonic resonances in the system [12]. Frequency scan analysis is based primarily on the linear characteristic of the impedance matrix to represent the system response as a function of frequency as

$$[V]_h = [Z_{bus}]_h [I]_h \tag{3}$$

where h is harmonic order, $[I]_h$ is the vector of the harmonic current source, $[V]_h$ the vector of the harmonic voltage and $[Z]_h$ the harmonic impedance matrix.

To estimate the voltage at the harmonic frequency considered, a current source with value of $1 \angle 0^0$ p.u is injected into the power system at the bus of interest. The voltage bus is proportional to the driving point impedance as seen at the bus respectively. Applying this procedure for the other harmonic frequencies, the curve plotted by the response of driving point impedance with corresponding frequencies helps identify harmonic resonances in the system at the bus of interest. The peaks represent parallel resonance frequencies; the valleys correspond to series resonance frequencies. In practice, the parallel resonance happens more frequently than the series resonance.

B. Harmonic Modeling

This section summarizes harmonic models used in this study. Major system component models are provided as follows.

1) Supply Source Impedance

Supply source impedance is modeled as a series impedance in [12] as

$$Z_s = R_s + jhX_s \tag{4}$$

where R_s, X_s are resistance and reactance at the fundamental frequency of the supply source, and h is harmonic order.

2) Overhead Line and Underground Cable

Overhead line and cable can be modeled as PI or distributed parameter model (exact PI model) [12][17]. The normal PI model is usually applied for short line; the distributed parameter model is used when considering long line effect. The accuracy of each applied model depends on the length of overhead line and cable. The most common difference between overhead line and underground cables is the shunt capacitance. As a result, the long line effect on cables is more significant than that on overhead lines. The parameters of PI model can be calculated from [12].

Normal PI model:

$$Z_{pi} = Z_{normal} = (R_0 + jhX_0)\ell \;;\; Y_{pi} = Y_{normal} = jhB_0\ell/2 \quad (5)$$

Distributed parameter model:

$$\gamma = \sqrt{2Z_{normal}Y_{normal}} \;;\; Z_{pi} = Z_{normal}\sinh(\gamma)/\gamma$$
$$Y_{pi} = Y_{normal}\tanh(\gamma/2)/(\gamma/2) \quad (6)$$

where R_0, X_0, B_0 are resistance, reactance and susceptance respectively of overhead line and cables, and ℓ is length of overhead line and cables.

The skin effect is considered in the simulation. Electricite de France (EDF) proposed equations for overhead lines and cables at the voltage below 225 kV including 150 kV and 90 kV, and harmonic orders are above the second harmonic as follows [12].

Overhead line: $$R = R_0\ell\left(1 + \frac{0.646h^2}{192 + 0.518h^2}\right) \quad (7)$$

Cables: $$R = R_0\ell(0.187 + 0.532\sqrt{h}) \quad (8)$$

3) Transformer

The classical model is used for modeling a transformer in harmonic analysis. The model for a single-phase transformer with no coupling between phases is represented as [12].

$$Z = R\sqrt{h} + jhX \quad (9)$$

where R and X are resistance and reactance of a transformer at fundamental frequency, and h is harmonic order.

IV. Results and Discussion

A. Test System

The test system used in the study is the first wind farm in Thailand, which is located in Nakorn Rachasrima province in Northeastern part of Thailand. The wind farm has capacity of 90 MW consisting of 45 wind turbine generators with rated power of 2.3 MW each. All wind turbine generators are full rated converter induction generator type with terminal voltage of 690 V. The wind farm is connected to medium voltage system at 33 kV by 45 pad-mount transformers and underground cables. The power from wind farm is transmitted to the main grid through two step-up transformers 33/115 kV at the substation and 115 kV transmission overhead line system. The tie circuit breaker at 33 kV bus bar is operated as normally open (NO). Wind farm feeders and network impedance are tabulated in TABLEs II and III in appendix.

There are four meters (M1, M2, M3 and M4) equipped in the wind farm and high voltage system. The procedure of measurement is in accordance with IEC 61400-21standard [10]. The data are collected during 8 days between 5th of March and 13th of March 2013 with 10-minute average data recorded. All the collected data are processed by using MATLAB software for harmonic analysis.

The system configuration and the meter location are illustrated in Fig. 2.

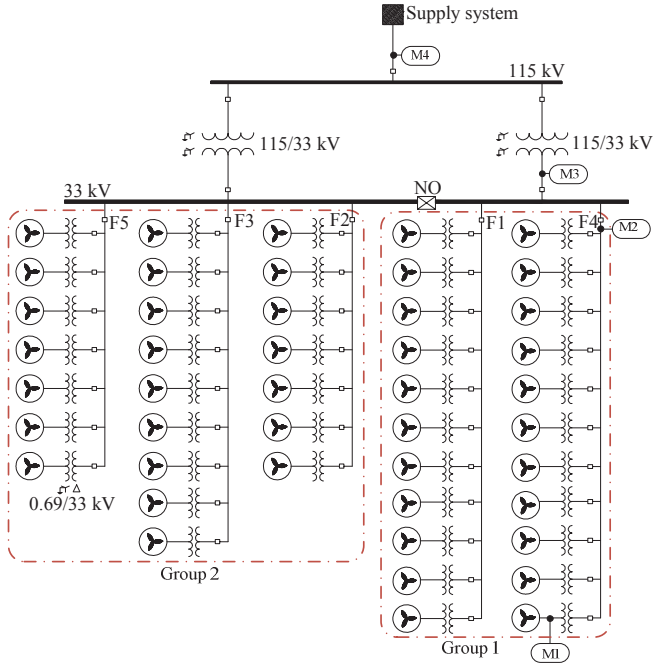


Fig. 2. Test system configuration

B. Simulation Results and Verification

Simulation results using the equivalent model proposed by NREL are performed in two approaches due to different numbers of grouped feeders. The 1st approach models feeders individually; the 2nd approach models wind power plants as two groups corresponding to two groups of the 33 kV bus bar as shown in Fig. 2. The different approaches aim to assess the effect of the number of grouped feeders on harmonic resonances. The 115 kV bus is point of common coupling (PCC), and the 33 kV bus is a collecting bus for the wind power plant. Thus, frequency scan is conducted at these buses. The simulation results performed by using the methods mentioned are crosschecked with the actual measurement data.

1) Frequency Scan Results

The developed program in MATLAB is used for the simulation. The simulation results are shown in Figs. 3 to 5 and summarized in TABLE I.

TABLE I. Harmonic Resonant Frequencies

Method	Harmonic Frequency (p.u)		
	At 115 kV bus	*At 33 kV bus connected to Feeders 1&4*	*At 33 kV bus connected to Feeders2,3&5*
1st approach NREL method	6.7; 24.9; 38.5	6.7; 18.6 ; 25	6.7; 18.6;24.9
2nd approach NREL method	6.7; 24.9; 38.7	6.7; 18.6 ; 25	6.7; 18.6;24.9
Reduction method	6.9	6.9 ; 47.9	6.9
IEC without lumped capacitors	11.2 ; 33.5	11.2 ; 33.5	11.2 ; 33.5
IEC with lumped capacitors	8.2; 23.8;36.3	8.2 ; 36.3	------

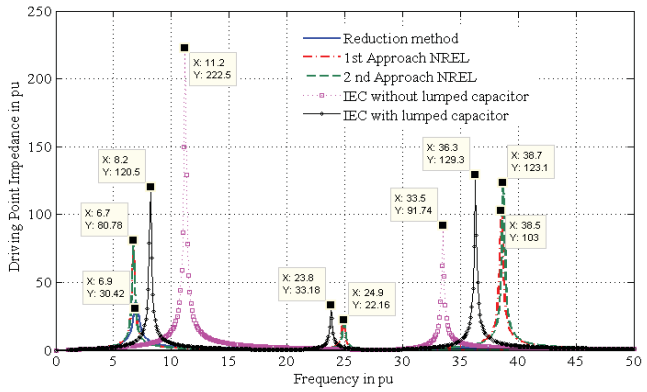


Fig. 3. Frequency scan analysis at 115 kV bus

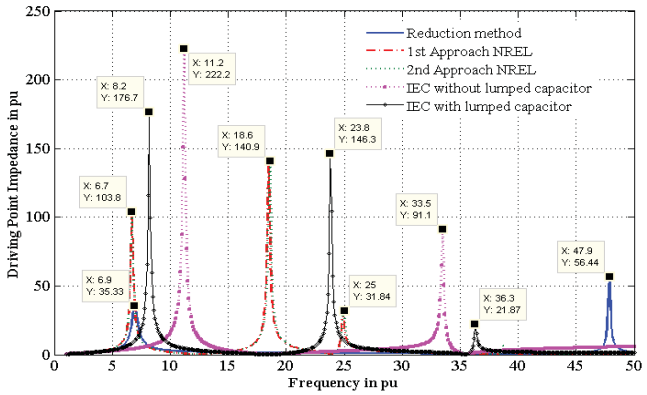


Fig. 4. Frequency scan analysis at 33 kV busbar connected to feeders 1&4

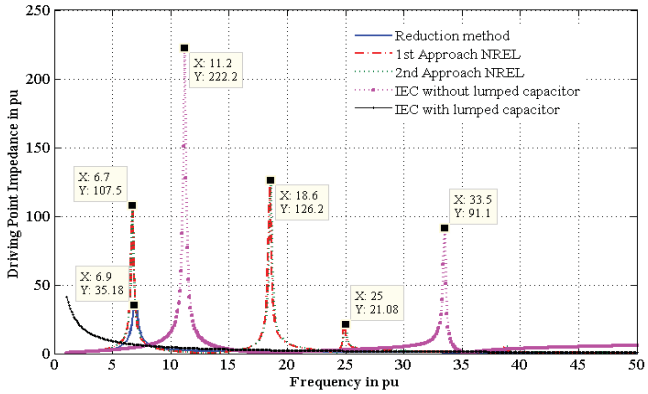


Fig. 5. Frequency scan analysis at 33 kV busbar connected to feeders 2,3&5

As shown in Figs. 3 to 5, harmonic resonances observed show that number of grouped feeders in NREL's model does not affect to harmonic resonance analysis. There are consistent results found by using different approaches in modeling the wind farm. Those resonant frequencies performed by NREL and reduction methods are a good match at frequency between the 6th and 7th harmonics. Nonetheless, besides the frequency at 6.7, there are two other resonant frequencies found by using NREL's model as summarized in TABLE I, and frequency scan analysis also points out another harmonic resonance at 47.9 p.u when the wind power plant model is based on the reduction method. As mentioned in previous sections, those results found by the methods based on IEC standards in harmonic resonance analysis only help conjecture possible harmonic resonant points. As a result, it seems that these results are close agreement to converge to the resonant frequencies found by using the reduction and NREL methods in modeling the wind power plant in both low and high frequency ranges. All the resonant frequencies are crosschecked with actual measurement data in the next section.

2) *Verification and Discussion*

The total harmonic distortion is plotted with active power as shown in Figs. 6 and 7. This index is used as a guide for identifying harmonic resonances.

As observed in Figs. 6 and 7, there is a small fluctuation of the harmonic voltage distortion between 0.5 and 1% with the increase of active power. This observation from Figs.6 and 7 also shows that an impact of power generation output on harmonic distortion is insignificant. Consequently, observing those high voltage distortion points to point out possibility of harmonic resonances is reasonable; these points are analyzed more in details.

The detail observation of changing voltage distortion in the range of output generation less than 10% rated power shows that those high voltage distortion points occurs at the lowest output generation. In other words, there is no impact of the wind farm on harmonic resonances. This phenomenon can be explained when considering effects of the system impedance that have been mentioned in [18].

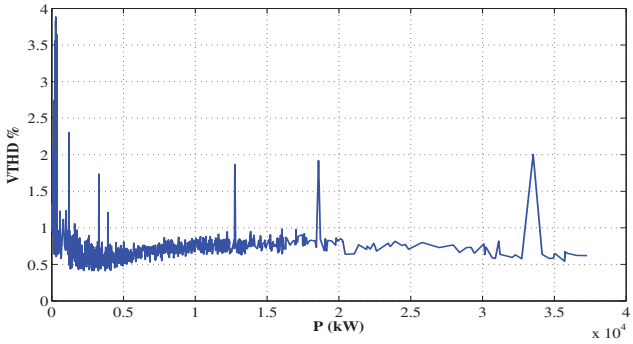


Fig. 6. Variation of harmonic distortion and active power at 33 kV bus - meter M3

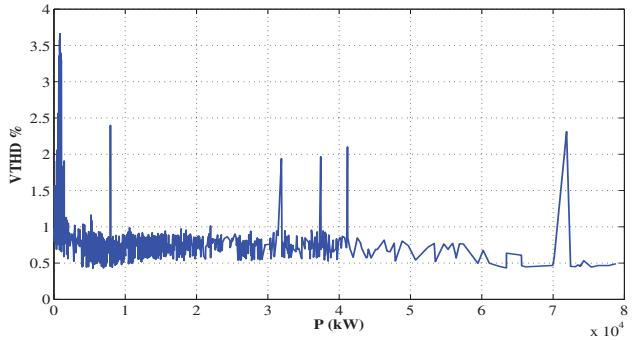


Fig. 7. Variation of harmonic distortion and active power at 115 kV bus - meter M4

Additionally, the influence of the wind farm operation on voltage fluctuation diminishes gradually when output generation increases. As observed in Figs. 6 and 7, when noticing that wind farms generate high frequency harmonics at low power, harmonic resonance assessment will be given at output generation which is greater than 17 MW at 33 kV bus and 40 MW at 115 kV bus.

Fig. 8 indicates that harmonic resonances happening at the 6th, which is identified by both the reduction and NREL methods. Moreover, the 39th harmonics reaches agreement with the simulation results obtained by using NREL's model. The harmonic resonances are most observable at the 6th harmonic, but the observation of resonances at the 25th harmonic is not obvious, which can be implied from the simulation results. As shown in Fig. 9, the content of high frequency harmonics at 115 kV level is considerably small; therefore, resonances observed at these frequencies are not clear in comparison with the observation at the 6th harmonic. The fact that harmonic resonances also are observed at the other even harmonics in low frequency range, but the most observable resonance is at the 6th harmonic as mentioned. This is explained because the wind farm generates a significant amount of harmonics at low frequencies.

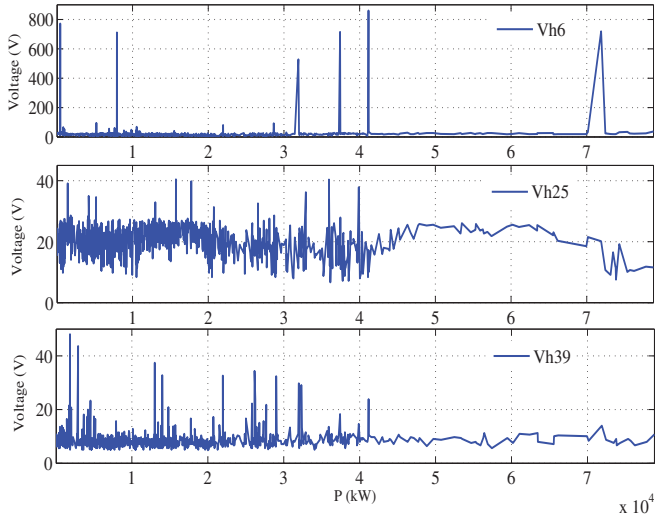


Fig. 8. The 6th, 25th and 39th harmonic voltages at 115 kV

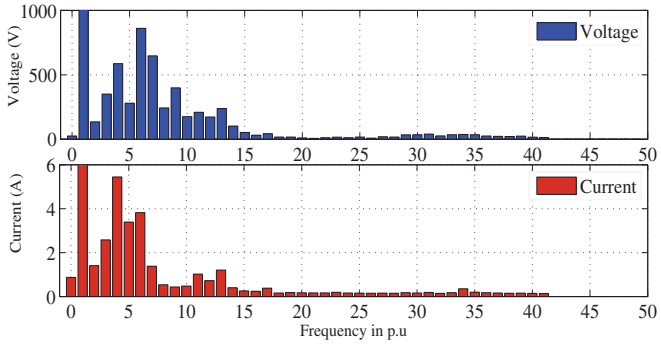


Fig. 9. The harmonic spectra at 115 kV bus collected by meter M4 when active power recorded is 42.992 MW

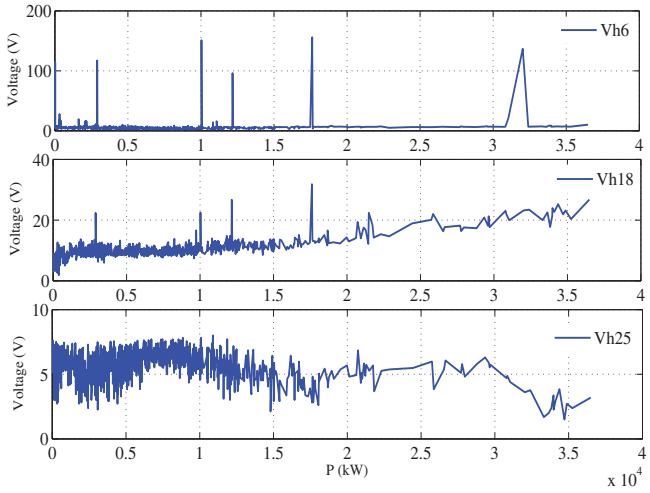


Fig. 10. The 6th , 18th and 25th harmonic voltages at 33 kV bus

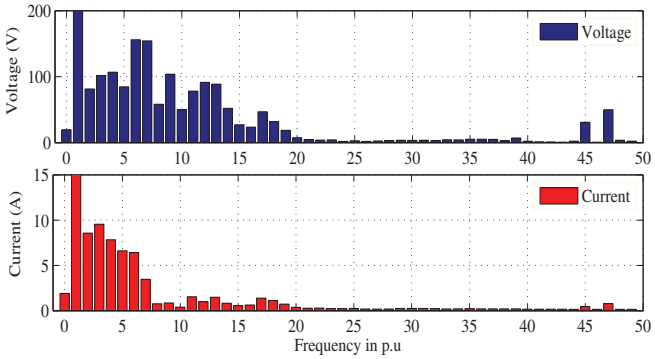


Fig. 11. The harmonic spectra at 33 kV bus collected by meter M3 when active power recorded is 17.617 MW

By analogy, the frequencies at 33 kV bus are verified as in Fig. 10, and harmonic spectra are depicted in Fig. 11. The resonances also are observed at the 6th and 18th harmonics, but there is no resonance at the 25th harmonic and between the 47th and 48th harmonic as found by reduction method.

In verification of the used methods with actual measurement data, the observation shows that the simulation results found by using the model proposed by NREL are worth considering. The identification of harmonic resonance in low frequency range is considerable accurate. In high frequency range, the method also figures out the possible harmonic resonance. There is still some deviation from actual data at the 25th harmonic because the wind farm does not generate an appreciable content of harmonic at the 25th.

The frequency scan analysis method used in study is still restrictive to harmonic resonance analysis. The actual measurement data point out that the magnitude of driving point impedance does not reflect exactly the possibility of harmonic resonances. In other words, frequency scan analysis provides insignificant information but harmonic resonant frequencies.

V. CONCLUSIONS

The study summarizes the pronounced aggregate methods for modeling wind power plants in harmonic resonance analysis. Identification of harmonic resonances has been crosschecked with the actual measurement data in order to validate the methods. The verification with actual data indicates that the obtained results by using the model proposed by NREL in modeling wind power plants is worthwhile for harmonic resonance study.

APPENDIX

A. Network Data

TABLE II. NETWORK IMPEDANCES IN P.U BASED ON 100 MVA

	R	X	B	Section
System	0.02706	0.1736	0	-
Transmission Line	0.0192	0.1198	0.03377	1
	0.0019	0.0119	0.00338	2
Transformer 115/33 kV	0.0033	0.1333	0	-

B. Wind Power Plant Data

TABLE III. WIND FARM IMPEDANCES IN P.U BASED ON 100 MVA

	R	X	B	No.WTGs
Feeder 1	0.011	0.008	0.0024	5
	0.0117	0.0132	0.0012	1
	0.0104	0.0117	0.0011	1
	0.006	0.0067	0.0006	1
	0.008	0.009	0.0008	2
	0.053	0.0598	0.0055	1
Feeder 2	0.0072	0.0066	0.0005	3
	0.0105	0.0119	0.0011	2
	0.028	0.0316	0.0029	2
Feeder 3	0.0355	0.0325	0.0026	3
	0.0074	0.0084	0.0008	4
	0.0525	0.0593	0.0054	2
Feeder 4	0.0118	0.0071	0.0004	2
	0.0095	0.0087	0.0007	1
	0.0133	0.015	0.0014	1
	0.0189	0.0213	0.0019	2
	0.02	0.0226	0.0021	2
	0.0139	0.0157	0.0014	3
Feeder 5	0.0096	0.0087	0.0007	3
	0.0064	0.0072	0.0007	1
	0.0046	0.0052	0.0005	1
	0.0164	0.0185	0.0017	1
	0.0517	0.0583	0.0053	1

The impedance of each section along each feeder is listed from the furthest section to the collector bus. No.WTGs stands for number of wind turbine generators.

ACKNOWLEDGMENT

The authors would like to acknowledge Provincial Electric Authority (PEA) Thailand for providing data. The paper is supported by AUN/Seed-Net, JICA scholarship.

REFERENCES

[1] K. Yang and M. Bollen, "Theoretical Emission Study of Windpark Grids," in 11th International Conference on Electrical Power Quality and Utilisation (EPQU), 2011.

[2] F. Ghassemi and K. Koo, "Equivalent Network for Wind Farm Harmonic Assessments," IEEE Trans.Power.Del., vol. 25, no. 3, pp. 1808–1815, 2010.

[3] E. Muljadi, S. Pasupulati, A. Ellis, and D. Kosterov, "Method of equivalencing for a large wind power plant with multiple turbine representation," IEEE PES General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, pp. 1–9, Jul. 2008.

[4] J. Li, N. Samaan, S. Williams, and N. S. Member, "Modeling of large wind farm systems for dynamic and harmonics analysis," IEEEPES Transmission and Distribution Conference and Exposition, pp. 1–7, 2008.

[5] R. Zheng, M. H. J. Bollen, and J. Zhong, "Harmonic resonances due to a grid-connected wind farm," in Proc. 14th International Conference on Harmonics and Quality of Power - ICHQP, pp. 1–7, 2010.

[6] B. Badrzadeh, M. Gupta, N. Singh, A.Petersson, L.Max, and M. Høgdahl, "Power System Harmonic Analysis in Wind Power Plants-Part I□: Study Methodology and Techniques," IEEE, Industry Application Society Annual Meeting (IAS), pp. 1–11, 2012.

[7] G. A. Mendonça, H. A. Pereira, and S. R. Silva, "Wind Farm and System Modelling Evaluation in Harmonic Propagation Studies," in International Conference on Renewable Energies and Power Quality (ICREPQ), 2012.

[8] S. A. Papathanassiou and M. P. Papadopoulos, "Harmonic Analysis in a Power System with Wind Generation," IEEE Transactions on Power Delivery, vol. 21, no. 4, pp. 2006–2016, 2006.

[9] R. King and J. Ekanayake, "Harmonic modelling of offshore wind farms," IEEE PES General Meetig, pp. 1–6, 2010.

[10] IEC Standard 61400-21 Wind Turbines - Part 21: "Measurement and assessment of power quality charateristics of grid connected wind turbnines," 2008.

[11] IEC Standards Part 3-6: "Assessment of emission limits for the connection of distoring installation to MV, HV and EHV power system," 2008.

[12] J. Arrillaga and N. Watson, Power System Harmonics, 2nd ed., Wiley, 2004, pp.261-305.

[13] E. Muljadi, C. P. Butterfield, A. Ellis, J. Mechenbier, J. Hochheimer, R. Young, N. Miller, R. Delmerico, R. Zavadil, and J. C. Smith, "Equivalencing the collector system of a large wind power plant," IEEE PES General Meeting, 2006.

[14] J. Brochu, S. Member, C. Larose, and R. Gagnon, "Validation of Single- and Multiple-Machine Equivalents for Modeling Wind Power Plants," IEEE Transactions on Energy Conversion, vol. 26, no. 2, pp. 532–541, 2011.

[15] J. Brochu, S. Member, C. Larose, and R. Gagnon, "Generic Equivalent Collector System Parameters for LargeWind Power Plants," IEEE Transactions on Energy Conversion, vol. 26, no. 2, pp. 542–549, 2011.

[16] C. Larose, R. Gagnon, P. Prud'Homme, M. Fecteau, and M. Asmine, "Type-III Wind Power Plant Harmonic Emissions: Field Measurements and Aggregation Guidelines for Adequate Representation of Harmonics," IEEE Transactions on Sustainable Energy, pp. 1–8, 2013.

[17] W. Xu, "Component modeling issues for power quality assessment," Power Engineering Review, IEEE, vol. 21, no. 11, pp. 12–17, 2001.

[18] Ł. H. Kocewiak, J. Hjerrild, and C. L. Bak, "Wind Farm Structures ' Impact on Harmonic Emission and Grid Interaction," in Proc. European Wind Energy Conference & Exhibition, pp. 1–8, 2010.

Optimized GDPWM based on Spontaneous Evolutionary GA for Reducing Switching Losses on Inverter

Ony Asrarul Qudsi[1], Novie Ayub Windarko[2], Ardyono Priyadi[1], and Mauridhi Hery Purnomo[1]

[1]Department of Electrical Engineering
Institut Teknologi Sepuluh Nopember
INDONESIA

[2]Department of Electrical Engineering
Politeknik Elektronika Negeri Surabaya
INDONESIA

***Abstract*—This paper presents analytical techniques for reducing switching losses of voltage source inverter (VSI) using Generalized Discontinuous PWM (GDPWM). The switching losses of inverter is influenced by the angle on the modulation of GDPWM. This problem will be optimized using a new optimization method. This method is called as Spontaneous Evolutionary GA (SEGA). The inverter switching losses is formulated as objective function to optimize the angle. At this optimization process, angle values will be determined to minimize the inverter switching losses. Thermal module of Power Simulator (PSIM) is used to verify the optimized angle of GDPWM. The simulation was performed using a three-phase voltage source inverter (VSI) and an inductive load. Simulation results confirm the method could minimize the losses of inverter.**

Keywords— Generalized Discontinuous PWM (GDPWM), inverter, Spontaneous Evolution Genetic Algorithm, switching losses.

I. INTRODUCTION

Recently, voltage source inverter (VSI) is widely used, especially for AC motor drives, utility interface, and a DC to AC converter. For example, inverter is capable to provide a smooth rotation of motor drive. As shown in Figure 1, VSI has a simple electrical circuit. VSI consists of semiconductor switches that are arranged so they can produce AC voltage. The AC output voltage can be varied according to the DC input voltage or to vary the ratio of the DC input voltage and AC output voltage called the gain (gain) inverter. Generally VSI is used for generating AC voltage and controllable frequency using pulse-width modulation (PWM).

PWM technique is applied in power electronics technologies that have fast transition time. PWM in inverter switching can be set with the duty cycle. The duty cycle is used to control the switching semiconductors. To get the duty cycle, the triangular signal will be compared with the reference signal so that the duty cycle can be obtained to control the inverter switching in accordance with the desired output. Sinusoidal pulse-width modulation (SPWM) is one of the commonly used PWM strategies. The basic concept of this strategy is the PWM triangular signal will be compared with a sine signal. In this way the duty cycle will be obtained for switching the inverter [1-3].

There are several types of PWM for inverter application such as traditional PWM, SPWM, harmonic/s injected PWM, SVPWM, GDPWM, random PWM etc. These kinds of PWM type have their own advantages and disadvantages. Generally, those types of PWM have been studied for reducing its harmonics [4]. While the increasing of applications in renewable energy need further study to reduce loss in power converter system [5].

The selection of PWM strategies for switching should be based on several factors, including the output signal quality and losses. In [6], it has been described that by adding a sequence of zero or non-sinusoidal signals in the modulation process, power quality improves, losses is reduced due to switching inverter, and linear modulation range is extended. If the addition zero sequence waveform is continuous it will produce Continuous PWM (CWPM). Otherwise, if the addition of zero sequence waveform is not continuous, it will produce Discontinuous PWM (DPWM). They are shown in figure 2 and figure 3. A carrier based generalized PWM method comprising of all DPWM schemes is considered as generalized PWM scheme (GDPWM) [6-9].

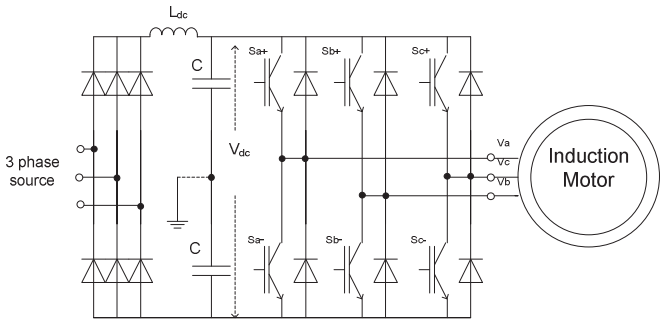


Figure 1. Electrical circuit VSI

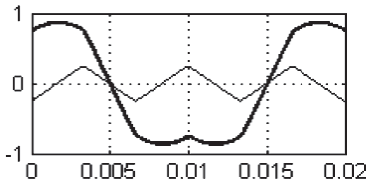
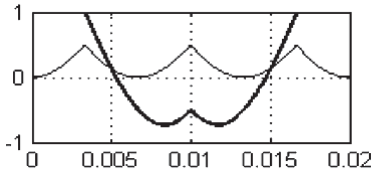
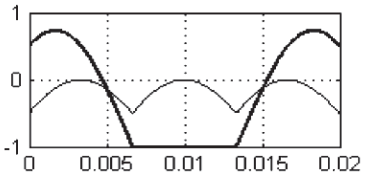


Figure 2. Continuous PWM (CWPM)



(a)



(b)

Figure 3. Discontinues PWM (DPWM), (a) DWPM MAX, (b) DWPM MIN [9]

GDPWM is effective to reduce the losses that are caused by inverter switching [7],[8]. In this research, it also concern about the use of GDPWM for reducing losses caused by inverter switching. However, the switching losses of inverter is influenced by the angle on the modulation of GDPWM. Hence, in this research the optimization on GDPWM proposed to obtain the minimum value of losses. Optimization proposed in this study using Spontaneous Evolutionary GA (SEGA). This is new method which combines artificial neural network (ANN) and GA (Genetic Algorithm). By using this optimization method, the optimal value of inverter switching losses could be obtained.

II. THE GDPWM FOR INVERTER SWITCHING

Any types of zero sequence signals can be injected to the reference signal. Figure 4 illustrates the technique of zero-sequence signal injection. From the figure it can be seen that there is reference signal that inject a zero sequence signal then compared with triangular signal. That signal will produce switching signals for the inverter. Selection of the proper zero sequence signals can improve the output signal quality (ripple) and reduce losses. As illustrated in Figure 3, there are a variety of output signals with the addition of zero-sequence signal technique to signal references.

GDPWM is a method by combining several DPWM. This includes DPWM 1 and DPWM 2 as shown in Figure 5. To get GDPWM, zero sequence signal will be added to the signal references. This is illustrated according Figure 6.

From illustration Figure 6, the zero signal sequence is shaded dark signal with saturation signal limit $\frac{V_{dc}}{2}$ to $-\frac{V_{dc}}{2}$, where V_{dc} is the voltage generated by the VSI dc-link.

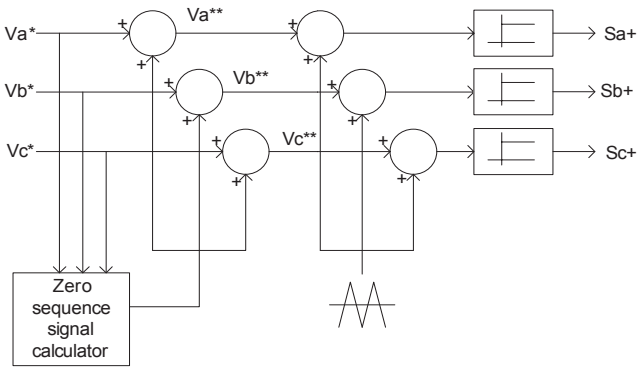


Figure 4. The zero-sequence injection principle [6]

Based on Figure 6, it can be seen that there are three references signal V_a^*, V_b^*, and V_c^*. Three reference signals have angle difference 120^o for each. Then it is modulated by the addition of a zero sequence signal. Then it will be done at

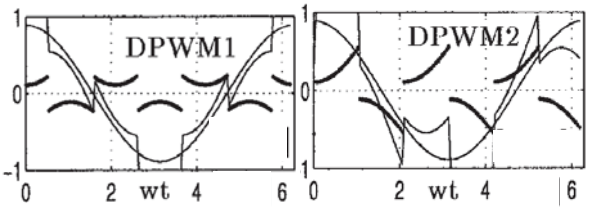


Figure 5. DPWM 1 and DPWM 2 [6]

maximum magnitude test with ψ shift, so producing three new signals V_{ax}^{**}, V_{bx}^{**}, and V_{cx}^{**}. Furthermore V_o will be elected. Assuming:

$$|V_{ax}^{**}| \geq |V_{bx}^{**}|, |V_{cx}^{**}| \quad (1)$$

So it can be obtained using equation (2):

$$V_o = \left(sgn(V_a^*)\right)\frac{V_{dc}}{2} - V_a^* \quad (2)$$

Then the reference signal V_o will be modulated with the reference signals V_a^*, V_b^*, dan V_c^* to obtain three new V_a^{**}, V_b^{**}, dan V_c^{**} called GDPWM. Figure 7 illustrates GDPWM which has produced by injecting zero-sequence signal.
GDPWM is easily to be implemented on digital signal processing (DSP) or microprocessor. This is due GDPWM operation only requires a phase shift and some comparisons. In this process, shifting angle ψ is very useful for analysis and plotting of GDPWM. Dq0 transformation and expanding the terms is in a manner to minimize the computational

requirements. The rotation calculation can be determined in the following equations:

$$V_{ax}^{*} = V_{a}^{*}\cos(\psi_{m}) - \frac{(V_{b}^{*} - V_{c}^{*})}{\sqrt{3}}\sin(\psi_{m}) \quad (3)$$

$$V_{bx}^{*} = V_{b}^{*}\cos(\psi_{m}) + \left(\frac{1}{2}\frac{(V_{c}^{*} - V_{b}^{*})}{\sqrt{3}} - \frac{\sqrt{3}V_{a}^{*}}{2}\right)\sin(\psi_{m}) \quad (4)$$

$$V_{cx}^{*} = -V_{ax}^{*} - V_{bx}^{*} \quad (5)$$

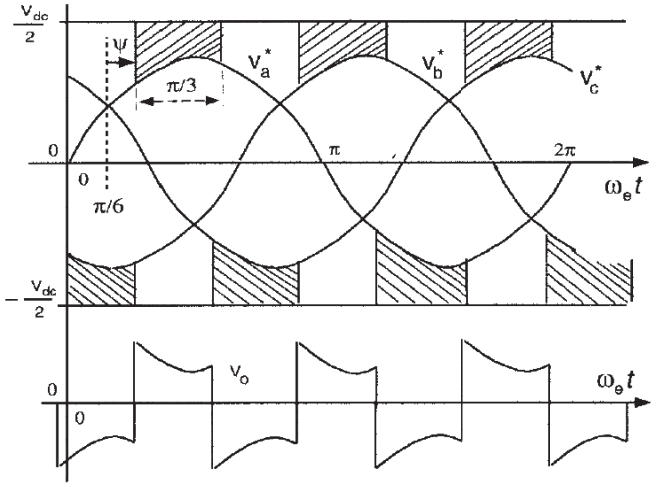


Figure 6. GDPWM zero-sequence signal generation method [6]

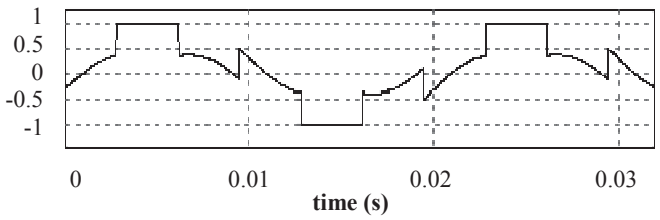


Figure 7. GDPWM

After GDPWM obtained duty cycle, inverter switching can be determined. Furthermore, the duty cycle is used for the switching inverter with smaller losses.

III. SPONTANEOUS EVOLUTIONARY GA

Spontaneous Evolutionary GA (SEGA) is a advanced development of the methods of optimization Genetic Algorithm (GA). Essentially the selection and optimization is used in SEGA algorithm similar as GA, but the generation of new populations during the selection process is the main thing that distinguishes SEGA with GA. When certain generation of GA fitness is not according with fitness setting, then SEGA will be raised up a new population to replace the existing population of GA.

New population will be generated from the existing population then trained base on the best individual results of GA. Training method used is neural network. This makes the resulting new population will be more homogeneous and have a better chromosome quality from existing population. The new population will be used to replace the old population in the GA iterations. The new population is generated based on the fitness value of the GA standard. If fitness is less than the GA standard, the neural network will be work.

After the data from the initial population of GA trained using neural network, a new population is obtained by an amount equal to the initial population. This population will be used for the next iteration in the selection process. New population of neural network replaces the existing GA population. Figure 8 shows the SEGA algorithm is used in this research.

Real number encoding scheme used to encode chromosome. On this scheme, the value of gen is in the interval [0, *R*], with *R* is a positive real number. A gen "g" is used to represent a real number between 0 and 1. By using a certain interval, the lower limit (r_b) and upper limit (r_a), decoding can be performed in the following equation:

$$x = r_b + (r_a - r_b)g \quad (6)$$

An individual will be evaluated based on a specific function as the measure its performance. Individuals that have high fitness will be alive. While individuals that have low fitness will be die. In this research, optimized problem (minimized problem) is losses generated due to switching inverter so that fitness that used is:

$$f = \frac{1}{h} \quad (7)$$

Where *f* is the fitness value and *h* is the minimized problem (inverter switching losses). But this will be a problem if h is 0, so that *f* is infinite. To handle it *h* need plus a small number so the fitness value becomes:

$$f = \frac{1}{h+a} \quad (8)$$

Where *a* is a number that is considered to be very small and varies according to the problem to be solved. After obtaining the value of fitness, then will be ranked according following equation:

$$f(i) = f_{max} - (f_{max} - f_{min})\left(\frac{R(i)-1}{N-1}\right) \quad (9)$$

where, *R (i)* declare the *i*-th individual rank. . Next will be selected using the roulette wheel method. This method

imitating roulette wheel where each chromosome occupies a circular piece on the roulette wheel in proportion with The fitness value. Chromosome has greater fitness value will be occupies pieces a larger circle than the low fitness value of chromosomes.

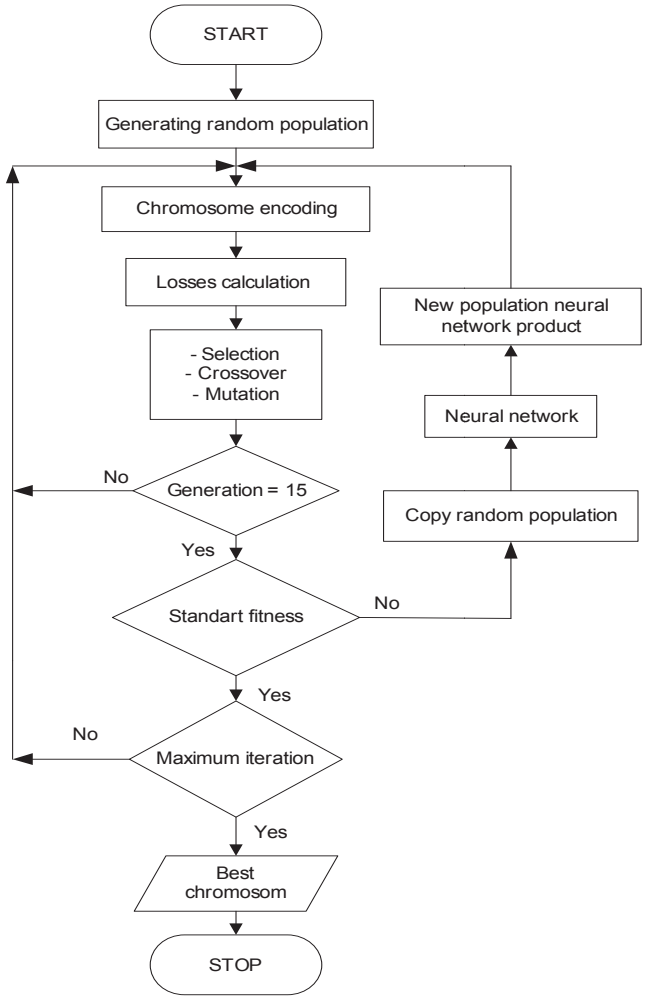


Figure 8. Flow chart losses switching inverter calculation using SEGA

One of the most important components in the genetic algorithm are crossovers. A chromosome that leads to a good solution can be obtained from the crossover process on two chromosomes. Crossovers can be done in several ways and the simplest is the cut point crossover (one point crossover). A cut point chosen at random, then the first part of the parents 1 combined with the second part of the parents 2.

Mutation procedure is very simple. For all existing gen if the random number generated is less than a specified probability of mutation P_{mut} then change the gen into the opposite value. P_{mut} usually is set as $\frac{1}{n}$ where n is the number of gen in the chromosome. It means mutation only occurs in about one gene alone. In a simple genetic algorithm, P_{mut} value is the same in evolution.

IV. SIMULATON AND RESULT

Simulations were conducted to verify specific and modeling methods that have been designed. Simulation is done by implementing GDPWM on the switching inverter using insulated gate bipolar transistor (IGBT) *SEMiX151GD066HDs* of Thermal Module in PSIM. Based on references [6], IGBT has a linear relationship between the losses and the with load current magnitude. If the load current magnitude become greater, then losses become greater. Moreover, the losses in the switching inverter is also affected by the power factor and the modulation method used. This has been studied in references [6].Therefore the selection of modulation methods must be considered in order to reduce losses to the maximum. Assuming the inverter switching device has a current turn on and turn off that linear with respect to time, so that the calculation of losses due to inverter switching can only be modeled on basis component.

GDPWM has lower losses than SPWM. Figure 9 shows general illustration about the difference of GDPWM losses and SPWM losses for inverter switching. This simulation result shows GDPWM able to reduce losses caused by inverter switching. Furthermore, the angle of GDPWM can be set to reach the minimum value of losses.

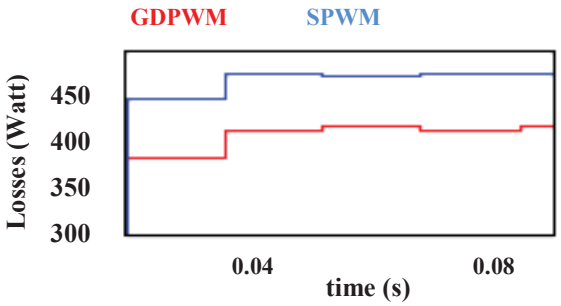
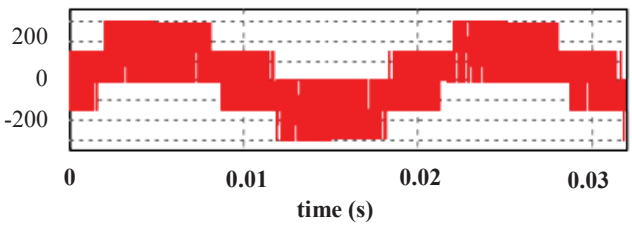


Figure 9. SPWM losses and GDPWM losses on inverter switching

Figure 10 illustrates the of voltage output inverter, harmonic distortion, GDPWM with different ψ, and losses using GDPWM. Shift angle ψ causing losses value changing. It is as illustrated in figure 10 (d), (e), (f), (g). Based on the illustration figure 10 (d), (e), (f), (g), inverter switching losses with $\psi = 0$, $\psi = \frac{\pi}{6}$, $\psi = \frac{\pi}{4}$, and $\psi = \frac{\pi}{3}$ is different.



(a)

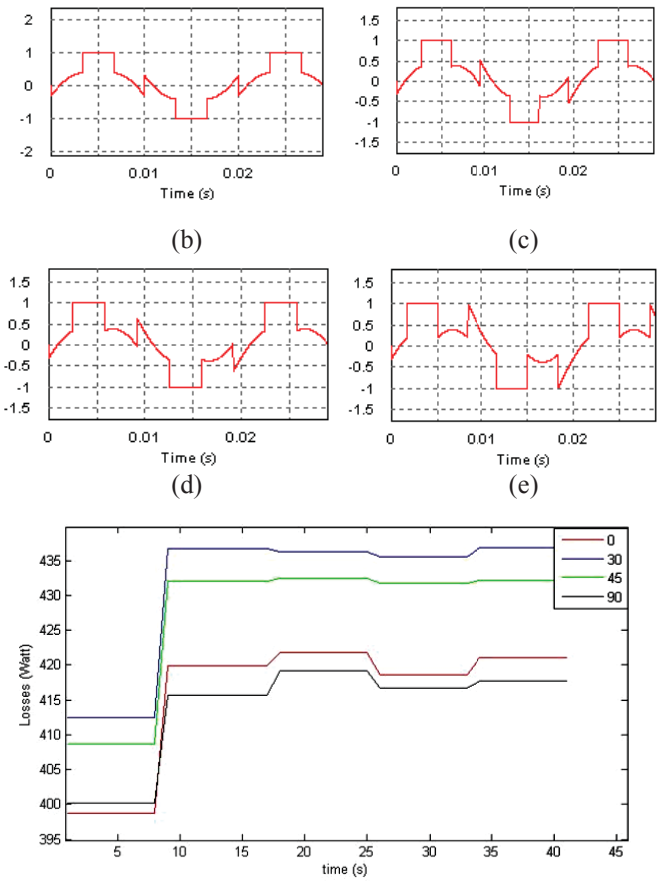


Figure 10. (a) The inverter voltage output, (b) GDPWM with $\psi = 0$, (c) GDPWM with $\psi = \frac{\pi}{6}$, (d) GDPWM with $\psi = \frac{\pi}{4}$, (e) GDPWM with $\psi = \frac{\pi}{3}$, (f) Losses GDPWM

In this research, optimization of GDPWM by varying the angle ψ to obtain the optimal value losses (minimum value). To get the optimal value of losses used optimization method SEGA. This method will generate 100 chromosomes for each population. Next will be the evaluation, selection, crossover, and mutation of the individual to get the best. In this research, the best individual represents the best angle to get the optimal inverter switching losses. Figure 11 illustrates the value of using GDPWM inverter losses that have been optimized with $\psi = 71^{o}$.

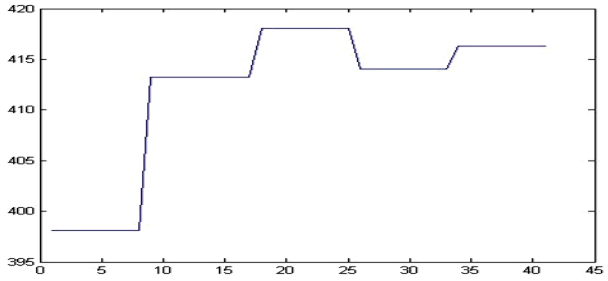


Figure 11. optimized GDPWM

Table I shows the losses of switching inverter with different ψ. Optimized GDPWM using SEGA have lowest losses amounting 411.9617 watt.

TABLE I. GDPWM LOSSES

No	ψ (^{o})	Losses (Watt)	Losses (%)
1	0	416.1421	2.080711
2	30	431.7105	2.158553
3	45	427.5633	2.137817
4	90	413.9914	2.069957
5	71	411.9617	2.059809

V. CONCLUSION

SPWM method is the conventional method for switching the inverter. However, these methods result significant losses in the inverter switching. In this research the method was developed and optimized using GDPWM SEGA for reducing losses in the inverter switching. The proposed method provides an effective solution to minimize the switching losses in the inverter. In addition, the inverter output current and voltage generated using this method has a good enough quality. Furthermore, the proposed GDPWM can be applied for some application such as solar inverter, inverter drive, PWM rectifier etc.

ACKNOWLEDGMENT

First author would to tanks to Fresh Graduate Scholarship for giving the financial support to study in Magister program at ITS, Surabaya.

REFERENCES

[1] J. Holtz, "Pulse width modulation – A Survey", *IEEE Trans. Ind. Electron.,* vol. 39, No. 5, December 1992.

[2] J. Holtz, "Pulse width modulation for electronic power conversion," *Proc. IEEE,* vol.82, no.8, pp 1194-1214, Aug.1994.

[3] Ahmet M. Hava, Russel J. Kerkman and Thomas A. Lipo, "Simple Analytical and Graphical Methods for Carrier –Based PWM-VSI Drives" *IEEE Trans.PowerElectronics*,vol4,no.1,pp.49-61,Jan,1999.

[4] Shi, K.L, and Hui Li,"Optimized PWM strategy based on genetic algorithms,", IEEE Transactions on Industrial Electronics, Vo.52, no. 5,pp. 1458 – 1461, 2005.

[5] Kakigano, H., Nomura, M., and Ise, T.," Loss evaluation of DC distribution for residential houses compared with AC system," Proc. Of International Power Electronics Conference (IPEC), pp. 480 – 486, 2010

[6] Ahmet M. Hava, Russel J. Kerkman and Thomas A. Lipo, "A High Performance Generalized Discontinuous PWM Algorithm", *IEEE Trans Ind appl,* vol34, No.5, pp.1059-1071September/October 1998.

[7] Olorunfemi Ojo, "The generalized discontinuous PWM scheme for three-phase voltage source inverters" *IEEE Trans. Ind. Electron.,* vol. 51, no. 6, pp. 1280-1289, Dec 2004.

[8] T. Brahmananda Reddy, J. Amarnath, D. Subba Rayudu and Md. Haseeb Khan, "Generalized Discontinuous PWM Based Direct Torque Controlled Induction Motor Drive with a Sliding Mode Speed Controller" *IEEE Proc. Power Electronics, Drives and Energy systems for Industrial Growth, PEDES'06*, New Delhi, India, paper no. 3D-11, Dec, 2006.

[9] K. Sri Gowri, T. Brahmananda Reddy, C. Sai Babu, "A Novel High Performance Generalized Discontinuous PWM Algorithm for Reduced

Current Ripple and Switching Losses Using Imaginary Switching Times"
TENCON 2008 - IEEE Region 10 Conference , Publication Year: 2008
[10] Guo Pengfei, Wang Xuezhi, Han Yingshi, "The Enhanced Genetic Algorithms for the Optimization Design" 3rd International Conference on Biomedical Engineering and Informatics, 2010.
[11] Shan Tai, "An Improved Genetic Algorithm and its Blending Application with Neural Network" .2010.
[12] Purnomo Hery Mauridhi, Kurniawan Agus" Supervised Neural Network dan Aplikasinya", Graha Ilmu, Yogyakarta: 2006.

Regenerative Braking Performance Analysis on GANG CAR Electric

Wahyu Parbowo, Agus Purwadi
Electrical Power Engineering - School of Electrical Engineering and Informatics,
Bandung Institute of Technology
Jl. Ganeca 10, Bandung 40132, Indonesia
wahyupe@gmail.com
apurwadi57@gmail.com

Abstract– **Increasing the driving range of an electric vehicle can be achieved by applying regenerative braking. In this paper, how much regenerative braking affects the vehicle will be discussed. The vehicle prototype on this study is Gang Car Electric, a conversion from internal combustion engine car type. The regenerative braking configuration enables the feedback energy active as soon as the throttle pedal being released and emulate engine retarding function of conventional car. With that goal, we still want the vehicle to be able to cruise in a constant speed on the cruise mode. The proposed configuration is included in this paper.**

Keywords –**Regenerative, braking, electric car conversion, Brushless DC motor**

I. INTRODUCTION

Electric car is a clean technology that has zero emission. Along with the development of power plant technology, electric car will have a cleaner energy source. The constraint of an electric car is the limited driving range because of the difficulty to save electric energy.

Braking is one of the most important features on a vehicle.The conventional braking dissipates kinetic energy that was propelling the car forward as heat and become useless. That wasted energy could have been used to do work.Electric car, which has electric motor for propulsion, enable to convert the kinetic energy from braking to be an electric energy for battery charging. This technology is called regenerative braking and can extend the driving range of an electric vehicle.

Urban driving cycle has frequent start and stop due to the crowded streets. This will allow a lot of braking and a lot of energy recovery that can be used to charge the battery.

In this study, the urban electric car prototype is made of an internal combustion engine called Gang Car. Gang Car has a light weight platform characteristic, so it is suitable to be converted into an electric car. The car is expected to be able to stop in short distances and in accordance with the driver's feel. To achieve this, the regenerative braking needs to be combined with mechanical braking from the car.

II. BRUSHLESS DC 4 QUADRANT OPERATION

Brushless dc motor is a synchronous motor with permanent magnet with a trapezoidal-shaped back-emf. Brushless dc is widely used due to the absence of brush, making easier maintenance, and a great power in a small size.

Because of the permanent magnet, brushless dc motor capable to operates as generator without excitation. It needs the controller to be able to fetch the feedback current, the same controller needed to drive the motor.

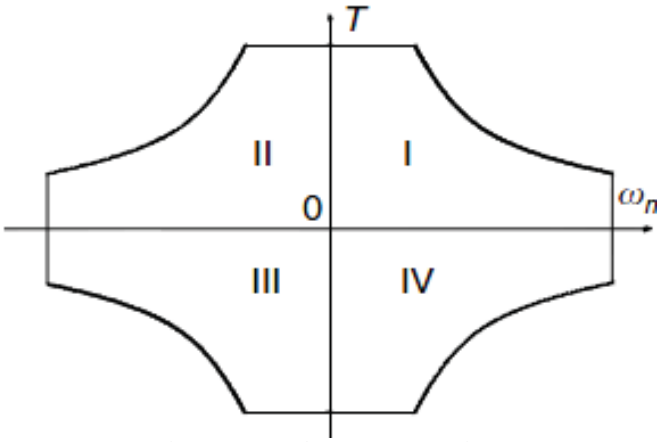


Fig. 1. Four Quadrant Motor Operation

To get the forward braking operation (Quadrant IV), we need thetorque to be negative in positive speed. This will happen when the controller fetches the current from the rotating Brushless DC due to the kinetic or potential energy. Thus, the brushless dc operates as a generator. [1-5]

III. REGENERATIVE BRAKING ON ELECTRIC CAR

Regenerative braking is a method that converts the energy during braking into an electrical energy and fed it back into the battery. The controller must have a bidirectional converter to be able to let the current flow both ways.

In Fig. 2 and Fig. 3 we can see a basic per-phase operation of a bidirectional inverter. On the normal mode, switches S1 and S4 are on simultaneously. The inductor current i_{ab} would be increased due to the current i_{on}. $e_{ab}$is the back-emf from motor. When the switch S1 off and S4 still on, the inductor current will flow into the freewheeling diode D2 and switch S4, which makes a discharging current path i_{off}.

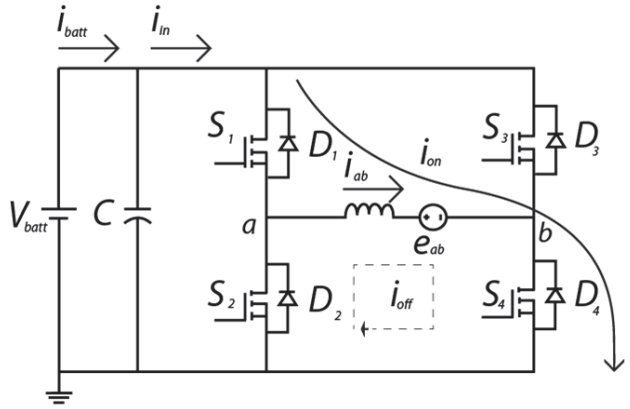


Fig. 2.Converter (per phase) in normal mode [6]

In the regenerative mode, controller will utilize the back-emfe_{ab} as a voltage source. When S2 and S3 are on, the winding will be energized. The inductor voltage will be equal to V_{batt} + e_{ab} and the current i_{in} is equal to $-i_{ab}$ or i_{on}. During the turn off period of S2 and S3, the current i_{in}will flow through the freewheeling diode D1 and D4 and used to charge the battery.[6-8]

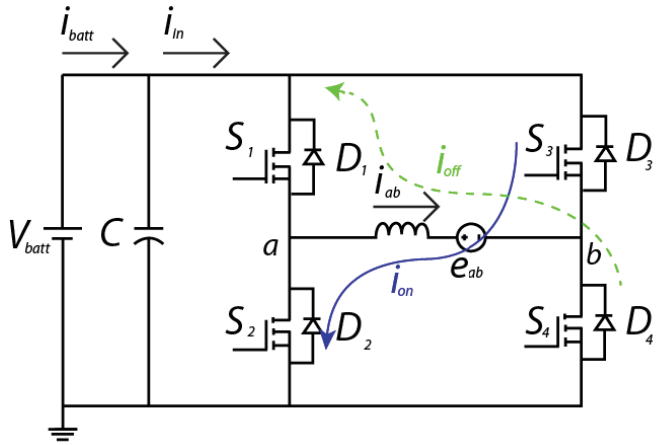


Fig. 3.Converter (per-phase)[6]

To be able to utilize the back-emf to charge the high voltage battery, dc-dc converter is needed.

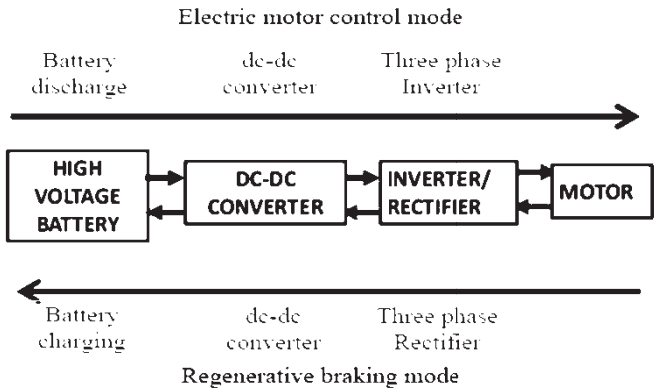


Fig. 4.Regenerative braking scheme

To increase the safety factor of the car, combination of electrical and mechanical brake needed. The braking force distribution on the front and rear axle of the car need to follow I curve on Fig. 5 below.

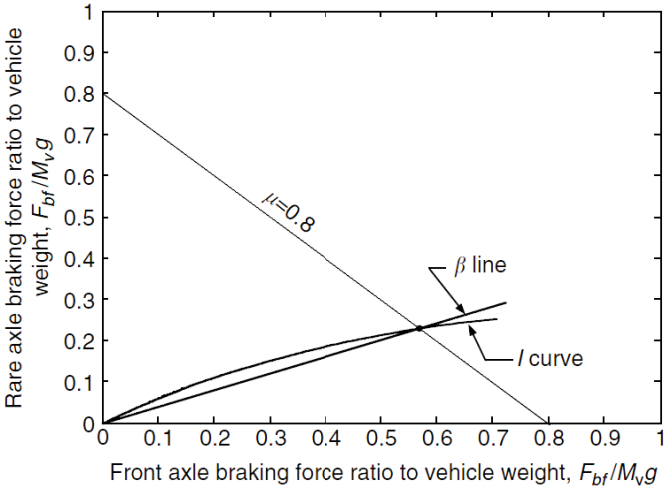


Fig. 5. Braking force distribution curve[2]

Series brake means the use of regenerative braking and mechanical braking is not simultaneously. When the braking command from the driver (with brake pedal) can be met by regenerative braking force, the mechanical brake will not active. This regenerative brake applied on the driven axle will emulates the engine retarding function in conventional vehicles. When the commanded deceleration is greater than the available regenerative braking force, the electric motor will operate to produce its maximum braking torque, and the remaining braking force is met by the mechanical brake system. The braking force distribution is illustrated in Fig. 6. [2]

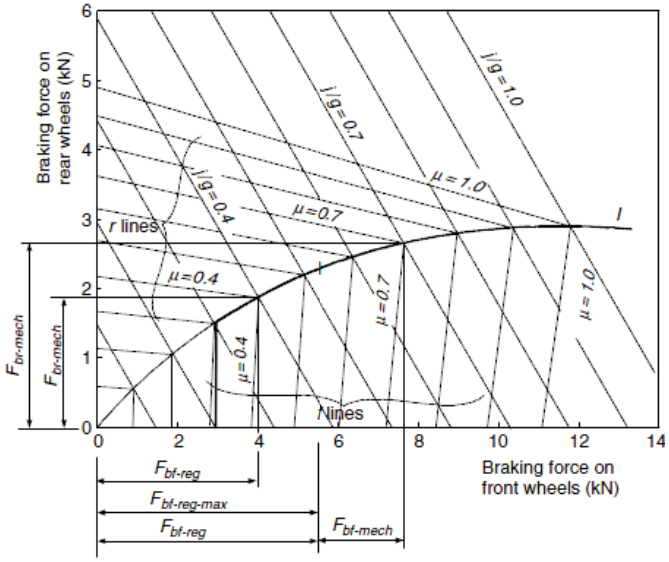


Fig. 6. Series braking force distribution [2]

The counterpart of series brake is parallel brake. It uses the mechanical and electrical brake simultaneously so the need of complex control for the mechanical brake will disappear. Regenerative braking used when the commanded

deceleration is low, if it's high then the mechanical brake will be used simultaneously with the electric regenerative brake. With the simpler control, parallel brake has less recovery energy from braking because the mechanical brake still dissipates the kinetic or potential energy of the car. [2]

The braking force distribution is illustrated in Fig. 7.

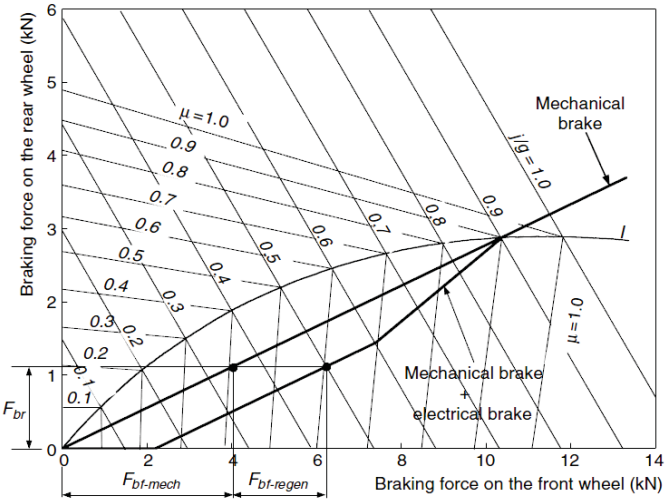


Fig. 7. Parallel braking force distribution [2]

IV. GANG CAR CONVERSION

Gang Car Electric, the prototype used in this study, is a conversion from an internal combustion engine car. Gang Car originally produced by PT. Dirgantara Indonesia in 2003, a small and lightweight car for four passengers. The lightweight characteristic made this car suitable for conversion into an electric car. Motor, controller, and battery are Golden Motor product that already bought before for laboratory experiment.

The motor used for the car is a 10 kW Brushless DC, 72 Volt, 3000-6000 RPM, 4 poles. The controller is HPC500H72500, 72 Volt, 350 Ampere. And the battery pack is $LiFePO_4$, 72 Volt, 100 Ah.

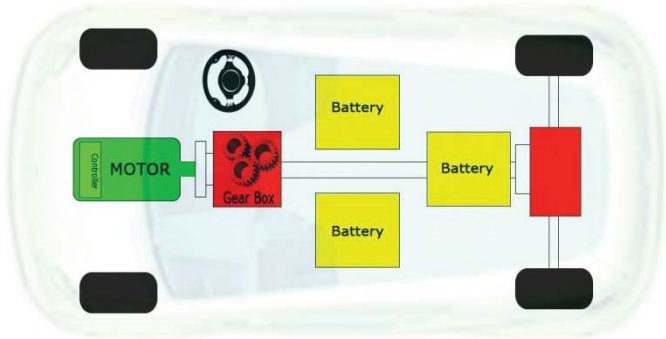


Fig. 8.Gang Car Electric Configuration

V. CONTROLLER AND BRAKING SCHEME

The controller scheme for Gang Car Electric is illustrated in Fig. 9 below.

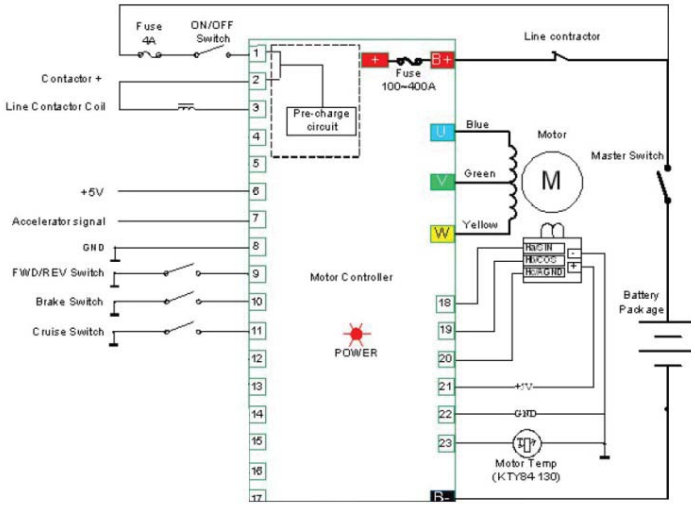


Fig. 9. Controller scheme for Gang Car Electric

The braking proposed for this study is parallel braking because of its simplicity and low cost. To maximize the energy recovery from the regenerative braking, the regenerative mode is activated as soon as the driver releases the throttle pedal. The current feedback to the battery is set to the point where this brake emulates the engine retard function in conventional vehicle.

Normally closed switch for braking is used in the throttle pedal, so when the throttle is pressed, the switch will be opened and it deactivates the regenerative brake mode. Because the throttle won't be activated if the brake-switch still close, it is important to ensure that the brake switch is open when the throttle switch began to be closed.

To be able to use the cruise feature on the controller, the scheme as Fig. 10 is proposed. The driver can choose to use the regenerative mode or cruise mode while driving.

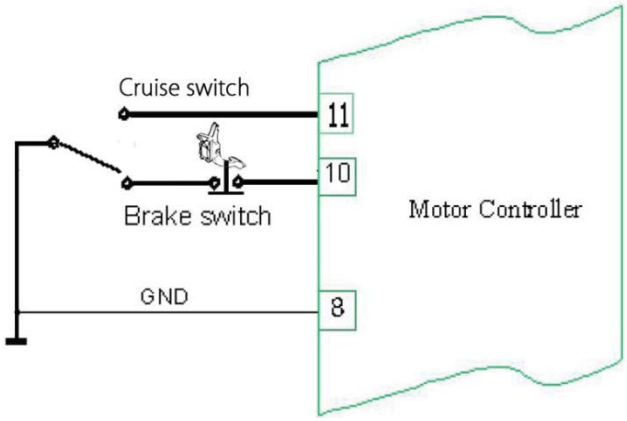


Fig. 10. Braking Scheme

VI. TESTING METHOD

To verify the regenerative braking performance, Gang Car Electric will be tested on a track with NEDC (New European Driving Cycle). NEDC is used to simulate the driving cycle

in a city. The NEDC driving cycle is illustrated in Fig. 11 below. This test is conducted on a track as in Fig. 12 instead of on a dynometer.

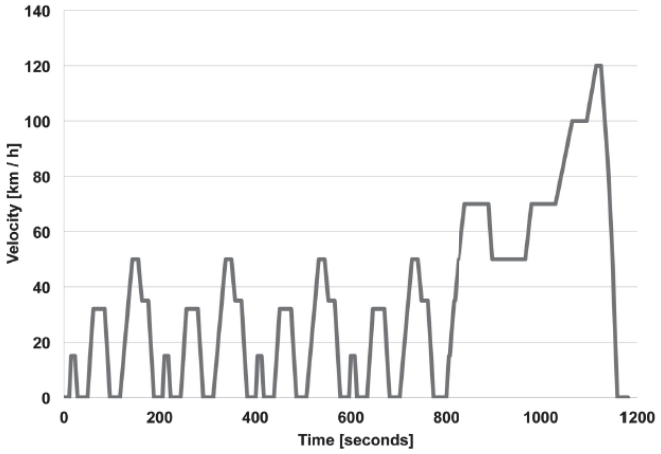


Fig. 11. New European Driving Cycle

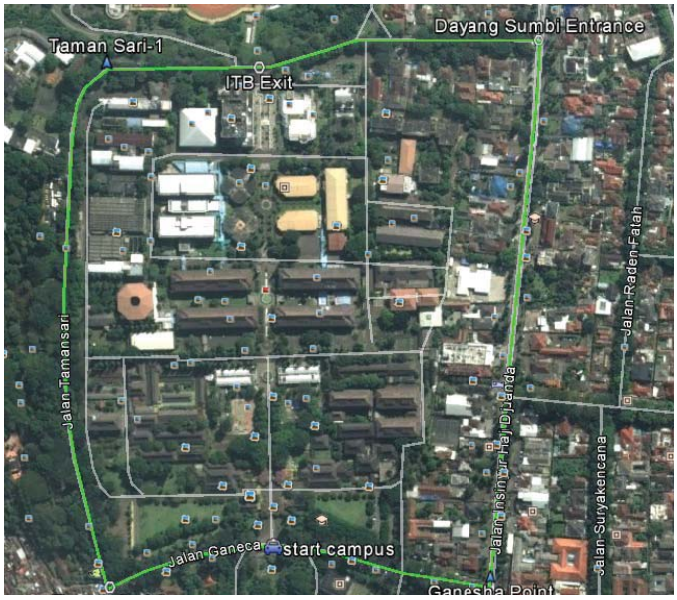


Fig. 12.The track for driving cycle examination

Tests were also performed on a steep road to see how much energy recovery can be obtained. These tests were performed on Sabuga downhill. The Google earth's imaging is shown in Fig. 13.

Fig. 13.Sabuga down track

The data logger built in the controller was used to obtain the rotational speed of the rotor, voltage, and current of the battery. The data logger was calibrated first in the laboratory with existing measuring instruments.

VII. DATA AND INTERPRETATION

A. *Driving Cycle Test*

The driving cycle test with NEDC can be used to determine the fuel economy of Gang Car Electric. The speed profile from the test is shown in Fig. 14.

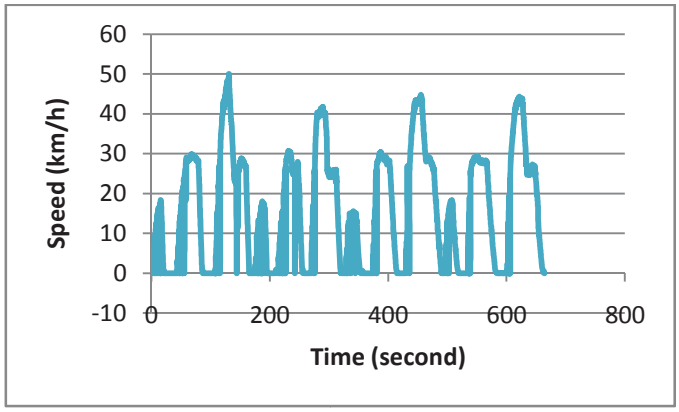


Fig. 14. Speed profile test result

TABLE I
TEST RESULT SUMMARY

Testing Time	18 May 2013, 2:44 (GMT +7)
Total distance covered	2973,84 meter
Total time	664 second
Average rotor rotation	585 rpm
Average speed	16,11 km/h
Total car mass	790 kg (with 4 passenger)
Total energy consumption	272,30 Wh
Feedback setting:	
Voltage	72 Volt
Current	20 Ampere
Total energy recovery	7,98 Wh

NEDC test usually performed on a dynometer, according to reference [10] though it has represent a city driving cycle, energy consumption could be greater in reality.

This test was performed on a real road track, so the speed profile in Fig. 14 will suffice.Power from the battery during the test can be seen in Fig. 15 below.

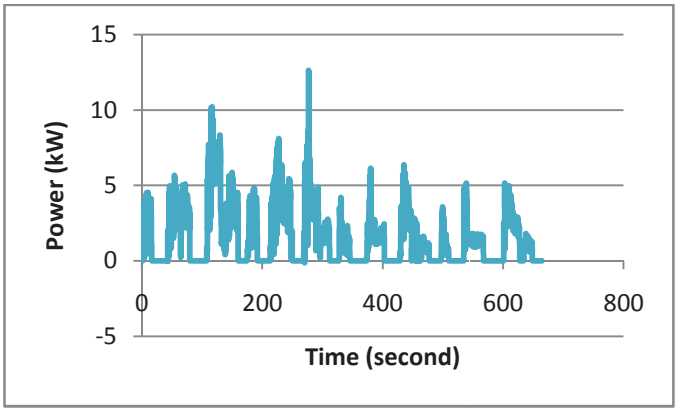


Fig. 15. Power from battery without regenerative

Amount of energy required to complete the driving cycle is **278,298 Wh**. Thus, the energy efficiency of the car is as follows:

$$Energy\ Efficiency = \frac{Energy\ Consumption}{Total\ distance\ covereed} = \frac{272{,}298\ Wh}{2{,}973\ km} = \mathbf{91{,}564\ Wh/km}$$

The feedback energy obtained from the test can be seen in Fig. 16. Keep in mind, this driving cycle testing did not performed on a flat trajectory. The height profile of the track is illustrated on Fig. 17.

Despite the uphill road, the motor still able to produce energy feedback when the driver releases the throttle pedal to decelerate.

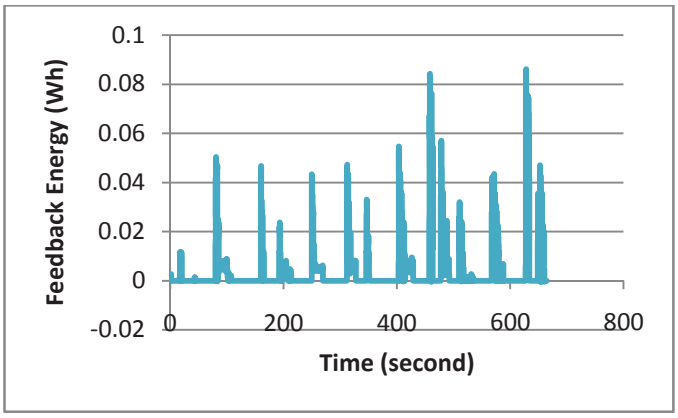


Fig. 16. Regenerative braking energy recovery

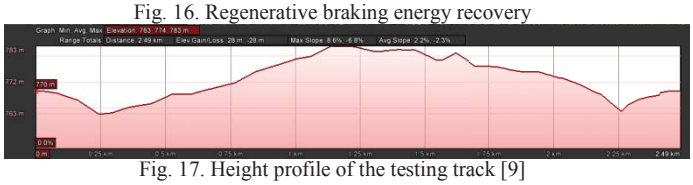


Fig. 17. Height profile of the testing track [9]

Total energy obtained from regenerative braking is 7,98Wh or 3% of its total energy consumption. With the 72V 100Ah battery, Gang Car Electric will have the driving range as in Table 2 below:

TABLE 2
DRIVING RANGE OF GANG CAR ELECTRIC

Without Regenerative Braking	78,63 km
With Regenerative Braking	81,01 km

The regenerative braking driving range is not that significant with just 3% addition. This is likely occurring due to the feedback setting on the controller. The setting used in this test is for optimal driver feel to emulate the engine retarding function of a conventional vehicle.

The amount of kinetic energy that can be converted to electricity to charge the battery is calculated as below:

$$\eta = \frac{P\ \Delta t}{\frac{1}{2}\ m\ v^2} = \frac{28743{,}35\ Watt.s}{169126{,}6\ Joule}\ x\ 100\% = \mathbf{16{,}99\%}$$

Where:

η : Regenerative efficiency
m : Car mass(kg)
v : Car speed (m/s)
P : *feedback power* (watt)
Δt : braking time (s)

16,99% energy which usually wasted on conventional car can be converted to electrical energy to charge the battery by this braking configuration.

B. *Downhill Test*

This test obtained the energy feedback data when the car drove on a downhill. Moreover, we used a much higher setting for feedback current to the battery on controller setting as in Fig. 18.

The track has an average slope of -9,5% and -17,6% maximum steepness. The total length of the track is 311 meters. The current feedback reference in controller is set to 20 Ampere (optimal driver feel) and 50 Ampere (optimal recovery energy - 0,5C of the battery).

Fig. 18. Controller feedback current setting

Table 3 show the summary of the test and Fig. 19 show the speed profile.

TABLE3
DOWNHILL TEST RESULT SUMMARY

Parameter	20 Ampere Setting	50 Ampere Setting
Total time	75 seconds	115,99 seconds
Track length	311 meters	311 meters
Height	24 meters	24 meters
Total mass	790 kg	790 kg

	(4 passenger)	(4 passenger)
Average rotor rotation	1236 rpm	712 rpm
Total feedback energy	9,47 Wh	17,30 Wh

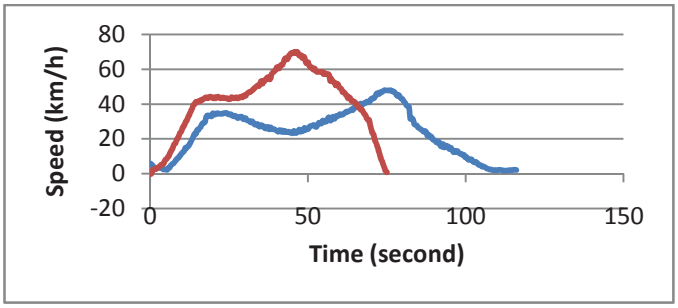


Fig. 19.Speed profile of downhill test

The red line is for 50 A setting and the blue one is for 20 A setting. The results show that with the larger current feedback setting, the larger feedback energy obtained. Because the load draw a larger current, the negative torque needed is larger, thus the braking force is larger that make the speed for the test lower.

Calculation of the amount of potential energy converted to electricity is as follows:

- For 20 Ampere Setting

$$\eta = \frac{P\,\Delta t}{m\,g\,h} = \frac{34085{,}3\ Watt.s}{185922\ Joule}\ x\ 100\% = \mathbf{18,33}\ \%$$

- For 50 Ampere Setting

$$\eta = \frac{P\,\Delta t}{m\,g\,h} = \frac{62776{,}6\ Watt.s}{185922\ Joule}\ x\ 100\% = \mathbf{33,50}\ \%$$

With the same trajectory, a higher feedback current setting can convert more potential energy to charge the battery. According to reference [6] regenerative braking can utilize 8-25 % of wasted energy. The regenerative efficiency is in the range with the optimal driver feel 20 Ampere setting. The insignificant adding to the driving range could be due to the losses from the battery and the mechanical losses from the car.

Battery of the Gang Car Electric has 100 Ah capacities, so it is actually fine to be charged with 50 Ampere current (0,5C). But, the braking force produced from the setting is too large and just concentrated in rear axle. The distribution of this braking force could be dangerous for the car and the passenger in it.

With the 20 Ampere setting in this study, Gang Car Electric can emulate the retarding engine of internal combustion engine car. This regenerative braking force is applied to rear axle and enough to decelerate the car without making a huge different from conventional car.

VIII. Conclusion

The conversion of Gang Car into an urban electric car prototype is successfully done. The Gang Car Electric with 10kW Brushless DC motor has 91,56Wh/km energy consumption.

Parallel regenerative braking system applied to Gang Car Electric can function properly as designed. As soon as the driver releases the throttle pedal, regenerative braking mode is activated. The cruise mode can still be used up to the driver.

The regenerative braking scheme can increase the driving range of Gang Car Electric. It's 78,6 km without regenerative braking and 81,01 km with regenerative braking. Regenerative braking efficiency for 20 Ampere setting is 18,33% and 33,50% for 50 Ampere setting. So, to increase the recovery energy the feedback current setting can be set higher, but it will make a badbraking force distribution.

References

[1] D. Torres, P. Heath, "*Regenerative Braking of BLDC Motors",* [online document], Microchip Technology Inc., 04 April 2013, Available: http://ww1.microchip.com/downloads/en/DeviceDoc/Regenerative%20Braking%20of%20BLDC%20Motors.pdf

[2] Ehsani. Mehrdad et al, *Modern Electric, Hybrid Electric, and Fuel Cell Vehicle: Fundamentals, Theory, and Design,* USA: CRC Press LLC, 2005.

[3] S. Leithman, B. Brant, *Build Your Own Electric Vehicle,* Ed. 3, Singapore: McGraw Hill Education, 2013.

[4] A. E. Fitzgerald, C. Kingsley Jr., S.D. Umans, *Electric Machinery*, Ed. 6, Singapore: McGraw Hill, 2003.

[5] Steve, J. Chapman, *Electric Machinery Fundamental*, Ed. 2, Singapore: McGraw-Hill, 1991.

[6] M.K. Yoong, et.al., *"Studies of Regenereative Braking in Electric Vehicle",* IEEE, 2010.

[7] Ned Mohan, *Power Electronics,* Jon Wiley & Sons Inc., Singapore, 2003.

[8] R. Krishnan, *Electric Motor Drives*, Prentice Hall International Inc., London, UK, 2001.

[9] Dozeno. Jimmy, *UjiKinerja Motor Listrik BLDC 10kW danSistemKendalinyapadaPrototipe Mobil Listrik ITB-1,* TugasAkhir, TeknikTenagaListrik ITB, 2012.

[10] M. Peter, G. John, B. Anup, dan R. Iddo, *Discrepancies Between Type-approval and "real-world" Fuel-consumption and CO2 Values*, ICCT, no. 2012-02, 2012.

SIMULATION OF MAGNETIC FIELD DISTRIBUTION OF OPPOSITE-POLES SINGLE-DISC PERMANENT MAGNET ROTOR

Prih Sumardjati Mulyaseputra, Suharyanto, Sasongko Pramono Hadi, Danang Wijaya
Dept of Electrical Engineering and Information Technology
UniversitasGadjahMada (UGM)
Yogyakarta, Indonesia
prihsumardjati_s3_12@ugm,
suharyanto@ugm.ac.id,sasongko@te.ugm.ac.id,danangwijaya@ugm.ac.id

Prih Sumardjati Mulyaseputra
Department of Electrical Engineering
Bandung State Polytechnic
Bandung, Indonesia
sumardjatiprih@yahoo.co.id

***Abstract*—The study development regarding Axial Flux Permanent Magnet Generator (AFPMG) has been increasingly interesting, since its construction is simple and easy to construct; permanent magnet type NdFeB has high density, available everywhere and relatively low cost; without using brushes and slip ring, thus the maintenance cost can be low, long usage duration, direct drive connection. These will affect in the increasing efficiency and reliability. One of the problems which remain to be the main issue in AFPMG is rotor disc dimension influenced by distance between magnets in rotor disc. This paper discussed magnetic field distribution simulation between permanent magnet NdFeB type N40 with diameter dimension 40 mm and 10 m in height on different pole rotor disc in order to obtain the minimum disc dimension with maximum magnetic flux distribution. Simulation was performed by Finite Element Method Magnetics (FEMM). The simulation result was presented in the form of magnetic field distribution geometry on two permanent magnets inside rotor disc. It showed that the magnetic field strength in the midpoint between different pole single rotor was very low (44 mT) on the distance of 25 mm.**

Keywords—magnet distance, magnet rotor disc dimension, FEMM

I. INTRODUCTION

Axial Flux Permanent Magnet Generator (AFPMG) has been increasingly more interesting to develop, particularly on certain application with its compact and portable dimension. This generator uses wind turbine or pico turbine/micro hydro for its prime over. Therefore, AFPMG can meet the renewable resources in terms of reducing global warming effect. On its installation, it needs construction design as commpatible to its application. Axial flux machine configurations can be constructed by referring to the stator position toward rotor position and the winding arrangement give freedom to choose the machine structure compatible to its application [1]. Axial flux configuration enables a simple, easy to make generator construction and relatively low cost [2] - [8]. The permanent magnet type NdFeB (Neodymium-Ferry-Boron) usage which has high density, easy to get and relatively low cost was compared to magnet from other material.The usage of this magnet type can also used to avoid brushes and slip ring [9], [10]. This can result in the elimination of routine maintenance cost. Rotor and stator construction in this study will use non metal material to reduce the weight and beside that, it has corrosive resistance [9], [11] - [13]. The previously massive and space consumed construction can recently become compact and the cogging torque can be reduced [14],[15].Its construction was used to be in a big manufacturer but now, it can be produced in small scale industry or even in home industry.

One of AFPMG main problems is rotor disc diameter dimension influenced by the distance between magnets in rotor disc. Generator is designed to operate on low speed, thus it's needed a large amount of magnet elements and disc diameter is directly proportional with the number of magnet elements used. This means that the disc dimension is getting bigger. The second problem is that there's still magnet leakage so that it can reduce efficiency.

The alternative solution used to solve magnet leakage is by arranging the distance between magnets in rotor disc or constructing isolation (magnetic shielding). The second alternative is more difficult to conduct, as the isolator material in Indonesia has been still difficult to obtain and the price is also expensive. The distance between magnets is also the most appropriate solution for this model. In this research, tool for analysis was Finite Element Method Magnetics (FEMM) software [16].

II. MAGNETIC FIELD MATHEMATICAL MODEL

Constant magnetic field density (B) or constant magnetic field intensity (H) can be obtained through several ways, namely from permanent magnet and an electrical field that linearly changes toward time.

The assumption in this study will be rotor using permanent magnet to generate magnetic field. Gauss law for magnetism:

$$\oint_S B.dS = \int_V (\nabla.B).dV = 0 \quad (1)$$

Where:

B = magnetic field density *((Wb/m^2)=Tesla)*,
dS = surface integration element, and
dV = volume integration element.

This result enables the determination of potential vector magnetic field density:

$$B = \nabla \times A \quad (2)$$

Whilst Ampere law for magnetic field in vacuum space:

$$\nabla \times B = \mu_0 J \quad (3)$$

Where: J = current density *(A/m^2)*.

Substitution of (2) to Ampere equation (3) results in:

$$\nabla^2 A = -\mu_0 J \quad (4)$$

Giving solution to potential magnetic:

$$A = \frac{\mu_0}{4\pi} \int_V \frac{J}{R} dV \quad (5)$$

By defening magnetic dipole moment: $m = IS\hat{r}$ thus this potential vector gives the vinal result:

$$A = \frac{\mu_0}{4\pi} \frac{m \times \hat{r}}{r^2} \quad (6)$$

So that the magnetic field density (2):

$$B = \frac{\mu_0}{4\pi} \left(\frac{\nabla \times m \times \hat{r}}{r^2} \right) \quad (7)$$

From the final solution (7), the result meets with the intuition/assumption that magnetic field density (B) inversely relate to distance (r^2), directly proportional with its magnetic moment dipole (m).

III. Flux Magnetic Influence Measurement

A magnet element as shown in Fig.1 and Fig. 2suggested how the magnetic field around it could be traced with the help of compass. Magnetic field lines outside of magnet circulated out from North Pole (N: North) and moved toward South Pole (S: South).

If two magnet elements were closed each other with the position of similar pole magnetic ends, they are NorthPole with North Pole or south with South Pole, so that their magnetic field pattern will have repulsive or attractive force.

Whilst two non similar magnet elements namely north and south pole, if the position of each magnetic end was brought to close, so that its magnetic field pattern as in Fig. 4 will experience attractive forces.

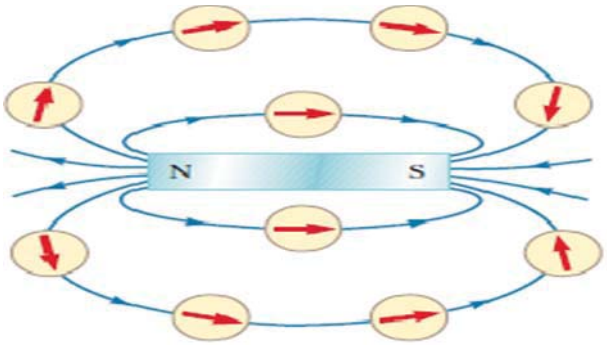


Fig. 1 Flux magnetic direction surrounding a bar magnet [18]

(2)

On rotor disc, there were permanent magnets arranged with the position N-S-N-S, as shown in Fig. 2.

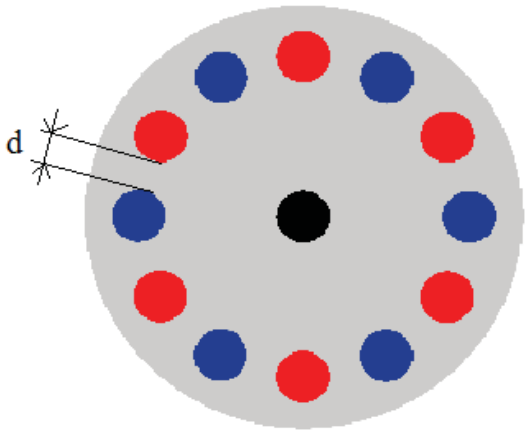


Fig. 2 Permanent magnet rotor disc

If rotor disc was seen on horizontal position, two magnet elements which were closed each other could be described as shown in Fig. 3.

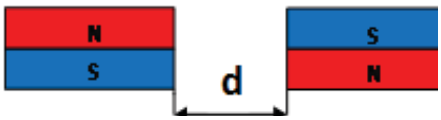


Fig. 3 Distance d between magnets

IV. Method

This study was focused on magnetic field distribution measurement around magnetic element. In order to do it, we needed three activity steps. First step was making study step as shown in Fig. 4. Second, it needed to validate magnetic distribution on a magnetic bar, two non similar magnets, and two similar magnets. The third step was performing magnetic distribution simulation on a rotor disc. Prior to perform either second or third step, before conducting simulation for each one, it created a geometry form first as presented in Fig. 8 that is a

geometric image on a magnetic bar. Fig. 9 shows a geometric image on two non similar magnetic bars, and Fig. 10 which was an image of two similar magnetic bars, as well as Fig. 11 that was a magnetic field distribution geometry image on a single rotor disc.

In creating a magnetic bar geometry, it's determined that its magnetic type was NdFeB in the form of cylinder with the diameter dimension 40 mm, thickness 10 mm, and permanent magnet type N40. This geometric form used point coordinates with software FEMM 4.2.

Prior to conducting magnetic field distribution measurement simulation between two magnets in a single rotor in order to be independent from the influence of each magnet, thus it's necessary to take the following process steps:

1) *Determining distance between two magnets by 10 mm; 15 mm; 20 mm; 25 mm dan 30 mm.*
2) *Forming geometry on FEMM 4.2 software by using coordinates (cylindrical magnet dimension with diameter 40 mm, height 10 mm).*
3) *Put in the function from each figure (the function used was NdFeB permanent magnet type N40, and the air as its medium).*
4) *Performing meshing process on that geometry.*
5) *Conducting run analysis to simulate geometric result that had been made.*
6) *Looking at the simulation result on View Result (presented as Flux Density and Plot Vector).*

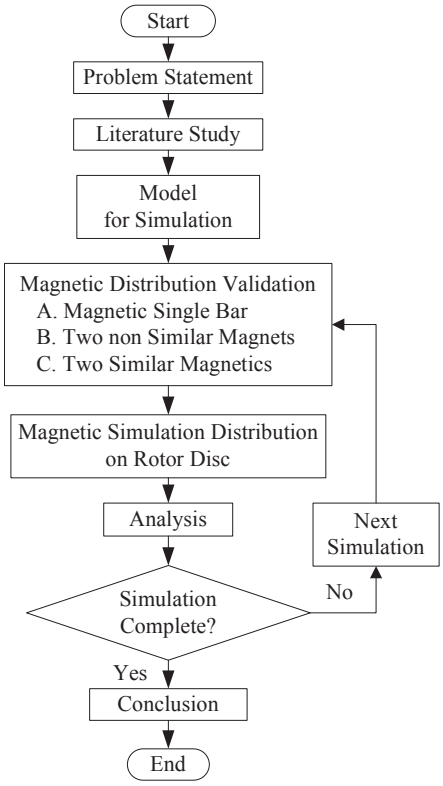


Fig. 4 . Research design

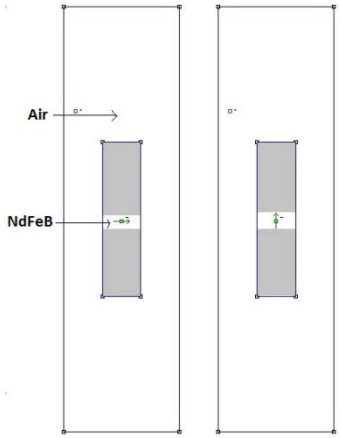


Fig. 5. Magnetic bar geometry

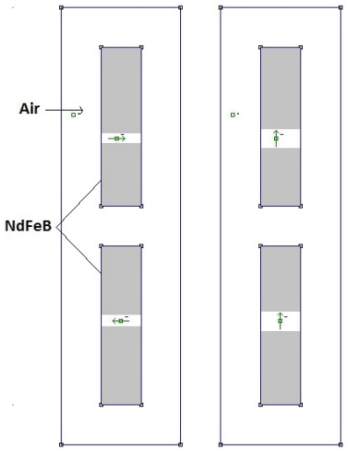


Fig. 6. Geometry of two magnets with opposite poles

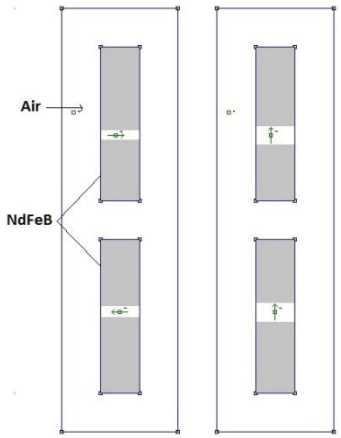


Fig.7. Geometry of magnets with same poles

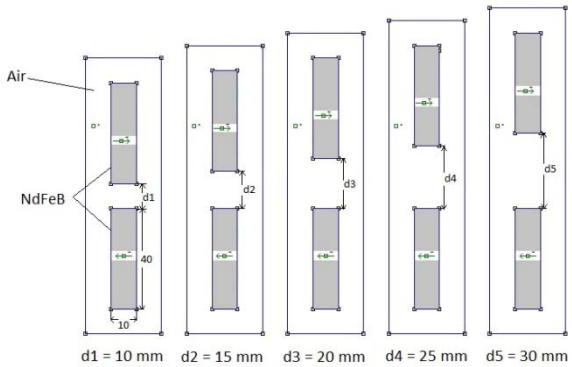


Fig. 8. Geometry of opposite-poles single-disc permanent mabnet rotor with distance between magnet of 10mm, 15mm, 20mm, 25mm and 30mm.

V. RESULT AND DISCUSSION

5.1. *Magnetic Field Disribution Simulation on Single Bar*

By utilizing coordinate on FEMM 4.2 software, from geometric shape on Fig. 12, it was obtained simulation result as the picture below.

Fig. 12(a) clearly showed that magnetic field lines were same with the arrow direction to the right which means that the side on the magnetic bar left by arrow was the north pole, and the side in which the arrow went toward it was the south pole. Similar to Fig. 12(b), the arrow direction was to the top, so that the upper side bar was north pole, and the lower side bar was south pole. Magnetic field strength magnitude on the middle surface of cylindrical magnet (Fig. 12(a)) with 40 mm in diameter and 10 mm in cylindrical thickness/height, or at Fig. 12(b) (10 mm in diameter and 40 mm in height) was around 50% from magnetic bar edge.

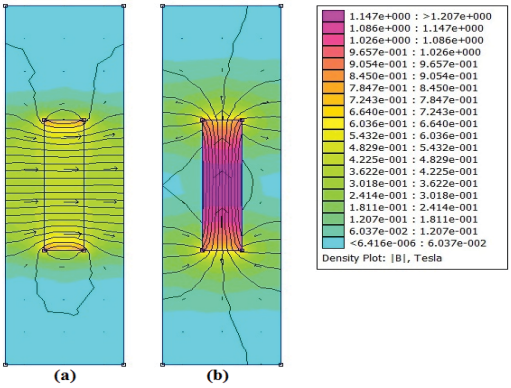


Fig. 9. Simulation of magnetic field distribution at a single bar magnet

5.2. *Magnetc Field Distribution Simulation on Two Magnet Bars with Opposite Pole*

From Fig. 6, it could be obtained simulation result as shown in Fig. 10. On the simulation figure, it's apparent that two non similar magnetic bars were brought to close each other (north pole with south pole or the contrary), thus (Fig. 10 (a)) in which the top magnetic bar had magnetic force lines directing to the bottom magnetic bar. It also occurred to the bottom magnetic bar, in which there were magnetic force lines directing to the top magnetic bar or each of them had magnetic force lines that attracted each other.

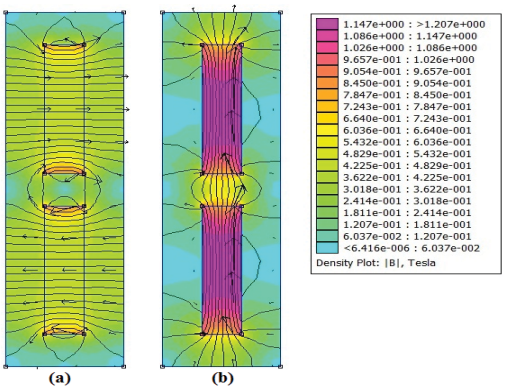


Fig. 10. Simulation of magnetic field distribution at two magnets with opposite poles.

5.3. *Magnetic Field DistributionSimulation on Two Magnet Bars with The Same Poles*

Whilst the magnetic field distribution simulation result on two non similar magnet elements on Fig. 11 (a), their magnetic force lines either from top or bottom magnet, each of them was to the left direction (avoiding) and on Fig. 11 (b) on the left side to the left direction and on the right side to the right direction (not fused).

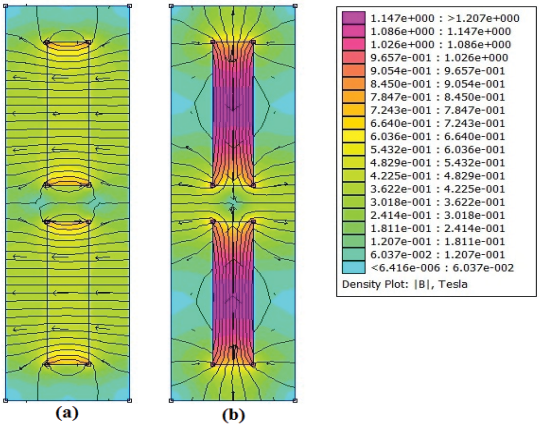


Fig. 11. Simulation of magnetic field distribution at two magnets with same poles.

It could subsequently be made a graph of association between magnetic field strength B with the distance d between magnets as in the following Fig. 13. From the figure, it could be suggested that on the distance 25 mm (mid point between top magnet and bottom magnet, i.e. a point on 12.5 mm from either top or bottom magnet) its magnetic influence was getting smaller, so that magnetic force lines on the top cylindrical magnet mostly directed to the right, likewise on bottom cylindrical magnet that they went toward left. The longer its distance (more than 25 mm), the lower/weaker was its magnetic field strength on the mid point between top and bottom magnet. On the simulation Fig. 12, magnetic field strength B magnitude on mid point 15 mm from top magnet and bottom magnet was 37 mT.

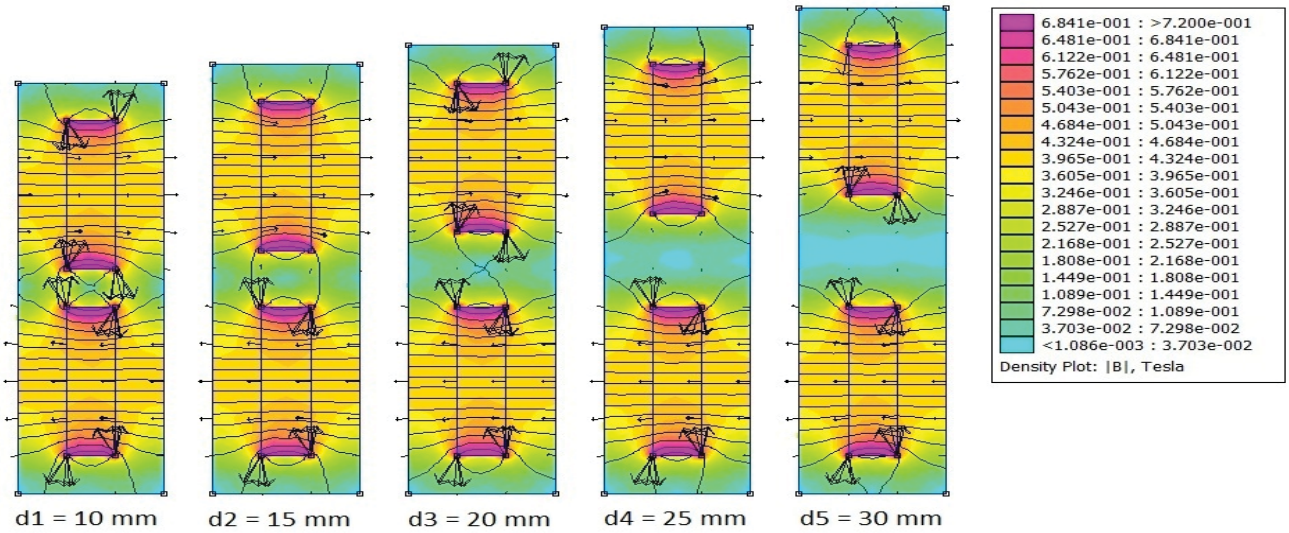


Fig. 12. Simulation of magnetic field distribution of opposite-poles single-disc peranent magnet rotor with distance between magnet of 10 mm, 15 mm, 20 mm, 25mm, and 30mm.

From the measurement result, it was obtained magnetic field strength value as listed on the following Table I.

DISTANCE BETWEEN MAGNETS D & MAGNETIC FIELD INTENSITY B

No	Distance d (mm)	Magnetic Field Intensity B (mT)
1	10	140
2	15	100
3	20	60
4	25	44
5	30	37

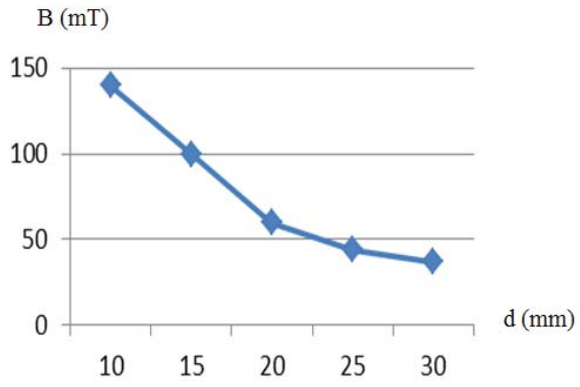


Fig. 13. Graph of the magnetic field (B) to the distance (d)

VI. CONCLUSION

Validation with magnetic bar, two non similar magnets, and two similar magnets showed that the simulation result met with magnetic field distribution simulation theory.

The simulation result exhibited that the greatest magnetic field magnitude was at the magnetic field pole edge. The further from the magnetic pole was, the weaker the magnetic field strength would be. Magnetic field strength at the middle position of magnetic bar was more or less 50 % of the magnetic field strength at the pole edge.

Magnetic field strength at the midpoint between magnets on single rotor disc permanent magnet with different pole was getting smaller with the increase of distance (the distance was more than 25 mm).

REFERENCES

[1] A. Parviainen, “Design of Axial-Flux Permanent Magnet Low-Speed Machines and Performance Comparasion Between Radial-Flux and Axial Flux Machines”, Thesis for Degree of Science)Technology) to be Presented with Due Permision for Public Examination and Criticsm in the Auditorium, 2005.

[2] Muljadi, Butterfield, Yih-Huei Wan, “Axial Flux, Modular, Permanenent Magnet Generator with a TorodialWinding for Wind Turbine Applications”, NREL/CP-500-24996 National Wind Technology Center, National Renewable Energy Laboratory, 1998.

[3] J.R. Bumby, “Torus: Aslottless,Torodial-Stator, Permanen Magnet Generator”, IEE Proceedings-B, 2006.

[4] J.R. Bumby, Stannard and R. Martin, “A Permanent Magnet Generator for Small Scale Wind Turbines”, Durham UK-DH 1 3LE, Paper 141, 2006.

[5] A.P. Ferreira, A.m. Silva, A.F. Costa, “Prototype of an Axial Flux Permanent Magnet Generator f Journal in School of Technology and Mana gement, Polytechnic Institute of Braganca or Wind Energy System Applications”, 2006.

[6] S.M. Hoseini, M. Agha-Mirsalim, M. Mirzaei, “Design, Prototyping and analysis of a low-cost disk permanent magnet generator with rectangular flat-shaped Magnets”, Iranian Journal Of Science & Technology, 2008.

[7] Howey, “Axial Flux Permanent Magnet Generators for Pico Hydropwer”, EWB-UKResearch Conference, 2009.

[8] A.P. Ferreira, A.F. Costa, “Direct Driven Axial Flux Permanent Magnet Generator for Small-Scale Wind Power Applications ",

[9] International Confrence on Renewable Energies and Power Quality, 2010.

[10] R.R. Walace, T.A. Lipo, L.A. Moran, J.A. Tapia, "Design and Construction of a Permanent Magnet Flux Sinchronous Generator", 0-7803-3946-097/S10.00-1997 IEEE, 1997.

[11] Enyuan, Guangxiao, Yongxing, Xiongying, Jiyan, "The Application of NdFeB in The Magnetic Force Actuator", Telkomnika Indonesian Journal of Electrical Engineering, 2012.

[12] F. Caricchi, Member, IEEE, F. Crecimbini, Member, IEE, O. Honorati, Member IEEE, G. L. Bianco, and E. Santini, Member, IEEE, "Performence of Coreless-Winding Axial Flux Permanent-Magnet Generator with Power Outputat 400 Hz, 3000r/min", 1998.

[13] J.R. Bumby and R. Martin, "Axial-Flux Permanent-Magnet Air-Cored Generator for Small-Scale Wind Turbines", IEE Proc.-Electr. Power Appl., Vol 152, No.5, September 2005.

[14] G.F. Price, T.D. Batzel, M. Comanescu, and B.A. Muller, "Design and Testing of A Permanent Magnet Axial Flux Wind Power Generator", Proceeding ofs of The 2008 IAJC-IJME International Conference ISBN 978-1-60643-379-0, 2008.

[15] M. Aydin, Z.Q. Zhu,T.A. Lipo, dan D. Howe, "Minimization of Cogging Torque in Axial-flux Permanent-Magnet Machines : Design consepts", IEEE Transaction On Magnetics, 2007.

[16] Tae-Uk Jung dan Jun-Seok Cho, "Electromagnetic Structural Design Analysis and Performance Improvement of AFPM Generator for Small Wind Turbine ", Journal of Magnetics, 2011.

[17] D. Meeker, Software Finite Element Method Magnetics FEMM Version 4.2., dmeeker@iee.org., 2010.

[18] D. Halliday and R. Resnick, "Physics", 3th Edition, John Wiley & Sons, Inc, 1987.

[19] J. Walker, D. Halliday and R. Resnick, "Fundamentals of Physics", 8th Edition, John Wiley, Incorported, April 2007.

Study of Excitation and Governor Control Effects of Superconducting Generator with High Response Excitation on the Stability of a SMIB Power System

F. Mayouf[1,3*](Adjeroud), F. Djahli[2], A. Mayouf[3], T. Devers[4]

[1]Department of Electrical Engineering, University of Sétif 1, Algeria.
[2]Department of Electronics, University of Sétif 1, Algeria.
[3]DIMMER Laboratory, University of Djelfa, Algeria.
[4]Department of Industrial Engineering and Maintenance, University of Orléans, France
[*]fa.adjeroud@gmail.com

***Abstract*— Superconducting generators (SCGs) are recently expected to substitute conventional machines in modern power systems. They are known for their many advantages such as light weight, small size and high efficiency. Self-excited SCGs, with high response excitation effect, have supplementary property that may be used for enhancing transient power system stability. Hence, the control of this type of generators becomes increasingly important. Because of the long time constant of the SCG, the control of excitation only is not sufficient. In this paper, we study the enhancement of the power system stability by implementing power system stabilizer (PSS) into the excitation (EPSS) and/or turbine governor (GPSS) systems of the SCG with high response excitation. Non-linear simulation results of a single machine infinite-bus power system, under different operating conditions, show the effectiveness of using exciter-based stabilizer in conjunction with the governor stabilizer (EGPSS).**

Keywords—Superconducting generator; high response excitation; power system stabilizer; single machine infinite-bus power system.

I. INTRODUCTION

Interconnected electrical power systems have been continuously expanding in scale and evolving into more complex structures in past times. Therefore, higher stabilities and efficiencies are more and more necessary. The stability of the power system is one of the most important aspects in electric system operation. This occurs from the fact that the power system must conserve frequency and voltage levels in the wanted level under any disturbance. Along with the considerably increase of electrical power demand, conventional generators of increased size are continually added to the system. However, the trend of these parameters tends to reduce stability and adversely affect system performance [1-6].

Superconducting generators (SCGs) are the optimal solution to overcome the uprating problems of conventional synchronous machines. They are expected to substitute these lasts in modern and complex power systems [6-9]. SCGs have various advantages such as small size and weight, low synchronous reactance, high efficiency and high power density. Therefore, this type of generators is very useful from viewpoints of steady state and transient stability and also from the viewpoints of reactive power supply [4-7].

Along with the progress of superconducting technologies, several types of superconducting machines have been developed. In general, SCGs can be subdivided into three categories: SCGs with low response excitation type A and B and SCGs with high response excitation which are different in characteristics and structures [10].

Several works concerning transient performance of SCGs have been published in the last few years [1-2,6-13]. However, there are few studies that have focused on their control and monitoring. Most of this works are interested to low response excitation SCGs type [6,14,15]. In this kind of generators, and because of the very long field time constant and the screening system, the control of the SCG via excitation system is not effective. Hence, the governor control loop becomes the sole solution for performance improvement of the system [6]. However, it is not the case for SCG with high response excitation where we can use the two solutions.

SCGs with high response excitation have an additional property that may improve transient power system stability. It has a power transfer limit greater than that of the SCG with constant excitation [13]. In the SCG with high response excitation, whose rotor has thermal radiation shield without damping, the field current can be changed rapidly according to the power system requirement [8-10]. Therefore, the exciter power is large enough to affect the conditions power system in self-excited operation [12]. Furthermore, the SCG with high response excitation has the so-called superconducting magnetic energy storage (SMES) effect. This is due to superconducting wire of the exciter field winding which is connected to the generator terminal through an AC-DC converter [8,9]. For these reasons, the excitation control system of SCGs with high response excitation is suggested. Tanzo Nitta et.al [7,11,12] have carried out experimental studies on the effect of SCG with high response excitation on power system stability and its excitation control system.

Generator excitation controls via power system stabilizers (EPSS), are a basic control but not the unique tool. Power system stability can be also enhanced by controlling the governor system which provides adjustment of the mechanical torque. The main advantage of applying a stabilizer into the governor loop (GPSS), lies in the fact that the turbine governor dynamics are weakly coupled with those of the rest parts of the system. Therefore, the settings of the GPSS are not dependents of the network parameters. In addition, in the case of SCG, the use of supplementary excitation stabilizing signal alone gives a little help to improve power system stability [16].

For conventional generators we have improved in previous work [17-20] the stability a single machine infinite-bus (SMIB) power system by implementing stabilizer in both excitation and governor loops (EGPSS).

In this paper, we extend this study to the superconducting generator with high response excitation. We study the effect of the proposed controller EGPSS on stability improvement of a SMIB power system under small and large perturbations. To show effectiveness of the EGPSS, we have compared its performance with other cases: open loop and EPSS and GPSS systems.

II. SCG WITH HIGH RESPONSE EXCITATION

A cross-section of the structure of SCG is illustrated in Fig. 1.

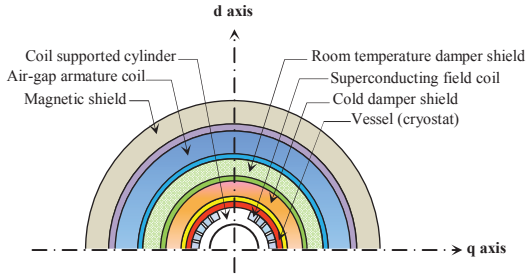


Figure 1. Schematic cross section of SCG.

Unlike conventional generators, there is no need for core structure in the stator and rotor of the SCG. This allows the armature to be place in a low permeance space. Consequently, this technology leads to low synchronous reactance, light weight and no flux saturation effect. The stator mainly consists of the air-gap armature coil. The only possibility is to use iron around the stator as a magnetic shield to protect the external environment from high fields [16,21]. The rotor in the SCG consists of superconducting field coil refrigerated in a cryogenic liquid like helium, coil support cylinder, vessel and dampers. In general, SCG has two dampers to shield the superconducting field coil, which are cold damper and warm (room temperature) damper [8-11]. The role of the first damper is to protect low temperature layer from being heated by the external layer [3]. Up to now, three types of SCG were developed: slow response excitation type A & B and high response excitation type [10].

Contrary to the two first types, in the SCG with high response excitation, the field current can be changed rapidly according to the requirement of the power system [8-10]. This is achieved by using special field winding materials that can generate very high magnetic field with reasonable speed and low losses. High-response excitation characteristic is achieved by changing the excitation voltage in order to change the induced voltage of the generator. In this type of SCG's, the field winding is a pulse magnet and the radiation shield does not play the role of a low temperature damper. Furthermore, the exciter power is large enough to affect the conditions power system in self-excited operation and the power system stability can be enhanced by controlling the field current. In order to well reflect the change of magnetic field to the stator, the SCG with high response excitation has only warm damper and its equivalent circuit can be reduced to that of the conventional generator [12,21-23].

III. MODELING

The proposed general control model of the SCG connected to an infinite bus through a step-up transformer and double circuit transmission line is shown in Fig. 2. The SCG is driven by a hydraulic turbine equipped with a governor.

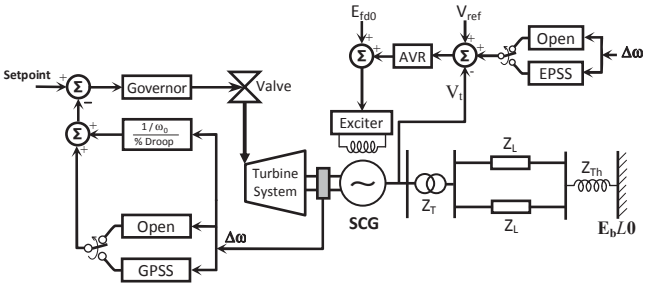


Figure 2. Proposed control model of SCG in a SMIB power system.

A. SMIB power system model

The SCG with high response excitation which has only warm damper is modeled with the so called 2.1 model. This model includes a field circuit (f) and one damper winding on the d-axis (D damper) and another on the q-axis (Q damper) representing the shield. The electric dynamic performance of the SCG in the d-q reference frame is described by the set of the following equations [23-24]:

Stator equations:

$$\begin{bmatrix} v_d \\ v_q \\ v_o \end{bmatrix} = \frac{-1}{\omega_b}\frac{d}{dt}\begin{bmatrix} \varphi_d \\ \varphi_q \\ \varphi_o \end{bmatrix} + \frac{\omega}{\omega_b}\begin{bmatrix} -\varphi_q \\ \varphi_d \\ 0 \end{bmatrix} - \begin{bmatrix} r_a & 0 & 0 \\ 0 & r_a & 0 \\ 0 & 0 & r_a \end{bmatrix}\begin{bmatrix} i_d \\ i_q \\ i_o \end{bmatrix} \quad (1)$$

Rotor equations:

$$\begin{bmatrix} v_f \\ 0 \\ 0 \end{bmatrix} = \frac{1}{\omega_b}\frac{d}{dt}\begin{bmatrix} \varphi_f \\ \varphi_D \\ \varphi_Q \end{bmatrix} + \begin{bmatrix} r_f & 0 & 0 \\ 0 & r_D & 0 \\ 0 & 0 & r_Q \end{bmatrix}\begin{bmatrix} i_f \\ i_D \\ i_Q \end{bmatrix} \quad (2)$$

where the flux linkages are expressed by:

$$\begin{bmatrix}\varphi_d\\ \varphi_f\\ \varphi_D\end{bmatrix}=\begin{bmatrix}x_d & x_{fd} & x_{dD}\\ x_{fd} & x_f & x_{fD}\\ x_{dD} & x_{fD} & x_D\end{bmatrix}\begin{bmatrix}I_d\\ I_f\\ I_D\end{bmatrix} \quad (3)$$

$$\begin{bmatrix}\varphi_q\\ \varphi_Q\end{bmatrix}=\begin{bmatrix}x_q & x_{qQ}\\ x_{qQ} & x_Q\end{bmatrix}\begin{bmatrix}I_q\\ I_Q\end{bmatrix} \quad (4)$$

To build a SCG's model adequate for use in simulation studies, we assume that:

$$x_d=x_l+x_{ad},\ x_{fd}=x_{ad},\ x_q=x_l+x_{aq},\ x_{qQ}=x_a \quad (5)$$

Applying the "shell constraints":

$$x_{fD}=x_D \quad (6)$$

The flux interlinkages become:

$$\begin{bmatrix}\varphi_d\\ \varphi_f\\ \varphi_D\end{bmatrix}=\begin{bmatrix}x_d & x_{ad} & x_{ad}\\ x_{ad} & x_f & x_D\\ x_{dD} & x_D & x_D\end{bmatrix}\begin{bmatrix}I_d\\ I_f\\ I_D\end{bmatrix} \quad (7)$$

$$\begin{bmatrix}\varphi_q\\ \varphi_Q\end{bmatrix}=\begin{bmatrix}x_q & x_{aq}\\ x_{aq} & x_Q\end{bmatrix}\begin{bmatrix}I_q\\ I_Q\end{bmatrix} \quad (8)$$

The mechanical equations of the rotor are:

$$\frac{d\delta}{dt}=\omega_b(S_m-S_{m0}) \quad (9)$$

$$\frac{dS_m}{dt}=\frac{1}{2H}\left[T_m-T_e-D(S_m-S_{m0})\right] \quad (10)$$

where:

$$T_e=\varphi_d i_q-\varphi_q i_d \quad (11)$$

To simplify the generator model, times constants subtransient and transient reactances are used. Then, the SCG equations written in a state-space model are as fllows:

$$\frac{dE'_q}{dt}=\frac{1}{T'_{do}}\left[E_{fd}-\alpha E'_q+(\alpha-1)E''_q\right] \quad (12)$$

$$\frac{dE''_q}{dt}=\frac{1}{T''_{do}}\left[E'_q-\frac{x'_d+x_e}{x''_d+x_e}E''_q+\frac{x'_d-x''_d}{x''_d+x_e}E_b\cos(\delta)\right] \quad (13)$$

$$\frac{dE''_d}{dt}=\frac{1}{T''_{qo}}\left[-\frac{x_q+x_e}{x''_q+x_e}E''_d-\frac{x_q-x''_q}{x''_q+x_e}E_b\sin(\delta)\right] \quad (14)$$

$$T_e=E''_d i_d+E''_q i_q+\left(x''_d-x''_q\right)i_d i_q \quad (15)$$

The stator flux linkages are:

$$\begin{cases}\varphi_d=x''_d\,i_d+E''_q\\ \varphi_q=x''_q\,i_q-E''_d\end{cases} \quad (16)$$

where:

$$E'_q=\frac{x_{ad}}{x_f}\varphi_f,\ E''_q=\frac{x_{ad}}{x_D}\varphi_D,\ E''_d=-\frac{x_{aq}}{x_Q}\varphi_Q,\ E_{fd}=x_{ad}\frac{v_f}{r_f} \quad (17)$$

and the subtransient and transient reactances are:

$$x''_d=x_d-\frac{x^2_{ad}}{x_D}\quad x'_d=x_d-\frac{x^2_{ad}}{x_f},\ x''_q=x_q-\frac{x^2_{aq}}{x_Q}, \quad (18)$$

the time constants are:

$$T'_{do}=\frac{x_f}{\omega_b r_f}\quad T''_{do}=\frac{x_D}{\alpha\omega_b r_D},\ T''_{qo}=\frac{x_Q}{\omega_b r_Q}, \quad (19)$$

and,

$$\alpha=\frac{x_d-x''_d}{x'_d-x''_d} \quad (20)$$

Neglecting the armature resistance, the stator transients and ignoring speed variations, the d-q machine terminal voltage equations related to the infinite bus are:

$$\begin{cases}v_d=r_e i_d+x_e i_q-E_b\sin(\delta)=E''_d-x''_q i_q\\ v_q=r_e i_q-x_e i_d+E_b\cos(\delta)=E''_q+x''_d i_d\end{cases} \quad (21)$$

and,

$$V_t=\sqrt{v_d^2+v_q^2} \quad (22)$$

where:

$$Z_e=r_e+jx_e=Z_T+Z_{Th}+\frac{Z_L}{2} \quad (23)$$

B. Exciter control system for SCG

High-response excitation is achieved by varying the excitation voltage which enables the rapid change of field current. Because of the large time constant of the field winding T'_{do} in SCG with high response excitation, the capability of the exciter is larger than that for the conventional generator. For a SCG with high response excitation and in self-excitation mode, the variation of exciting power is considerable large to act on the power system stability. The excitation system required for SCGs is bi-polar, in this study, a thyristorized exciter is used. The exciter coupled with the field winding made of wire superconducting, is considered as superconducting magnetic energy storage (SMES). During modeling, we suppose that the SMES effect is considered as a nonlinear load.

The exciter of the SCG is controlled by an automatic voltage regulator (AVR) which has a large gain. However,

large gain of AVR leads to system instability and therefore, a PSS is needed. To reach moderate rates of change of SCG field flux, larges per unit values of exciter ceiling voltage are required. However, large per unit values does not mean significant high physical values (in volts) [16].

The block diagram of a speed input conventional PSS applied to the excitation system (EPSS) is shown in Fig. 3. According to the large gain of the AVR, the stabilizer gain (K_{PSS}) for SCG is also larger than that of the conventional one.

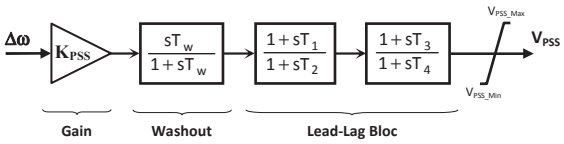


Figure 3. Synoptic of Lead-Lag Power System Stabilizer.

C. *Governor control system for SCG*

The structure of the proposed governor power system stabilizer (GPSS) is the same of the exciter-based stabilizer (EPSS). The stabilizer considered in this work is a lead-lag compensator. It has the same design of the conventional exciter-based power system stabilizer (EPSS) except that the gain of the GPSS (K_{GPSS}) must be larger than that of the EPSS because of its large time constant [25]. Furthermore, the gain K_{GPSS} for small disturbance is chosen $2.5\times K_{GPSS}$ for large disturbance. The output of the GPSS is injected in the governor loop to supply additional damping torque to the SCG.

IV. NONLINEAR SIMULATION RESULTS

The dynamic performance of single SCG with high response excitation, connected to an infinite bus power system, has been analyzed with the proposed PSS (Open loop, EPSS, GPSS and EGPSS), under various disturbances: small perturbation and three-phase fault. The parameters of the simulated system are as follows:

Superconducting generator constants:

$x_d = x_q = 0.36\,pu,\ x'_d = 0.256\,pu,\ x''_q = 0.2\,pu,$

H=3.7s, D=0, $T'_{do} = 15.88s,\ T''_{do} = 0.107s,\ T'_{qo} = 0.33s.$

Exciter: K_A=400, T_A=0.025s, -60 pu $\le E_{fd} \le$ 60pu.

Transmission line: R_L=0.08593pu, X_L=0.8125pu, B_c=0.1184pu.

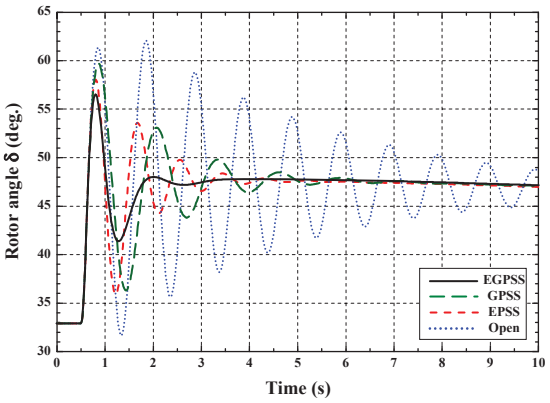
Transformer: R_T=0pu, X_T=0.1364pu, **Thevenin:** X_{Th}=0.13636pu.

EPSS: K_{PSS}=5, **GPSS:** K_{GPSS}=35 (for 3-phase fault).

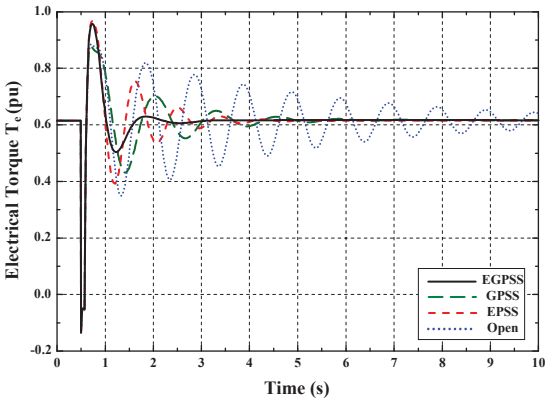
To investigate the effectiveness of EGPSS, we have simulated with EPSS, GPSS, EGPSS and without stabilizer a four-cycle three-phase fault (Fig. 4) and a step increase in mechanical input power (Fig. 5) in SMIB power system. Our results concern: rotor angle (δ), electrical torque (T_e) and speed deviation (Δω) for the four cases: open loop (without PSS), EPSS, GPSS and EGPSS.

A. *Large disturbance test*

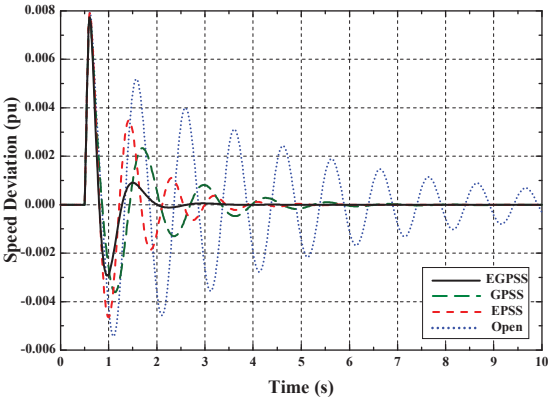
In Fig. 4, a four-cycle three-phase fault was applied at the sending end of one of the circuits of the transmission line at time t=0.5s. The fault is cleared by tripping the faulty line.



(a) Variation of rotor angle δ (deg.).



(b) Variation of electrical torque T_e (pu).

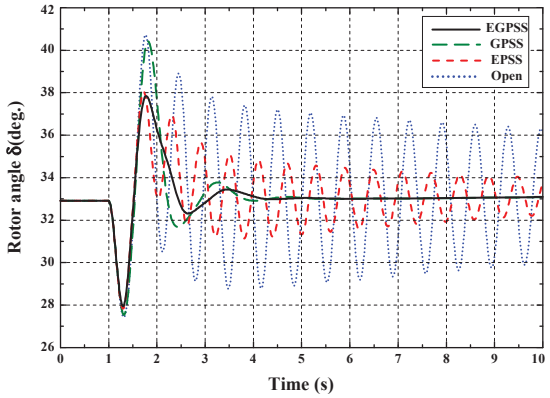


(c) Variation of speed deviation (pu).

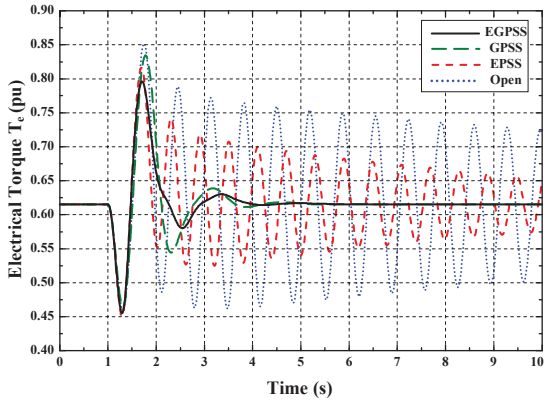
Figure 4. System Response to a four-cycle three phase fault.

B. *Small disturbance test*

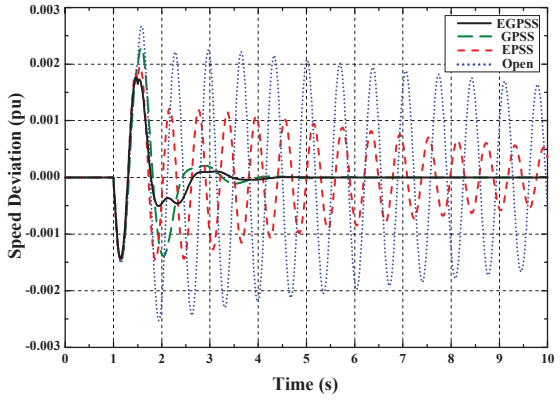
Fig. 5 presents the simulation of a 10% step increase in P_{mec} for 0.5s applied at time=1s.



(a) Variation of rotor angle δ (deg.).



(b) Variation of electrical torque T_e (pu).



(c) Variation of speed deviation (pu).

Figure 5. System Response to a 10% step incease in P_{mec} for 0.5s.

In the two types of disturbances, and in the absence of PSS (open loop), the system response presents important and dangerous oscillations. With the implementation of stabilizers whether in the exciter or governor systems, the oscillations are clearly reduced. Because the SCG has a long field time, the use of an exciter-based stabilizer (EPSS) alone, is not very sufficient to damp oscillations. Addition of a PSS applied to the governor system (GPSS) presents more improvements. It can be seen that the proposed stabilizer (EGPSS), which consists in using exciter-based stabilizer (EPSS) in conjunction with the governor stabilizer (GPSS), have damped the system oscillations effectively and hence have enhance clearly the stability of the SCG connected to an infinite bus power system.

V. CONCLUSION

In this paper, the implementations of PSS via the governor-turbine and excitation systems of a superconducting generator connected to an infinite bus power system have been studied. The results of analysis and simulation have shown that the EGPSS can be used to suppress low-frequency oscillations in power systems. This arrangement of stabilizer possesses better robustness to the changes of power system operating conditions and presents a useful tool for control of superconducting machines.

VI. REFERENCES

[1] M. A. A. S. Alyan, Y. H. A. Rahim, and S. M. Osheba, "Assessment of steady state stability of a multi-machine power system with and without a superconducting alternator", IEEE Trans. Power Systems, vol. 5, no. 3, pp. 730-736, Aug. 1990.

[2] S. M. Osheba, M. A. A. S. Alyan and Y. H .A. Rahim, "Comparison of transient performance of superconducting and conventional generators in a multimachine system". IEE proceedings, C, Generation, Transmission and Distribution, vol. 135, no. 5, pp. 389-395, sept. 1988.

[3] H. Shirato, G. Wu, S. Oikawa and H. Sisido, "Investigation of power system stability enhancement by superconducting generator with high response excitation considering its detailed excitation system", IEEE Int. Conf. on Power System Technology (POWERCON 2010), Oct. 24-28, 2010, Hangzhou, pp. 1-6.

[4] S. Woodruff, M. Steurer, and H. Boenig, "Experimental determination of dynamic parameters for a superconducting machine", IEEE Trans. Applied Superconductivity, vol. 15, no. 2, pp. 2154-2157, June 2005.

[5] Y. Shirai, T. Nitta, and M. Yamada, "Experimental study on system stability evaluation in parallel running of a superconducting generator and a SMES", IEEE Trans. Applied Superconductivity, vol. 12, no. 1, pp. 890-895, March 2002.

[6] M.A.A.S. Alyan, and Y.H. Rahim, "The role of governor control in transient stability of superconducting turbo-generators", IEEE Trans. Energy Conversion, vol. 2, no. 1, pp. 38-46, 1987.

[7] T. Nitta, Y. Shirai, T. Okada, Y. Imai, and Y. Kobayashi, "High response excitation control of superconducting generator for stability of superconducting field winding", IEEE Trans. Applied Superconductivity, vol. 10, no. 1, pp. 935-938, March 2000.

[8] W. Sae-Kok, A. Yokoyama, T. Nitta, "Excitation control system design of superconducting generator with high response excitation considering SMES effect for stability improvement", TENCON 2004. 2004 IEEE Region 10 Conference, vol. C, Nov. 24-26, 2004, Chiang Mai, pp. 512-515.

[9] W. Sae-Kok, A. Yokoyama, T. Nitta, "Excitation control system design of superconducting generator with high response excitation in consideration of SMES effect for improving stability in multi-machine power system", 15th Power Systems Computation Conference PSCC, Aug. 22-26, 2005, Liège, Belgium, 2005, pp. 1-7.

[10] K. Arai, M.Kazumori, T. Ichikawa, Y. Nakabayashi, S. Ohshima ,Y. Matsunobu, H. Sawazaki, Y. Yagi, A. Ueda, and T. Kitajima, "Development of 70Mw class superconducting generators", IEEE Trans. Magnetics, vol. 30, no. 4, pp. 1875-1878, July 1994.

[11] T. Nitta, Y. Shirai, and T. Kishida, "Experimental studies on field circuit of superconducting generator", IEEE Trans. Applied Superconductivity, vol. 7, no. 2, pp. 939-542, June 1997.

[12] T. NittaT, T. OkadaT, Y. Shirai, T. Kishida, Y. Ogawa, H. Hasegawa, K. Takagi, and H. Matsumoto, "Experimental studies on power system stability of a superconducting generator with high response excitation", IEEE Trans. Power Systems, vol. 12, no. 2, May 1997, pp. 906-912.

[13] K. Miyaike, T. Kitajima, and T. Ito "Development of 70 MW class superconducting generator with quick-response excitation" Cryogenics vol. 42, pp. 183-189, 2002, Translation of article originally published in Cryogenic Engineering (Journal of Cryogenic Association of Japan), vol. 36, pp. 125-130, 2001.

[14] M. A. A. S. Alyan and Y. H. Rahim, "A Discrete state-space controller for superconducting turbo-generators", IEEE Trans. Energy Conversion, vol. 3, no. 2, pp. 300-304, June 1988.

[15] R. A. F. Saleh and H. R. Bolton, "Genetic algorithm-aided design of a fuzzy logic stabilizer for a superconducting generator" IEEE Trans. Power Systems, vol. 15, no. 4, pp. 1329-1335, Nov. 2000.

[16] T. H. Einstein, "System performance characteristics of superconducting alternators for electric utility power generation", IEEE Trans. Power Apparatus and Systems, vol. PAS-94, no. 2, pp. 310-319, March 1975.

[17] F. Mayouf(Adjeroud), F. Djahli and A. Mayouf, "Study of excitation and governor power system stabilizers effect on the stability enhancement of a single machine infinte-bus power system", The 12th Int. Conf. on Environment and Elect. Eng. (EEEIC-2013), 5-8 May 2013, Wroclaw, Poland, pp. 534-538.

[18] F. Mayouf(Adjeroud), F. Djahli, A. Mayouf and T. Devers, "Study of excitation and governor power system stabilizers effect on the sability enhancement of a multimachine power system", The 9th Int. Symposium on Diagnostics for Electrical Machines, Power Electronics & Drives (SDEMPED 2013), 27-30 August 2013, Valencia, Spain, in press.

[19] F. Mayouf(Adjeroud), F. Djahli, A. Mayouf and T. Devers, "A new controller based on fuzzy logic and PSS for enhancing the stability of a single machine ifinte-bus power system", The 9th Int. Symposium on Diagnostics for Electrical Machines, Power Electronics & Drives (SDEMPED 2013), 27-30 August 2013, Valencia, Spain, in press.

[20] F. Mayouf(Adjeroud), F. Djahli, A. Mayouf and T. Devers, "Study of fuzzy logic control and power system stabilizers effect on the stability enhancement of a SMIB power system", The 5th Int. Conf. on Information Technology and Electrical Engineering (ICITEE 2013), 7-8 Oct. 2013, Yogyakarta, Indonesia, in press.

[21] H. Nakamura, H. Matsumoto, H. Hasegawa, K. Nakanishi, K. Inoue, M. Sakai, and S. Ohshima, "Static excitation test results of the partial rotor model for 70 MW class superconducting generator with quick response excitation", IEEE Trans. Magnetics, vol. 32, no. 4, pp. 2357-2360, July 1996.

[22] T. Nitta, S. Tanaka, and T. Okada, "Characteristics of superconducting generator at high response excitation", Electrical Engineering in Japan, vol. 112, no. 5, pp. 26-35, 1992, Translated from Denki Gakkai Ronbunshi, vol. 111-B, no. 4, pp. 389-395, April 1991.

[23] I. Muta and S. Inadomi, "Comparison of dynamic stability of superconducting and conventional synchronous alternators" Electrical Engineering in Japan, vol. 100, vo. 6, pp. 98-106, 1980, Translated from Denki Gakkai Ronbunshi, vol. 100B, no. 12, pp. 739-746, December 1980.

[24] J. L. Kirtley, "Large system interaction characteristics of superconducting generators", Proc. of the IEEE, vol. 81, no. 3, pp. 449-461, Mar. 1993.

[25] H. F. Wang, Y. S. Hao, B. W. Hogg, and Y. H. Yang, "An adaptive power system stabilizer based on turbine governor control", IEE International Conference on Advances in Power System Control, Operation and Management, APSCOM-93, 7-10 Dec. 1993, vol. 1, pp. 285-289.

Teaching the Large Synchronous Generator Dynamic Model under Unbalanced Steady-State Operation

Sugiarto
Electrical Engineering Department
Sekolah Tinggi Teknologi Nasional
Yogyakarta, Indonesia
sugiarto.kadiman@gmail.com

Sasongko Pramono Hadi, Tumiran, F. Danang Wijaya
Electrical Engineering Department
Gadjah Mada University
Yogyakarta, Indonesia
sasongko@te.ugm.ac.id, tumiran@te.ugm.ac.id,
danang@te.ugm.ac.id

Abstract — **This paper presents an attractive approach for teaching the large three-phase synchronous generator under unbalanced steady-state condition of the 500 kV EHV Jawa-Madura-Bali (or Jamali) system to which it is connected using Matlab's Graphical User Interface (GUI) capability. Whereas two unbalanced steady-state conditions of the grid are obtained by setting all of IBTs (or inter bus transformers), the Jamali's grid loads, into load imbalance of 5% and 10%. The main motivation for such a study is to develop user-friendly software for better teaching the behavior of the generator which is connected to the grid when unbalanced loads are present. An example is given to demonstrate the usefulness of the developed tool.**

Keywords — Synchronous generator, 500 kV EHV Jamali system, GUI, IBT and unbalanced load.

I. INTRODUCTION

Many researchers have already studied numbers of modeling of synchronous generator methods. Each of their studies is made according for which the machine is design. Nowadays, there is an increasing demand to replace or supplement the detailed of prototyping of synchronous generator with mathematical method and computer simulation because of the increasing cost of it. The crucial importance of a good model of synchronous generator taking in account dampers and other elements which are sometime neglected in simplified model are shown in [1], [2], [3], [4] and [5].

This paper presents an effective approach for teaching the large synchronous generator under unbalanced steady-state operation using a Matlab environment. This approach allows a student to simulate different levels of unbalanced operation conditions of the generator easily using basic Matlab instructions. In addition, the student can learn Matlab's GUI capabilities to construct a user-friendly software package for teaching the synchronous generator.

In Matlab software [6], the synchronous generator blocks are designed to be balanced model and connected to a three-phase balanced or unbalanced networks. Unfortunately, the student may be difficult to represent the framework of three-phase synchronous generator. It is why three-phase balanced synchronous generator complete model will be presented in this study. The proposed model is based on the magnetic circuit of the machine; in general, the coupled-circuit approach is simpler to understand and is more widely adopted. In addition, generator's voltage inputs, which are unbalance, are derived utilizing unbalanced three-phase load-flow, EDSA 2000 [7], applied on 500 kV EHV Jamali System.

As a basis for developing an easy-to-use educational software tool, some of Matlab's GUI functions are implemented in creating an active link with this model. A brief explanation about modeling of the synchronous generator is defined on Section II. Section III presents the demonstration. Section IV presents the final conclusion obtained with the present study.

II. MODELLING OF THE SYNCHRONOUS GENERATOR

Considering a synchronous generator with six windings that are three stators, one rotor and two dampers [8], the general scheme of the machine is presented in Fig. 1. The proposed modeling method is based on this global scheme.

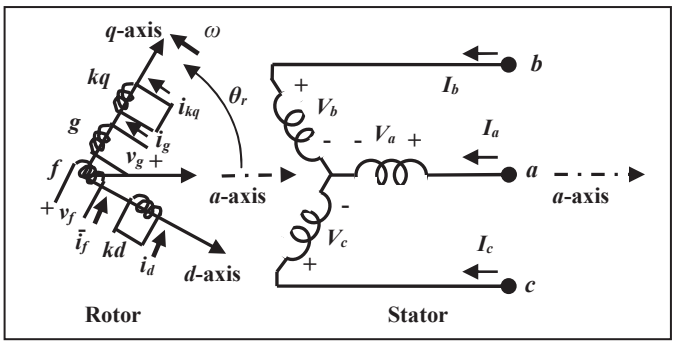


Fig. 1. Synchronous generator winding with dampers

A. Synchronous Generator Modeling by Electrical Circuits

The mathematical description or model develop is based on concept of an ideal synchronous generator. The fields produced by the winding currents are assumed to be sinusoidal distributed around the air-gap. This assumption of sinusoidal

field distribution ignores the space harmonics, which may have secondary effects on the machine's behavior. It is also assumed that stator slots cause no appreciable variation of any of the rotor winding inductances with rotor angle.

The machine equation in three-axes framework can be written as:

$$\begin{aligned} v_{abc} &= -r_s \cdot i_{abc} + \frac{d}{dt}\Psi_{abc} \qquad v_f = r_f \cdot i_f + \frac{d}{dt}\Psi_f \\ 0 &= r_{kq} \cdot i_{kq} + \frac{d}{dt}\Psi_{kq} \qquad 0 = r_{kd} \cdot i_{kd} + \frac{d}{dt}\Psi_{kd} \end{aligned} \tag{1}$$

Where i_{kd} and i_{kq} are the currents of direct and tranverse damper windings, Ψ_{kd} and Ψ_{kq} are the total flux of direct and tranverse damper windings, Ψ_{abc} is the stator total flux and Ψ_f is the main field total flux.

The two stator electromagnetic fields, both traveling at rotor speed, were identified by decomposing each stator phase current under steady-state into two components, one in phase with the electromagnetic field and other phase shift by 90^O. An air-gap field with its maxima aligned to *d*-axis; while other is aligned to the *q*-axis can be constructed by this framework. This method is called Park's transformation that gives the following relationship:

$$i_{dq0} = P \cdot i_{abc} \qquad i_{abc} = P^{-1} \cdot i_{dq0} \tag{2}$$

Where the current vectors are $i_{dq0} = \begin{bmatrix} i_0 & i_d & i_q \end{bmatrix}^T$ and $i_{abc} = [i_a \quad i_b \quad i_c]^T$ and Park's transformation matrix is:

$$P = \sqrt{\frac{2}{3}} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ i_a \cos\theta & i_b \cos\left(\theta - \frac{2\pi}{3}\right) & i_c \cos\left(\theta - \frac{4\pi}{3}\right) \\ i_a \sin\theta & i_b \sin\left(\theta - \frac{2\pi}{3}\right) & i_c \sin\left(\theta - \frac{4\pi}{3}\right) \end{bmatrix}$$

where i_a, i_b and i_c are the phase currents and θ is the angle between the phase current i_a and the current i_d.

Utilizing Park's transformation matrix, the differential equations of electrical dynamic that describe the stator and rotor windings are shown in (3) and are written in dq0 reference frame. The first three equations describe the stator winding (subscript *s*) and the following three equations describe the rotor winding (superscript *r*). The subscript *k* is used for the damping windings (*kq* for *q*-axis damping winding and *kd* for direct axis) while the subscript *f* is used for the field winding.

$$\begin{aligned} v_{qs}^r &= -r_s i_{qs}^r + \frac{\omega_r}{\omega_b}\psi_{ds}^r + \frac{p}{\omega_b}\psi_{qs}^r \\ v_{ds}^r &= -r_s i_{ds}^r + \frac{\omega_r}{\omega_b}\psi_{qs}^r + \frac{p}{\omega_b}\psi_{ds}^r \\ v_{0s}^r &= -r_s i_{0s}^r + \frac{p}{\omega_b}\psi_{0s}^r \qquad v_{kq}^r = -r_{kq} i_{kq}^r + \frac{p}{\omega_b}\psi_{kq}^r \\ v_{kd}^r &= -r_{kd} i_{kd}^r + \frac{p}{\omega_b}\psi_{kd}^r \qquad v_{fd}^r = -r_{fd} i_{fd}^r + \frac{p}{\omega_b}\psi_{fd}^r \end{aligned} \tag{3}$$

In these *v* represents the voltage of windings, *i* represents the electrical current flowing in the winding, ψ represents the magnetic flux linking the winding, *p* represents differential operator (d/dt), ω_r and ω_b are angular speed of the rotor referred to a two pole generator and reference angular speed corresponded to the rated frequency, respectively. The magnetic flux ψ for each winding is drawn in (4). The value of v_{kq}^r and v_{kd}^r are null since the damping windings are short-circuited.

$$\begin{aligned} \psi_{qs}^r &= -x_{ls} i_{qs}^r + x_{mq}\left(-i_{qs}^r + i_{kq}^r\right) \\ \psi_{ds}^r &= -x_{ls} i_{ds}^r + x_{md}\left(-i_{ds}^r + i_{fd}^r + i_{kd}^r\right) \\ \psi_{0s}^r &= -x_{ls} i_{0s} \\ \psi_{kq}^r &= x_{lkq} i_{kq}^r + x_{mq}\left(-i_{qs}^r + i_{kq}^r\right) \\ \psi_{kd}^r &= x_{lkd} i_{kd}^r + x_{md}\left(-i_{ds}^r + i_{fd}^r + i_{kd}^r\right) \\ \psi_{fd}^r &= x_{lfd} i_{fd}^r + x_{md}\left(-i_{ds}^r + i_{fd}^r + i_{kd}^r\right) \end{aligned} \tag{4}$$

where r_s, r_{kq}, r_{kd}, r_{fd}, x_{ls}, x_{lkq}, x_{lkd}, x_{lfd}, x_{mq} and x_{md} are the electrical fundamental parameters of synchronous generator. The direct-axis reactance x_d and the quadrature-axis reactance x_q are given by (5).

$$x_d = x_{ls} + x_{md} \qquad x_q = x_{ls} + x_{mq} \tag{5}$$

The mechanical part of the generator is described by two differential equations as described in (6).

$$\begin{aligned} p\delta &= \omega_r - \omega_s \\ \frac{2H}{\omega_s} p\omega_r &= T_m - \left(\psi_d i_{qs} - \psi_d i_{qs}\right) - T_{damp} \end{aligned} \tag{6}$$

In (6), H is an inertia constant of the turbine generator set, T_m is the mechanical torque of the turbine and T_{damp} is a damping torque. The damping torque represents the rotational losses of the rotating parts which consist of the magnetic losses and the mechanical losses.

B. Balanced Synchronous Generator Model

The balanced three-phase synchronous generator is shown in Fig. 2. It is clear that the generator is driven by the unbalanced voltage inputs. Fig. 3 shows the inside of generator block.

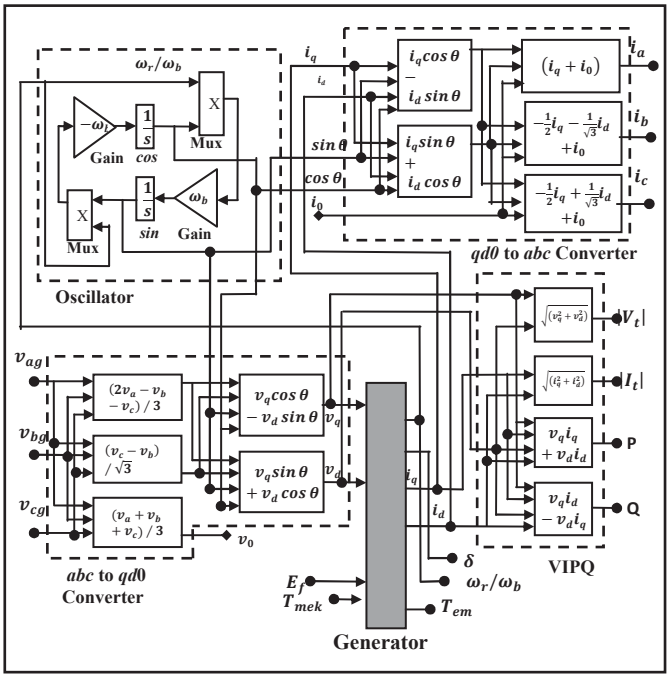


Fig. 2. Balanced Generator with Unbalanced Inputs

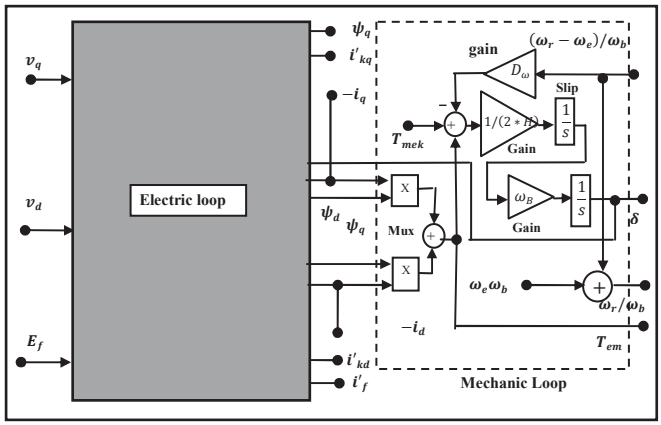


Fig. 3. Inside the Generator Block

C. *Studied System Description*

The studied generator is Tanjung Jati B which is one of generating plants of the 500 kV EHV Jamali System shown in Fig. 4. The grid consists of 9 generator nodes and 21 load nodes. The Paiton's bus is swing-node and others are PV nodes. System capacity is 100,000 MVA.

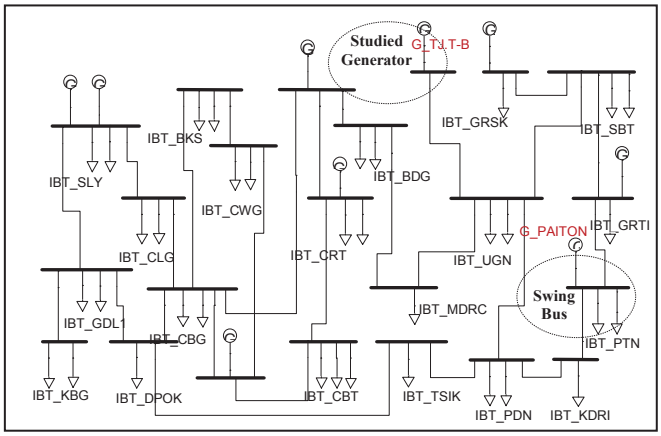


Fig. 4. *Single-line diagram* of 500KV EHV Jamali System

The student can access Matlab's GUI facilities to construct a software package for teaching synchronous generator under unbalanced steady-state condition. As an example of using Matlab's GUI capabilities, menu and plotting commands are implemented in a script file to provide interactive windows. The main menu, which is displayed after running the file, are shown in Fig. 5 and Fig. 6.

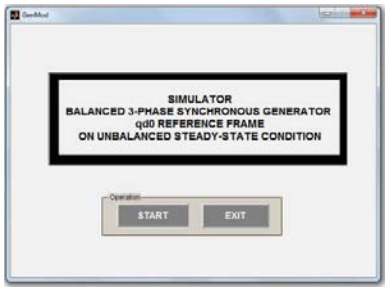


Fig. 5. The main window of the developed software tool

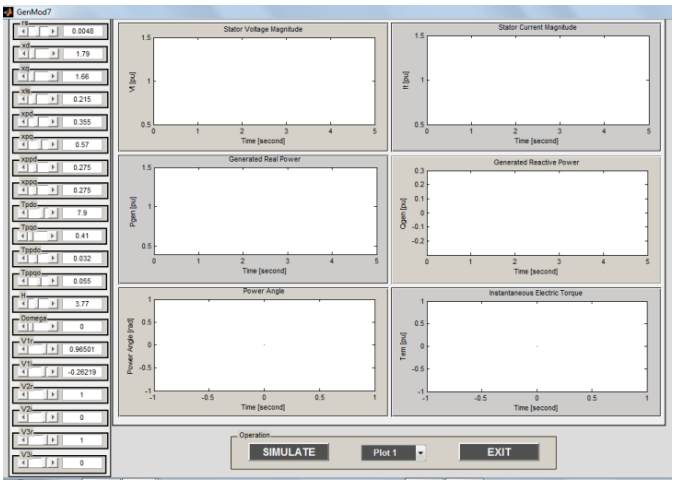


Fig. 6. The window of inserting the inputs for desired balanced synchronous generator and unbalanced inputs

III. DEMONSTRATION

The simulation of the proposed generator model is carried out by Matlab. As an inputs of this generator model, which are stator voltages, is derived by analyzing a single-line diagram of 500 kV EHV Jamali System on EDSA 2000 load-flow software. The process of numerical simulation method can be presented by the block diagram of Fig. 7.

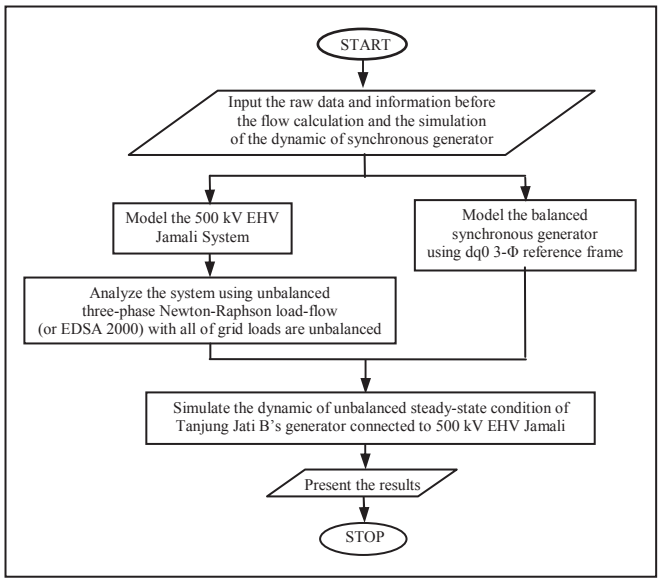


Fig. 7. Simulation flowchart

A. *Tanjung Jati B's Terminal Inputs Calculation*

Using EDSA 2000 software program one can get the flow calculation results from Fig. 4. Table I presents inter-phase voltage values of the test generator terminal, before and after loading condition. It is shown that under unbalanced loads condition, the phase angles of terminal generator voltage are deviated from its balanced value. The biggest deviation occurs when the grid operates under balanced load condition.

Table I. VALUES OF GENERATOR TERMINAL VOLTAGE

Conditions of Synchronous Generator	Phase	Tanjung Jati B Voltage [p.u]
Connected the grid and load balance	a	$1\angle -15^{o}$
	b	$1\angle 120^{o}$
	c	$1\angle 240^{o}$
Connected the grid and load imbalance of 5%	a	$1\angle 0.3^{o}$
	b	$1\angle 120^{o}$
	c	$1\angle 240^{o}$
Connected the grid and load imbalance of 10%	a	$1\angle -0.2^{o}$
	b	$1\angle 120^{o}$
	c	$1\angle 240^{o}$

B. Unbalanced Steady-Sate Simulation and Analysis

The ease of using the presented tool for the dynamic analysis of the generator has been demonstrated by examining the variations of dynamic parameters at different levels of unbalanced load. Its characteristics are presented in Table II.

This study is carried out utilizing the created GUI windows. Running the created GUI M-file, called "AwalGenMod", from Matlab workspace will display the main window shown in Fig. 5. The window presented in Fig. 6 will appear after clicking on the icon named "START". Setting the slider icons of generator's parameters and typical stator voltages and also clicking on the icon named "SIMULATE" will present the result menu shown in Fig. 6. This leads to the following figures which present the current and the voltage, generated real and reactive power, power angle, electric torque, field current and the current in phase.

Table II. GENERATOR RATINGS AND PARAMETERS

S_{rated}	= 820 MVA	V_{rated}	= 18 KV
N_{rated}	= 1800 rpm	Rated power factor	= 0.9
r_s	= 0.003 pu	x_{ls}	= 0.19 pu
x_d	= 1.70 pu	x_q	= 1.66 pu
x'_d	= 0.266 pu	x'_q	= 0.262 pu
x''_d	= 0.27 pu	x''_q	= 0.27 pu
T'_{d0}	= 7.8 s	T'_{q0}	= 0.41 s
T''_{d0}	= 0.033 s	T''_{q0}	= 0.055 s
H	= 3.77 s	D_{ω}	= 0 pu

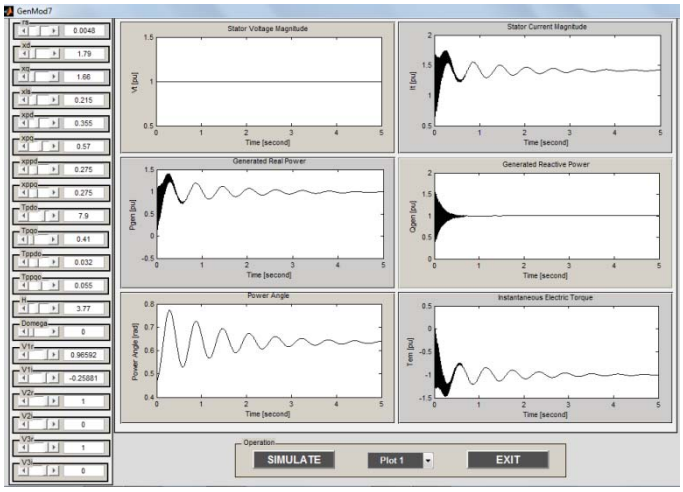


Fig. 8.a. Simulated load balance condition

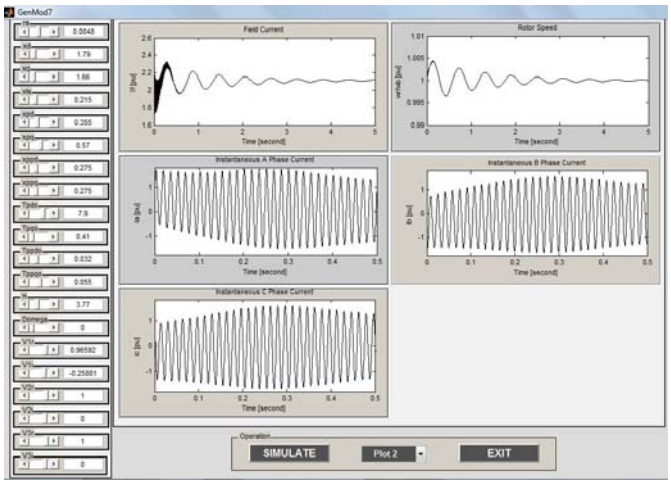


Fig. 8.b. Simulated load balance condition

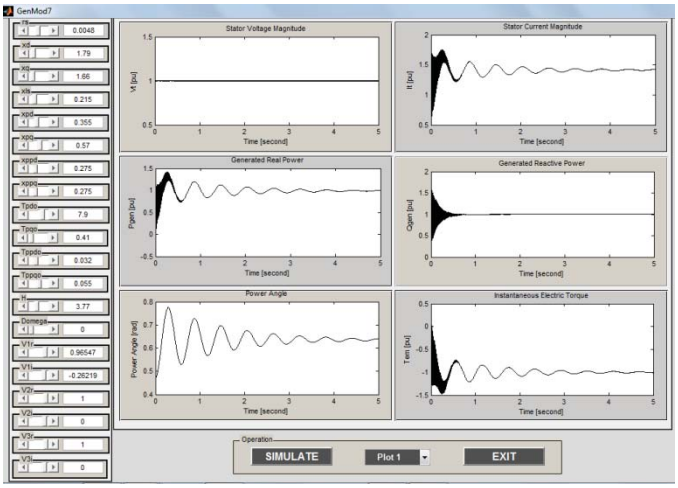


Fig. 9.a. Simulated load imbalance of 5% condition

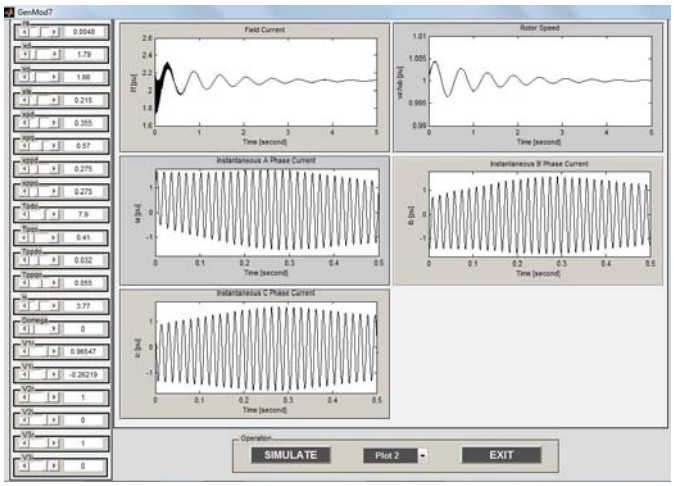


Fig. 9.b. Simulated load imbalance of 5% condition

As it can see on Fig. 8.a and 8.b, when the synchronous generator connects and delievers energy to the 500 kV EHV Jamali System as much as $P + jQ = (1 + j1)$ p.u, its current

in phase could become unbalanced even though all of the grid loads are balanced. Moreover, the generated powers, active and reactive, will become slightly to oscillate but the rotor speed tends to be constant. And the power angle peak goes up to 0.8 p.u.

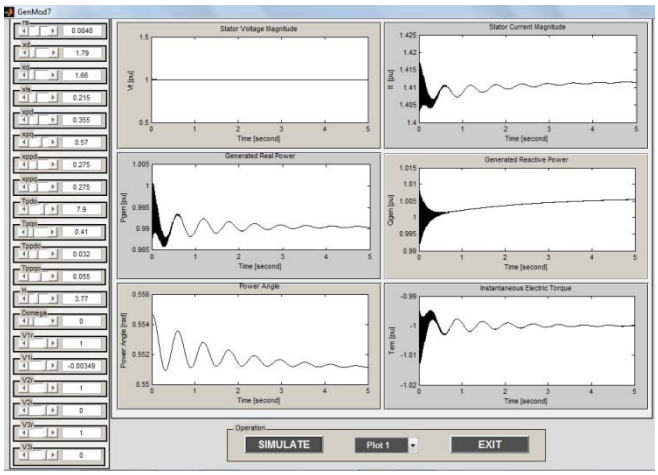


Fig. 10.a. Simulated load imbalance of 10% condition

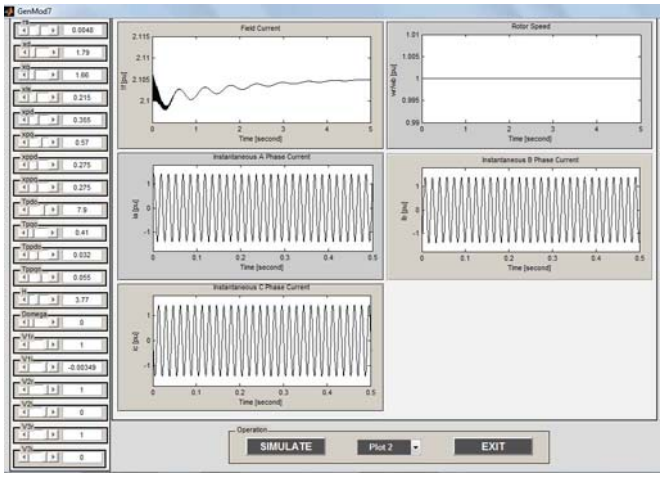


Fig. 10.b. Simulated load imbalance of 10% condition

As the level of unbalance is increasing up to 5%, the currents in phase are still unbalanced as shown in Fig. 9.b. But when the unbalanced portion is going up to 10% of the level, the currents in phase of the machine is almost balanced as shown in Fig. 9.b.

Comparing among Fig. 8.a, Fig. 9.a and Fig. 10.a, observably that the increasing of unbalanced percentage of load has an affect on the decreasing of the power angle and the magnitude of stator current oscillation, the generated active and reactive powers and the electric torque. On the contrary, it has no effect on the speed of the machine's rotor; the rotor speed is always constant.

IV. CONCLUSION

A useful approach for teaching a three-phase synchronous generator under unbalanced steady-state operation has been presented in this paper. Two operation conditions of the synchronous generator, load imbalance of 5% and 10%, are mathematically modeled then simulated using Matlab.

The simulation results state that the increasing level of unbalanced load has significant effect on the parameters of generator dynamic, except to the rotor speed. Surprisingly, the increasing level of unbalance on all of loads of 500 kV EHV Jamali system will change generator's phase currents to be in balance.

The developed tool is made easy to use by providing an active link with the simulated models using some of Matlab's GUI functions. The given examples demonstrate helpfulness of the developed tool for teaching the synchronous generator dynamic under unbalanced steady-state operation.

REFERENCES

[1] Y.K. Ching, B. Adkins, "Transient Theory of Synchronous Generators under Unbalanced Conditions', Institution Monograph, Supply Section, No. 85, 15th December, 1953.

[2] D.D. Robb, P.C. Krause, "Dynamic Simulation of Generator Fault Using Combined abc and 0dq Variables", IEEE Transaction on Power Apparatus and System, Vol. PAS-94, no. 6, November/December, 1975.

[3] J. van Amerongen, P.J. Buys, H.W.M. Barends, G. Honderd, "Modelling and Control a 180-MW Power System", Proceedings of 23rd Conference on Decision and Control, Las Vegas, NV, December, 1984.

[4] J.C.P. Suni, E. Ruppert, F. Fajoni, "A Guide for Synchronous Generator Parameters Determination Using Dynamic Simulations Based on IEEE Standards", XIX International Conference on Electrical Machine – ICEM, Rome, 2010.

[5] L. Yan, W. Dongmei, "Simulation of Synchronous Generator's Dynamic Operation Characteristics", The Tenth International Conference on Electronic Measurement & Instruments – ICEMI, 2011.

[6] Matlab® Ver. R2010B. Available: http://www.mathworks.com

[7] EDSA Technical 2000. Available : http://www.edsa.com

[8] P.C. Krause, O. Wasynczuk, S.D. Sudhoff, Analysis of Electric Machinery and Drive Systems, Willey-InterScience, A John Willey & Sons, Inc. Publishing, New York, 2002.

Thermal Unit Commitment Solution Using Genetic Algorithm Combined with The Principle of Tabu Search and Priority List Method

Sarjiya, Arief Budi Mulyawan, Apri Setiawan
Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada, Indonesia
sarjiya@ugm.ac.id, arief.abm@gmail.com, prisetiawan@gmail.com

Andi Sudiarso
Department of Mechanical and Industrial Engineering
Universitas Gadjah Mada, Yogyakarta, Indonesia
a.sudiarso@ugm.ac.id

***Abstract*— Unit commitment (UC) is one of optimization problem which is important in electrical power systems as effort to minimize generation cost by applying an effective scheduling. However, the size of search space and many constraints in this problem are becoming the problems. This paper will present hybrid algorithm which integrates genetic algorithm (GA) combined with the principle of tabu search (TS) and priority list (PL) methods to solve the UC problem. PL will be used for solving the unit scheduled problem. GA and the principle of TS are used for solving the economic dispatch problem. To optimize GA parameters, design of experiment (DOE) method will be used. The proposed algorithm is tested on the IEEE 10 unit systems for a one day scheduling periods. The results are compared with methodological priority list, shuffled frog leaping algorithm, hybrid particle swarm optimization, standard GA, integer coded GA, and Lagrange relaxation GA methods. This proposed hybrid method shows that the total cost of the unit commitment problem is better than other compared methods and near-optimal solution.**

Keywords—component; unit commitment; principle of tabu search; priority list; genetic algorithm; design of experiment

I. INTRODUCTION

Unit commitment (UC) is one of optimization problem which is important in electrical power systems. The main objective of this problem is to get an effective schedule so that it can minimize generation cost and fuel cost which is used. But, this main objective is restricted by some constraints [1]. In solving the UC problem, generally two main problems are involved: the unit scheduled problem and the economic dispatch problem. The unit scheduled problem involves to start up or shut down a given generating units during each hour while satisfying the constraints. The economic dispatch problem involves the allocation of the load demand and spinning reserve among the online generating units during each hour.

In order to obtain the minimum generating cost with UC in large system, large computation is needed, so that the analytic solution is impossible. In the developments are known various optimization methods based on programming and simulation using various optimization iterating methods on approach.

In the beginning, the best solution searching methods used to solve UC problem are conventional methods such as dynamic programming (DP) [2],[3], integer programming (IP) [3], and Lagrangian Relaxation (LR) [3]. Some of these methods give good results, such as LR, but the result is sub-optimal. While the other methods suffer the dimensionality problem like in DP.

In the development, heuristic methods are known such as expert systems (ES) [3] and priority list (PL) [3-5]. In heuristic methods a suboptimal solution may be obtained due to incomplete search of the solution space. Then, there are meta-heuristic methods such as tabu search (TS) [3],[6], artificial neural network (ANN) [3],[7], particle swarm optimization (PSO) [3][8-9], shuffled frog leaping algorithm (SFLA) [10], and genetic algorithm (GA) [3],[11-14]. The advantage of meta-heuristic methods is their effectiveness and it can be applied to different optimization problems with relatively few modifications to make them adapted to a specific problem. All these methods maintain a population of solutions that are evolved through random alterations and selection. There is no guarantee that the best solution found will be the optimal solution because of random evolved. But, meta-heuristic allowing the local search to escape from local optima so that the solution obtained using these methods is near global optimal [15].

In this paper, GA is developed for the solution of the UC problem. It uses the GA and the principle of TS combined with a PL method. PL will be used for solving the unit scheduled problem. GA is used for solving the economic dispatch problem. The GA parameters are optimized with design of experiment (DOE) method. DOE is one of accurate method for improving the quality and productivity [16]. As a result, the used GA parameters are the optimal combination. After GA is done, the principle of TS method will be used to improve the GA result.

II. UC PROBLEM FORMULATION

A. Notations

The notations used in this paper are:

S_{it} start up cost of i-th unit at t-th hour,

So_i	cool start up cost of i-th unit,
D_i, E_i	start up cost coefficient of i-th unit,
$F_{it}(P_{it})$	fuel cost function of i-th unit, with generation output, P_{it}, at the t-hour,
$TotalCost$	total cost of generation,
P_{it}	the generation output of the i-th unit at t-th hour,
U_{it}	the on/off status of the i-th unit at t-th hour, if online U_{it}=1, if offline U_{it}=0,
A_i, B_i, C_i	fuel cost coefficient cost of i-th unit,
$Toff_i$	total time of i-th unit during offline,
Ton_i	total time of i-th unit during online,
$Tdown_i$	minimum down time of i-th unit,
Tup_i	minimum up time of i-th unit,
D_t	load demand at t-th hour,
R_t	spinning reserve at t-th hour,
$Pmax_i$	maximum output power of i-th unit,
$Pmin_i$	minimum output power of i-th unit,
n	total generation,
t	time ,
X_0	average value from all running result,
X_1 - X_4	coefficient from 1st to 4th factor,
$F_1 - F_4$	1st to 4th factor with (+1, 0, or -1) value.

B. Objective Function

The objective function in this paper is to minimize total cost. Mathematically, it can be formulated as follows,

$$MinTotalCost = \sum_{t=1}^{T}\sum_{i=1}^{N}\left(U_{it}.F_{it}.\left(P_{it}\right) + U_{it}.S_{it}\right). \quad (1)$$

Total cost from (1) is specified as a sum of fuel cost and start-up cost from the generating units. Fuel cost equation can be defined as follows,

$$S_{it}\left(P_{it}\right) = A_i P_{it}^2 + B_i P_{it} + C_i. \quad (2)$$

Start-up cost equation can be defined as follows,

$$S_{it} = So_i\left(1 - D_i.\exp\left(-\frac{Toff_i}{Tdown_i}\right)\right) + E_i. \quad (3)$$

C. Constraints Subjected to UC problem

In UC problem, constraints are divided into two main constraints, that are system constraints and unit constraints.

1) System constraints

System constraints are constraints that affected to other variables in generation unit problem.

System constraints that used in this paper are power balance and spinning reserve constraints. The explanation about these constraints can be seen in below,

a) Power balance constraint

Power balance constraint is total power that are generated from online generating unit/s must be able to meet the demand at the specified time. The equation can be defined,

$$\sum_{i=1}^{N} U_{it}.P_{it} = D_t, 1 \le t \le T. \quad (4)$$

b) Spinning reserve constraint

Spinning reserve is the difference between total maximum power from all online generating units with total demand at the specified time. In this paper, spinning reserve is 10% from total demand. Generally spinning reserve constraint equation can be defined as follows,

$$\sum_{i=1}^{N} U_{it}.P\max_{it} = \left(D_t + R_t\right), 1 \le t \le T. \quad (5)$$

2) Unit constraints

Unit constraints are constraints applied in each generating unit as constraints that must be met before it's used to do the operation. Unit constraints that used in this UC problem are,

a) Power generation limit

Power that generated by each generating unit must be between minimum power and maximum power of each generating unit, and P value (power that generated) must be real values.

$$P\min_{it} \le P_{it} \le P\max_{it}, P \in R. \quad (6)$$

b) Minimum up and down time

Minimum up time is the minimum time when the generating unit had just turn on to go back in off mode. Meanwhile of minimum down time in UC is the minimum time when generating unit had just turn on to go back in online mode. From this definition, the minimum down time can be expressed in this equation,

$$\begin{aligned} &if\ D_t > D_{t-1}; if\ U_{it-1} = 0 \rightarrow U_{it} = 1 \\ &must\ fulfilled Toff_i \le Tdown_i \\ &1 \le i \le n; Toff_i < 0. \end{aligned} \quad (7)$$

For the minimum up time, the equation can be expressed as follows,

$$\begin{aligned} &if\ D_t \le D_{t-1}; if\ U_{it-1} = 1 \rightarrow U_{it} = 0 \\ &must\ fulfilled Ton_i \le Tup \\ &1 \le i \le n; Ton_i > 0. \end{aligned} \quad (8)$$

c) Initial status

Initial status is a precondition of generation schedule which contains information, which generation is on to serve the demand, and how long are the generations in on/off mode.

III. UC Problem Solution Methodology

In this paper, UC is solved by using GA and the principle of TS combined with PL methods. PL method is used to solve the unit scheduled problem. GA and the principle of TS are used together to solve the economic dispatch problem. The flowchart can be seen in Fig. 1.

A. Unit Scheduled Problem

To solve the unit scheduled problem, PL method is used. PL works in condition, if the load increases, then the generation with low operating costs is the priority to turn on. If the load decreases, then the generation with high operation costs in advance to turn off. In addition, when the load dropped only momentarily at a certain time, then 'hold on' status is done. Hold on status is status which works by hold the position of previous generating conditions so that it can be applied at the load time which the load is decreased momentarily. This status was done in anticipation of unpreparedness of generation when the load increases on the next time.

B. Economic Dispatch Problem

GA and the principle of TS are used to solve the economic dispatch problem. First, GA is executed. Then, after GA reaching the maximum generation, the principle of TS method is used. The detailed explanation can be seen below.

1) Genetic algorithm method

GA is an optimization method based on bioscience knowledges which adopted nature selection like crossover and mutation. In selection process using genetic algorithm, there are population, chromosome, generation, etc. Step from GA that will be used to solve economic dispatch problem are,

a) Population initialization

Population initialization is generation process step that generate population values randomly but all of them are satisfying the power balance constraint or above. This population consists of m chromosomes. The generated variables are consists of power generated from each generating unit in biner coding. Each variable is represented with 10 bits. Each of the generated chromosome is consists of M x N matrix. Where M value is 1 and N value is depends on total generating unit. The chromosome structure can be seen in Fig. 2.

In generating population step, if until n times repetition of generating chromosomes, the generated chromosomes still don't satisfying the power balance constraint or above, all of genes in the generated chromosomes are change into bit 1 so that it will satisfying the power balance constraint. The number of repetition in this paper is 100 times.

b) Chromosome decode

The function of chromosome decode is to decode the value from an existing chromosome to the value that will be used as input in objective function. The upper and lower limit from the generating unit will be included in this decoding so that it can't violate the power generation limit.

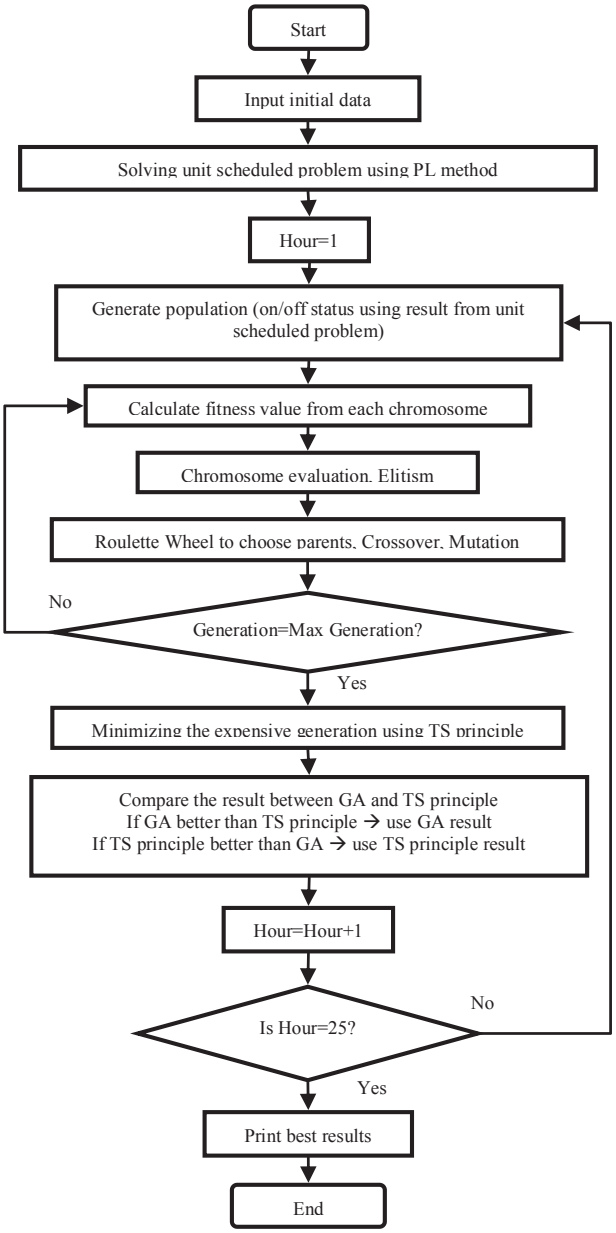


Fig. 1. Flowchart of the proposed method.

c) Fitness function

Fitness function divided into two types, minimization function and maximization function.

If the problem is maximization, then,

$$fitnessfunction = objectivefunction \quad (9)$$

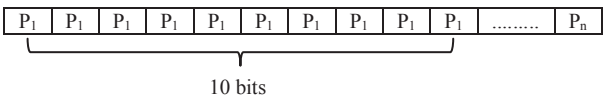


Fig. 2. Chromosome structure.

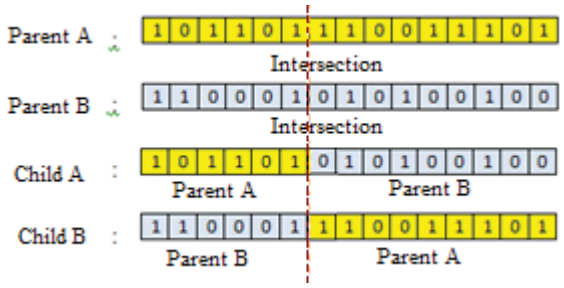


Fig. 3. Cross over illustration.

If the problem is minimization, then,

$$fitnessfunction = 1/(objectivefunction + a). \quad (10)$$

Where '*a*' is a small number used to anticipating division zero number.

d) Chromosome evaluation

Chromosome evaluation is a process where the chromosome from the generated population will be selected to get the best chromosome which will be used to represents the solution.

e) Elitism

Elitism is a duplicate process of the best chromosome as much as *N* copies. This step is done to keep the best chromosome so that the best chromosome will exist in the next generation without undergoing crossover and mutation.

f) Linear fitness ranking

Linear fitness ranking (LFR) is fitness value scaling method. It is used to form upper bound of each fitness values so that it can be used to select the parents.

g) Roulette wheel

Roulette Wheel is a method to select the parents in GA process. The principle of this method is like spinning the roulette wheel. It will stop at certain LFR value and chromosome which has a LFR value appropriate to stopping roulette wheel place would be chosen to become parents in the process of GA.

h) Cross over

Crossover is a process where new chromosome resulting from the exchange of genes from both parents who was selected so that the child or the resulting offspring will inherit the parent nature or value will be born. Probability to crossover is depent on crossover probability value. Cross over illustration can be seen in Fig.3.

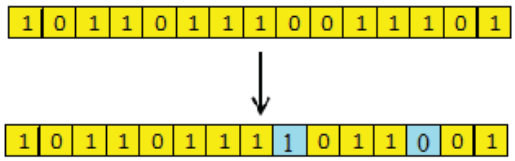


Fig. 4. Mutation illustration

i) Mutation

Mutation is a process to change a gene from 1 to 0 or from 0 to 1. Probability to change is depent on mutation probability value. The illustration of this process can be seen in Fig.4.

2) The principle of tabu search

Tabu search is an iterative improvement procedure that starts from some initial feasible solution and attempts to determine a better solution in the manner of a greatest-descent algorithm. Tabu search permits backtracking to previous solutions, which may ultimately lead, via a different direction, to better solutions [6]. This principle method will be used to obtain a better solution from the economic dispatch problem.

Actually, GA is smart enough to know which generation is the cheaper, but sometimes it doesn't maximize the potential of the cheaper generation. For the example, the maximum power generating from the cheaper generation is 85 MW, but the result from GA is 84.7 MW. This principle will maximize the generation potential to that maximum power which is 85 MW. So, the principle of tabu search method is used to maximize the power generated by the cheaper generation and to minimize the power generated by the expensive generation. The expensive generation is pushed to generate at the minimum power of that generation so that the cost will be decreased.

C. Design of Experiment (DOE) Formulation

DOE is one of an accurate method for improving the quality and productivity. Generally, experiment is done to find out the relationship condition between the affected of one variable to other variable in effect of increasing output [16].

From the exposure of experimental types, fractional factorial is adopted by DOE. This type is trying to combine half possibility from all the possibility variables that exist structurally. Dr. Taguchi has developed that coding structure and is known as 'Taguchi's table'. The combinations according to the DOE 4 factors 3 level of 'Taguchi's table' in Table I is used in this paper.

The advantages of doing experiment using DOE method is besides it can reduces the burden of trying all the possibilities that exist, it can be used to determine the prediction function too, so the best combination to obtain the best output values can be found out with this prediction function. The prediction function can be approximated with (11).

$$result = X_0 + X_1.F_1 + X_2.F_2 + X_3.F_3 + X_4.F_4, \quad (11)$$

TABLE I. DOE 4 FACTORS 3 LEVEL

		Variables				Average Results
		1	2	3	4	
Total Experiments	1	-1	-1	-1	-1	A
	2	-1	0	0	0	B
	3	-1	1	1	1	C
	4	0	-1	0	1	D
	5	0	0	1	-1	E
	6	0	1	-1	0	F
	7	1	-1	1	0	G
	8	1	0	-1	1	H
	9	1	1	0	-1	I

TABLE II. DOE MODELLING

Population Size			Crossover Prob.		
Min	Mid	Max	Min	Mid	Max
30	40	50	0.5	0.7	0.9
Symbol of DOE			Symbol of DOE		
-1	0	1	-1	0	1
Mutation Prob.			**Max. Generation**		
Min	Mid	Max	Min	Mid	Max
0.04	0.08	0.12	300	500	700
Symbol of DOE			Symbol of DOE		
-1	0	1	-1	0	1

$$X_0 = \frac{\sum_{i=1}^{9} averagecost}{N_{\exp eriments}}, \quad (12)$$

$$X_1 = \sum_{i=1}^{9} \left(symb.1st\,variable \times averagecost_i \right), \quad (13)$$

$$X_2 = \sum_{i=1}^{9} \left(symb.2nd\,variable \times averagecost_i \right), \quad (14)$$

$$X_3 = \sum_{i=1}^{9} \left(symb.3rd\,variable \times averagecost_i \right), \quad (15)$$

$$X_4 = \sum_{i=1}^{9} \left(symb.4th\,variable \times averagecost_i \right). \quad (16)$$

IV. SIMULATION AND RESULTS

This hybrid algorithm is tested in IEEE 10 unit system standard [9]. After unit scheduled problem is solved by PL method, the next step is to solve economic dispatch problem using GA. But in order to perform at the optimum combination, firstly, optimizing GA parameters is done by using DOE method. DOE model for this test can be seen in Table II.

After the DOE model is built, the combinations according to the DOE 4 factors 3 level of 'Taguchi's table' in Table I must be tested. The results from testing the combination can be seen in Table III.

The prediction equation can be obtained using the defined variables and average cost from Table III and using (11) – (16). The prediction equation is shown in (17).

$$result = 563339.98 - 36.56X_1 - 62.2701X_2 - 28.8006X_3 - 0.90236X_4. \quad (17)$$

TABLE III. DOE TESTING RESULTS

No.	Population Size	Max. Generation	Cross over Probability	Mutation Probability	Trial-1 Cost ($)	Trial-2 Cost ($)	Trial-3 Cost ($)	Average Cost ($)
1	-1	-1	-1	-1	563,729.38	563,376.51	563,674.88	563,593.59
2	-1	0	0	0	563,506.38	563,430.07	563,402.30	563,446.25
3	-1	1	1	1	563,081.31	563,385.72	563,485.99	563,317.67
4	0	-1	0	1	563,184.34	563,338.53	563,173.84	563,232.24
5	0	0	1	-1	563,164.87	563,292.57	563,298.56	563,252.00
6	0	1	-1	0	563,116.05	563,254.43	563,198.43	563,189.64
7	1	-1	1	0	563,623.34	563,076.28	563,525.10	563,408.24
8	1	0	-1	1	563,136.09	563,646.72	563,578.87	563,453.89
9	1	1	0	-1	563,063.02	563,216.92	563,219.06	563,166.33

TABLE IV. VALIDATION RESULTS WITH THE OPTIMUM GA PARAMETERS

Population Size	Max. Generation	Cross over Probability	Mutation Probability	Trial-1 Cost ($)	Trial-2 Cost ($)	Trial-3 Cost ($)	Average Cost ($)
50	700	0.9	0.12	562,979	563,082	563,147	563,069

Equation (17) indicated that the optimum GA parameters value is in [1 1 1 1] condition which means the population size = 50, crossover probability = 0.9, mutation probability = 0.12, and maximum generation = 700 with the total cost $563,211. Then, from this result, testing is done to validate the result. The validation is done three times and the best results from those validations is became the solution. From Table IV, the best results obtained in the first validation test with the total cost $562,979. The dispatched power of units in each time interval can be seen in Table V.

The best obtained result is then compared with other methods i.e. methodological priority list (MPL) [5], shuffled frog leaping algorithm (SFLA) [10], hybrid particle swarm optimization (HPSO) [9], standard genetic algorithm (SGA) [11], integer-coded genetic algorithm (ICGA) [14], and the combination of lagrangian relaxation and genetic algorithm (LRGA) [17]. From Table VI, it can be seen that the production cost of this hybrid algorithm is less than the other compared methods.

V. CONCLUSION

In this paper, a hybrid algorithm for the solution to the unit commitment problem was presented. This hybrid algorithm uses the combination of GA, the principle of TS combined and a PL method. PL will be used for solving the unit scheduled problem. GA and the principle of TS are used for solving the economic dispatch problem. To optimize the GA parameters, DOE method is used.

TABLE V. DISPATCH OF UNITS DURING THE OPERATING HOURS

Hour	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Unit 6	Unit 7	Unit 8	Unit 9	Unit 10
1	455.0	245.0	0	0	0	0	0	0	0	0
2	455.0	295.0	0	0	0	0	0	0	0	0
3	455.0	370.0	0	0	25.0	0	0	0	0	0
4	455.0	455.0	0	0	40.0	0	0	0	0	0
5	445.8	417.1	0	111.8	25.4	0	0	0	0	0
6	455.0	377.2	104.4	128.5	35.0	0	0	0	0	0
7	455.0	455.0	85.0	130.0	25.0	0	0	0	0	0
8	455.0	455.0	130.0	130.0	30.0	0	0	0	0	0
9	455.0	455.0	130.0	130.0	85.0	20.0	25.0	0	0	0
10	455.0	455.0	130.0	130.0	162.0	33.0	25.0	10.0	0	0
11	455.0	455.0	130.0	130.0	162.0	73.0	25.0	10.0	10.0	0
12	455.0	455.0	130.0	130.0	162.0	80.0	25.0	43.0	10.0	10.0
13	455.0	455.0	130.0	130.0	162.0	33.0	25.0	10.0	0	0
14	455.0	455.0	130.0	130.0	85.0	20.0	25.0	0	0	0
15	455.0	455.0	130.0	130.0	30.0	0	0	0	0	0
16	455.0	310.0	130.0	130.0	25.0	0	0	0	0	0
17	455.0	260.0	130.0	130.0	25.0	0	0	0	0	0
18	453.5	396.6	120.3	103.5	26.1	0	0	0	0	0
19	455.0	455.0	130.0	130.0	30.0	0	0	0	0	0
20	455.0	455.0	130.0	130.0	162.0	33.0	25.0	10.0	0	0
21	455.0	455.0	130.0	130.0	85.0	20.0	25.0	0	0	0
22	455.0	455.0	0	0	145.0	20.0	25.0	0	0	0
23	455.0	420.0	0	0	25.0	0	0	0	0	0
24	455.0	345.0	0	0	0	0	0	0	0	0

TABLE VI. COMPARISON OF FINAL PRODUCTION COST FOR 10 UNIT SYSTEM AMONG OTHER METHODS

Methods	Total Cost ($)
Proposed method	562,979
MPL	563,977
SFLA	564,769
HPSO	563,942
SGA	565,825
ICGA	566,404
LRGA	564,800

The performance of the proposed method was tested for a one day thermal unit commitment problem. The simulation result shows that the proposed method give better result than the other compared methods.

REFERENCES

[1] A.J. Wood & B.F. Wollenberg, "Power Generation, Operation, and Control," 1st ed. New York : Willey, 1984.

[2] Z. Ouyang and S. M. Shahidehpour, "An intelligent dynamic programming for unit commitment application," IEEE Transactions on Power Systems, vol. 6, no. 3, pp. 1203-1209, Aug. 1991.

[3] N. P. Padhy, "Unit Commitment-A bibliographical survey," IEEE Trans. Power syst., vol. 19, no. 2, pp. 1196-1205, May 2004.

[4] T. Senjyu, K. Shimabukuro, K. Uezato and T. Funabashi, "A Fast Technique for Unit Commitment Problem by Extended Priority List," IEEE Trans. Power Syst., vol. 18, no. 2, pp. 882–888, May 2003.

[5] Yang Tingfang, T.O. Ting, "Methodological Priority List for Unit Commitment Problem," International Conference on Computer Science and Software Engineering, vol.1, pp. 176-179, Dec 2008.

[6] A.H. Mantawy, Y.L. Abdel-Magid, S.Z. Selim, "Unit commitment by tabu search," IEE Proceedings-Generation, Transmission and Distribution, vol.145, pp. 56-64,Jan 1998.

[7] H. Sasaki, M. Watanabe, J. Kubokawa, N. Yorina, and R. Yokoyama,"A solution method of unit commitment by artificial neural network," in Proc. IEEE Power Eng. Soc. Summer Meeting, 1991.

[8] Zwe-Lee Gaing. Discrete particle swarm optimization algorithm for unit commitment. In 2003 IEEE/PES General Meeting Vol 1,pp13-17.July,2003.

[9] T. O. Ting, M. V. C., Rao, C. K. Loo, "A novel approach for unit commitment problem via an effective hybrid particle swarm optimization,". IEEE Trans. power syst., vol. 21, no. 1, pp. 411-418, 2006.

[10] Javad Ebrahimi, Seyed Hossein Hosseinain, Gevorg B. Harehpatian, "Unit Commitment Problem Solution Using Shuffled Frog Leaping Algorithm," IEEE Trans. Power Systems, vol.26, no.2, pp.573-581,May 2011.

[11] S A. Kazarlis, A. G. Bakirtzis and V. Petridis, "A Genetic Algorithm Solution to The Unit Commitment Problem," IEEE Trans. on PowerSystems, Vol. 11, No. 1, pp.83-92, Feb. 1996.

[12] G. B. Sheble et al., "Unit commitment by genetic algorithm with penalty methods and a comparison of Lagrangian search and genetic algorithm economic dispatch example," Int. J. Elect. Power Energy Syst., vol. 18,no. 6, pp. 339–346, Feb. 1996.

[13] K. S. Swarup and S. Yamashiro, "Unit commitment solution methodology using genetic algorithm," IEEE Trans. Power Syst., vol. 17, pp.87–91, Feb. 2002.

[14] Ioannis G. Damousis et al., "A Solution to the Unit-Commitment Problem Using Integer-Coded Genetic Algorithm," IEEE Trans.Power Syst.,vol.19,pp.1165-1171,May.2004.

[15] Christian Blum,Andrea Roli,"Metaheuristics in Combinatorial Optimization:Overview and Conceptual Comparison," ACM Computing Surveys (CSUR), vol.35, pp. 263-308,September.2003.

[16] A.Sudiarso, "Statistic for Industrial Engineer 2," unpublished.

[17] Chuan-Ping Cheng , Chih-Wen Liu, Chun-Chang Liu, "Unit commitment by lagrangian relaxation and genetic algorithms," IEEE Transactions on Power Systems,vol.15,pp.707-714,May.2000

Wind Speed Calculation by Using Electrical Output and Wind Turbine Power Curve

Agus Purwadi, M. Ikhsan, Nanang Hariyanto, Nana Heryana, Yanuarsyah Haroen
School of Electrical Engineering and Informatics,
Bandung Institute of Technology
Bandung 40132, Indonesia
agusp@konversi.ee.itb.ac.id, m.ikhsan.yusuf@gmail.com

***Abstract*—Measurements of wind speed usually done by using anemometer, these data can be used as a reference signal to the wind turbine control, or it can be used also for mapping the wind energy potential. Recently, the use of anemometer on wind turbines are often eliminated because of technical and cost reasons. This paper describes an alternative method to determine the upstream wind at the wind turbine without using the anemometer, but using only the electrical output (current and voltage) and the power curve data provided by the wind turbine manufacture. Further, the power curve is mathematically modeled by polynomial interpolation method so that the wind speed equation as a function of power can be determined.**

Keywords—Wind turbine; data logger; power curve, lagrange polynomial interpolation.

I. INTRODUCTION

Wind speed measurements using anemometer is the most accurate way to know the wind speed at a certain place. Anemometer is also often used to support control algorithms in optimizing electrical energy produced at wind power, such as has been done in [1] and [2] for the maximum power point tracking algorithms, the wind speed is compared to the tip speed ratio, then the error signal is used in the control process, which change the turbine speed to minimize the error.

Basically, the use of anemometer in wind turbine systems will increase the component and maintenance costs of the system. In addition, the anemometer can degrade system reliability because of the aerodynamic interference, shadow effects, and the turbulence caused by wind turbines downstream wind [3]. Besides using the anemometer, the wind speed data can also be obtained by using a small-scale test turbine, the method is then categorized into measurement by indirect methods.

The wind speed indirect measurement method had previously been investigated as in [4] and [5]. The authors use the output frequency of the generator to calculate the wind speed. This can be done because it will be proportional to the wind speed rotor angular velocity, and the angular speed of the rotor will be proportional to the alternator's frequency.

However, there are disadvantages when the frequency is used as a basis of wind speed indirect measurements, when the system is generating a harmonic waves, frequency readout is becomes difficult and complex [6]. It has been confirmed that the static power converters such as rectifiers, inverters, choppers, etc. are the main sources of harmonics on power systems [7]. Analytical methods need to be applied so that the fundamental frequency can be read accurately. However, it needs to be processed on a computing machine (CPU or DSP) which is relatively more expensive than the price of anemometer. Calculation of the wind speed proposed in this paper is using the wind turbine electrical output and power curves which are mathematically modeled by Lagrange polynomial interpolation method, thus the wind velocity equation as a function of power (electrical output) can be determined.

II. WIND TURBINE POWER CURVE

Wind turbine power curve is the expression between wind speed and power output for each specific wind speed. Power curve is typically determined from measurements in the field, anemometer mounted near to the wind turbines at the appropriate places (not disturbed by the turbulence of the rotor). At the same time, the power output of wind turbines were also measured or recorded. This can be done by using a data logger, an electronic device that records data over time or in relation to location with a built in sensor [8]. If these values are plotted, the wind turbine power curve will be obtained.

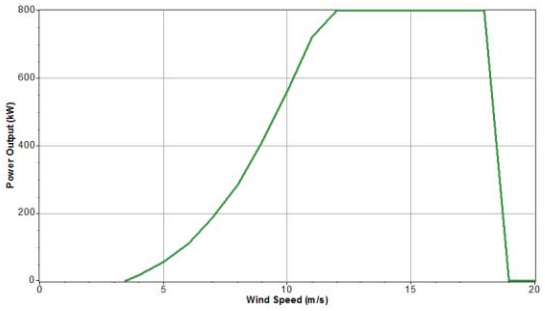


Fig 1. Gamesa MADE AE-59/800kW power curve

Fig. 1 is the power curve for Gamesa wind turbine model MADE AE-59/800kW.

However currently, many of the wind turbine manufacturers make up the power curve which is not based on field experiments, but only by using the coefficient of performance prediction [9].

III. POLYNOMIAL INTERPOLATION

Interpolation is a method to construct a point/new data from a discrete data set known [10]. Interpolation is also commonly used for data analysis, design of industrial products, as well as signal processing (digital-analog conversion) [11]. If it meets a discrete point in the form of a polynomial function, then a work that interpolates data points to a polynomial function is called polynomial interpolation [12]. Polynomial interpolation often provides a solution to a mathematical problem in engineering sciences. Polynomial interpolation technique is one of the tools that can be used to process the data (in the form of functions, curves, processes, etc.), and then development of the data in a wider domain. There are many polynomial interpolation method commonly used in numerical analysis, such as polynomial Lagrange. This formula was discovered by Edward Waring in 1779, but was first published by Joseph Louis Lagrange in 1795.

If it is assumed there is a set of data points (x_i , y_i), where i=0,1,2,…,n which is obtained from the function $f(x)$, so that $y_i = f(x_i)$, i=0,1,2,…,n. the appropriate interpolation of $I(x)$, can be expressed as [13],

$$I(x) = \sum_{i=0}^{n} L_i(x) \cdot f(x_i) \tag{1}$$

$$= L_0(x) \cdot f(x_0) + L_1(x) \cdot f(x_1) + \ldots + L\,(x) \cdot f(x_n)$$

The $L_i(x)$, i=0,1,2,…,n function is chosen so that it satisfy,

$$L_i(x) = \begin{cases} 0 & x = x_0, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n \\ 1 & x = x_i \end{cases} \tag{2}$$

Before defining the function $L_i(x)$, should be understood first the implications of (1) and (2). The best way to solve this equation is to select a value for "n" and write the equation results. Suppose there are four data points [x_i , $f(x_i)$], i=0,1,2,3. from (1) with n=3 so the interpolation function $I(x)$ become,

$$I(x) = \sum_{i=0}^{3} L_i(x) \cdot f(x_i) \tag{3}$$

$$= L_0(x) \cdot f(x_0) + L_1(x) \cdot f(x_1) + L_2(x) \cdot f(x_2) + L_3(x) \cdot f(x_3)$$

$I(x)$ became identical with $f(x)$ when x is one of the four pieces of data. If $I(x)$ evaluated at x_0, x_1, x_2, and x_3, then,

$$I(x_0) = L_0(x_0) \cdot f(x_0) + \ldots + L_3(x_0) \cdot f(x_3) \tag{4.a}$$

$$I(x_1) = L_0(x_1) \cdot f(x_0) + \ldots + L_3(x_1) \cdot f(x_3) \tag{4.b}$$

$$I(x_2) = L_0(x_2) \cdot f(x_0) + \ldots + L_3(x_2) \cdot f(x_3) \tag{4.c}$$

$$I(x_3) = L_0(x_3) \cdot f(x_0) + \ldots + L_3(x_3) \cdot f(x_3) \tag{4.d}$$

based on the (2), $L_0(x_0)$ = 1 and $L_1(x_0) = L_2(x_0) = L_3(x_0)$ = 0 , (4.a) through (4.d) can be simplified to,

$$\begin{aligned} I(x_0) &= L_0(x_0) \cdot f(x_0) + \ldots + L_3(x_0) \cdot f(x_3) \\ &= 1 \cdot f(x_0) + 0 \cdot f(x_1) + 0 \cdot f(x_2) + 0 \cdot f(x_3) \\ &= f(x_0) \end{aligned} \tag{5}$$

$I(x_1) = f(x_1), I(x_2) = f(x_2)$, so that means the interpolation function, $I(x)$ is through the data points are held. The analytical Form $I(x)$ will depend with $L_i(x), i = 1,2,\ldots,n$ function which satisfies (2). The function then called a Lagrange polynomial coefficient and defined as follows:

$$i = 0,\ L_0(x) = \frac{(x - x_1)(x - x_2)(x - x_3)}{(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)} \tag{6.a}$$

$$i = 1,\ \ L_1(x) = \frac{(x - x_0)(x - x_2)(x - x_3)}{(x_1 - x_0)(x_1 - x_2)(x_1 - x_3)} \tag{6.b}$$

$$i = 2,\ L_2(x) = \frac{(x - x_0)(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_2)(x_2 - x_3)} \tag{6.c}$$

$$i = 3,\ L_3(x) = \frac{(x - x_0)(x - x_1)(x - x_2)}{(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)} \tag{6.d}$$

Note that each Lagrange coefficient in the above equation is a third order polynomial. For the general case where there are N +1 data points, the Lagrange polynomial coefficient $L_i(x)$ at (1) s a polynomial of n^{th} order, as well as the interpolation function $I(x)$ $I(x)$ Thus the polynomial interpolation $I(x)$ on (1) can be represented as $f_n(x)$ and expressed as follows,

$$f_n(x) = f_0(x) L_0(x) + \ldots + f_3(x) L_3(x) \tag{7}$$

IV. GENERATING THE WIND TURBINE POWER CURVE

The wind turbines used in these paper is Gamesa MADE AE-59/800kW which have a synchronous generator with independent excitation and brushless. Nine pairs of data contained in the product datasheet (Table 1) is used as in (6.a-d) and (7) for the Lagrange polynomial interpolation. By using MATLAB the interpolation results can be seen as Fig.2.

TABLE 1. THE DATA OF GAMESA MADE AE-59/800KW OUTPUT POWER VERSUS WIND SPEED FROM MANUFACTURE [14]

Wind Speed (m/s)	Output Power (kW)
3.5	0
4	17.3
5	55.2
6	109.9
7	186.7
8	284.9
9	409.6
10	560.3
11	722.3
12	800
13	800

It appears that the line in Fig.2, which is the results of the interpolation polynomial, is coincides with all of the nine round dots of data from the wind turbine datasheet. The results of polynomial interpolation as in Fig.2 gives a meaning that there will be a specific wind speed for each electrical power output.

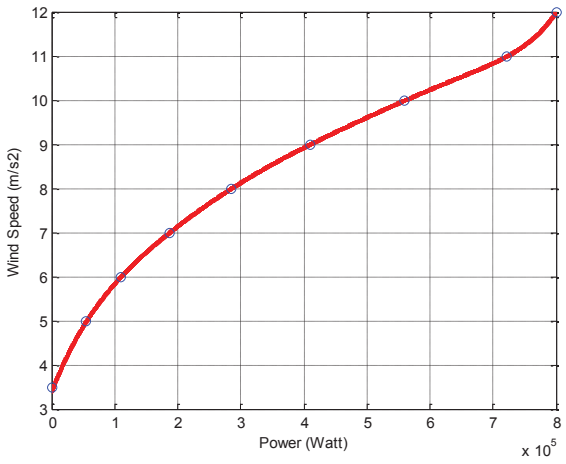


Fig 2. The results of Lagrange polynomial interpolation

V. EXPERIMENT

The experiment is done by using field data in reference [15] as shown in table 2. The data is taken every 1 hour in one day on the Balearic Island which has four units of Gamesa wind turbine model MADE AE-59/800kW as shown in Fig. 3. The wind speed measurements performed by using the anemometer.

Fig 3. Es Milà Wind Farm at Balearic Islands, Spain

The goal of the experiment is to compare the results of wind speed measurement readings between anemometer measurement (direct method) and calculation method (indirect method).

TABLE 2. THE MEASUREMENT DATA OF MADE AE-59/800KW OUTPUT POWER VERSUS WIND SPEED

Data No.	Electric Power (kW)	Actual (m/s^2)	Indirect (m/s^2)	Error (%)
1	0	1.85	3.41	45.71
2	0	2.27	3.41	33.39
3	0	3.17	3.41	6.98
4	15	4.07	3.55	14.81
5	502	4.85	6.23	22.14
6	226	3.40	5.03	32.39
7	131	3.97	4.45	10.79
8	687	9.50	6.83	39.18
9	2742	12.17	10.75	13.22
10	2943	13.12	11.11	18.08
11	2875	14.38	10.97	31.05
12	2986	11.32	11.21	0.96
13	2862	11.80	10.95	7.78
14	2856	12.63	10.94	15.47
15	2600	11.42	10.54	8.37
16	2108	10.70	9.79	9.31
17	2157	11.03	9.87	11.79
18	2731	11.05	10.73	2.97
19	2674	9.93	10.65	6.73
20	2482	9.62	10.37	7.19
21	2134	8.00	9.83	18.62
22	1134	5.50	7.99	31.14
23	357	6.78	5.67	19.68
24	892	6.95	7.40	6.03

VI. RESULT AND DISCUSSION

The comparison between direct measurement and calculation result using the proposed indirect method (table 2) can be seen as in the Fig.4.

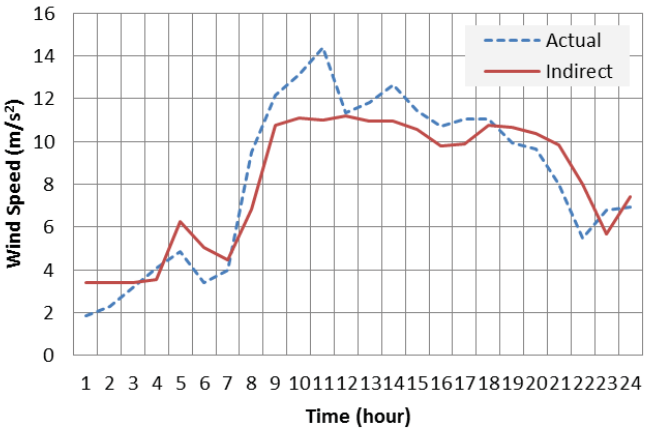


Fig 4. Measurement comparison between direct and indirect measurement

Based on table 2, the errors that occur between two measurement methods is still relatively large, e.g. in the data

number 1. There are several things that can cause a large error, such as:

- The indirect measurement method cannot calculate the wind speed when there is no power produced by the wind turbine. This occurs for instance in the data number from 1 to 3.
- The wind energy is very dependent on the contours of the surrounding environment. Within a radius of a few kilometers, it can have a significant difference [16]. The power measured in Table 2 is the sum of the four wind turbines power on large wind farms, while the wind speed measurement is only performed using a single anemometer at a single particular point around the wind turbines. Consequently, four of the wind turbines and the anemometer have a probability to get a different wind speed which impacted on a large error.

VII. CONCLUSION

This paper describes a method to calculate the wind speed of the wind turbine without using anemometer, but using only the electrical output (current and voltage) from the generator and the mathematical model of wind turbine power curves which can be approximated by polynomial interpolation method. There are several advantages when polynomial interpolation is used to get the equation of the power curve, namely:

- The wind turbines manufactures always gives the datasheet equipped with a power curve, it means: for each particular wind speed the power to be generated will be known, thus polynomial interpolation will be possible to apply
- Coefficient of wind turbine design is not required: the data regarding the design coefficient was never mentioned in the wind turbine datasheet. Using polynomial interpolation, the use of coefficient data can be avoided.
- Accurate: the result of the mathematical approach will be very close to the power curve provided by the wind turbine manufacture.
- Flexible: can be applied to all types of wind turbines without need to consider the mathematical approach of generators that is used.

There are several work still to be done in the future, such as, comparing the indirect measurements result from a single wind turbine with a near-placed-single anemometer.

Beside Lagrange, there are many polynomial interpolation methods that have been developed. In the future, these methods needs to be compared so that the mathematical approaches of the wind turbine power curve obtained has the smallest error value and a simpler calculation.

ACKNOWLEDGMENT

This research was funded by Director General of Higher Education Republic of Indonesia through DIKTI decentralization program year 2013.

REFERENCES

[1] Mutoh, N., Nagasawa, A., 2006. A maximum power point tracking control method suitable for compact wind power generators, PESC Record - IEEE Annual Power Electronics Specialists Conference, art. no. 1712214

[2] Patel, Mukund R. 1999. Constant Tip-Speed Ratio Scheme, in "*Wind and Solar Power Systems*". CRC Press, Florida, p. 101

[3] Rocha, R., 2011. A sensorless control for a variable speed wind turbine operating at partial load, Elsavier Renewable Energy, 36 (1), pp. 132-141.

[4] Budisan, N., Groza, V., Prostean, O., Filip, I., Biriescu, M., Szeidert, I., Stern, M., 2008. Rotation speed and wind speed indirect measurement methods for the control of windmills with fixed blades turbine, Conference Record - IEEE Instrumentation and Measurement Technology Conference, art. no. 4547166, pp. 912-916

[5] De Broe, A.M., Drouilhet, S., Gevorgian, V., 1999. A peak power tracker for small wind turbines in battery charging applications, IEEE Transactions on Energy Conversion, 14 (4), pp. 1630-1635

[6] Begovic, Miroslav M., Djuric, Peter M., Dunlap, Sean, Phadke, Arun G., 1993. Frequency tracking in power networks in the presence of harmonics, IEEE Transactions on Power Delivery, 8 (2), pp. 480-486.

[7] Chanedeau, G., Gaubert, J.-P., Rambault, L., Tissier, J.-F., 2012. Frequency measurement with high accuracy and method for harmonics identification in polluted distribution power networks, 2012 IEEE International Conference on Industrial Technology, ICIT 2012, Proceedings, art. no. 6209983, pp. 474-479.

[8] Purwadi, A., Haroen, Y., Farianza Yahya Ali, Heryana, N., Nurafiat, D., Assegaf, A., 2011. Prototype development of a Low Cost data logger for PV based LED Street Lighting System, Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, ICEEI 2011, art. no. 6021693,

[9] Ani, S.O., Polinder, H., Ferreira, J.A., 2013. Comparison of energy yield of small wind turbines in low wind speed areas, IEEE Transactions on Sustainable Energy, 4 (1), art. no. 6216474, pp. 42-49.

[10] Sadkhan, Sattar B., Ruma, K.H., 2006. "Evaluation of Polynomial Reconstruction Problem Using Lagrange Interpolation Method," Information and Communication Technologies - 2006. ICTTA '06. 2nd, p. 1399 - 1403

[11] MIT OpenCourseWare, "Introduction to Numerical Analysis, Chapter 3*:* Interpolation", Spring 2012, http://ocw.mit.edu/ courses/ mathematics /18-330-introduction-to-numerical-analysis-spring-2012/lecture-notes/MIT18_330S12_Chapter3.pdf, 18 January 2013

[12] Munir, R., "Informatics Special Course I IF4058: Polynomial Interpolation" http://informatika.stei.itb.ac.id/ ~rinaldi.munir/ MetNum/ 2010-2011/Interpolasi%20Polinom.pdf. 15 November 2012.

[13] University of Central Florida, "Section 4: Lagrange Interpolating Polynomials", http:// pegasus.cc.ucf.edu/ ~klee/ EGN3420/ Notes/ Interp_Lagrange.pdf. 20 January 2013.

[14] Gamesa MADE AE-59/800 kW Datasheet

[15] Montoya Moyá, J. 2009: Study of data of a wind farm. (Student paper). Linköpings universitet.

[16] Chiras, Dan. 2006: The Homeowner's Guide to Renewable Energy, Canada, New Society Publisher, 215.

Session 4

Electronics, Circuits, and Systems

Another Approach to Design a Sinusoidal Oscillator: Ensuring the Oscillation Stability

Dzuhri Radityo Utomo
Department of Electrical Engineering
and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
Email: dzuhriru@gmail.com

***Abstract*—Another approach to design a sinusoidal oscillator is proposed. This design is proposed to improve the oscillation stability of sinusoidal oscillator. The ideal model of this oscillator does not have any specific oscillation criterion. This property ensures the oscillation stability of this oscillator. Any changes in oscillation frequency does not violate the oscillation stability. This oscillator design can be implemented using op-amps and passive components. But the characteristics of op-amps influence the oscillator, thus change the characteristics of this oscillator. So a more realistic model is needed to predict the behaviour of this oscillator circuit more accurately. The realistic model is built by considering the characteristics of op-amps. Based on this model, a gain adjustment is needed by this circuit. Without any proper adjustment, a stable oscillation does not occur in this circuit. The realistic model is also used to predict the oscillation frequency of this oscillator and it produces a result that is similar to the simulation result.**

Keywords*—Sinusoidal Oscillators, Oscillation Criterion, Oscillation Stability, Variable Frequency, Gain Adjustment.

I. Introduction

Sinusoidal oscillator (oscillator) has numerous applications. Several examples of its applications are as test oscillators or signal generators for testing of radio receivers, measurement of standing-wave ratio, signal-to-noise ratio, transducer oscillators, measurement of an unknown capacitance in some oscillator-based schemes, realization of very low frequency (VLF) oscillations for biomedical, geophysical, or control applications, and also realization of voltage-controlled oscillators (VCOs) useful for analog phase-locked loops, A/D converters, frequency response display systems and spectrum analyzers (see [1]).

There are several proposed oscillator circuits. One of those oscillators is known as Wien-Bridge Oscillator circuit in [2]. The improvement of Wien-Bridge Oscillator was proposed in [3] by considering the nonlinear characteristic of Operational Amplifier (op-amp) such as slew rate to improve the frequency range of this oscillator. Another oscillators were also proposed in [4] (with total about 18 novel oscillators were proposed), [5] and [6]. Those oscillators use op-amp as an amplifier. Another oscillator were proposed in [7] by using the op-amp as a buffer to optimize its bandwidth. Recently, D.R. Bhaskar and Raj Senani investigated the use of Current Feedback Operational Amplifier (CFOA) in oscillator circuit design (see [1]).

Each oscillator has its own specific criterion that makes the circuit produces a sinusoidal output. This specific criterion in most oscillators constrains its frequency controllability. Frequency changes in oscillators such as in [2] and [3] might violate the oscillation criterion. In these circuits, frequency changing is achieved by changing the value of two resistors simultaneously in order to keep the oscillation criterion. Oscillator designs were proposed in [4], [7] and [1] where their frequency are controlled only by a single element without violating the oscillation criterion.

In this paper, we proposed another design of oscillator. Theoretically, it has an advantage that it does not have any specific criterion in order to make this circuit produces a sinusoidal output. This property ensures the oscillation stability of this oscillator. In this oscillator, output frequency can be controlled by 4 elements, instead of just one element.

This oscillator is implemented using op-amps, resistors and capacitors. However, the characteristics of op-amps influence the characteristics of the proposed oscillator. A realistic model that also considers the characteristics of op-amps is needed to predict the behaviour of this oscillator more accurately.

In this paper, part II-A explains the basic idea of the proposed oscillator. Then part II-B explains system design of the proposed oscillator. This ideal system design is referred as the ideal model. The circuit implementation of the oscillator is shown in part II-C. The construction of this oscillator's realistic model is discussed in part III. The design is verified by performing simulations to verify the theoretical prediction. Simulations in part IV is performed using both the ideal model and realistic model.

II. Oscillator Design

A. General Overview

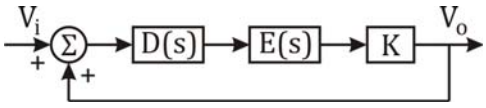


Fig. 1. System with Positive Feedback

Consider a system consists of two subsystems $D(s)$, $E(s)$ and gain controller K as shown in Fig. 1. The transfer function of this closed loop system is :

$$A_v(s) = \frac{T(s)}{1 - T(s)} = \frac{KD(s)E(s)}{1 - KD(s)E(s)} \qquad (1)$$

where $T(s)$ is the open loop transfer function of this system. Then we choose the transfer functions $D(s)$ and $E(s)$ so that:

$$T(s) = K\left[\frac{A(s) - B(s)}{A(s) + C(s)}\right] \tag{2}$$

The Barkhausen criterion states that oscillation occurs when $1 - T(s) = 0$. If we let $K = 1$, the oscillation occurs when:

$$1 - T(s) = \frac{C(s) + B(s)}{A(s) + C(s)} = 0 \tag{3}$$

or

$$C(s) + B(s) = k_2 s^2 + k_0 = 0 \tag{4}$$

where $C(s)$ and $B(s)$ are defined as $C(s) = k_2 s^2$ and $B(s) = k_0$. This system then produces a sinusoidal output with frequency of :

$$\omega_o = \sqrt{\frac{k_0}{k_2}} \tag{5}$$

B. System Design

In order to build a system where its transfer function satisfies (2), we propose a system with the transfer function of $D(s)$ and $E(s)$ are defined as:

$$D(s) = \frac{G_1}{G_1 + G_2} \tag{6}$$

$$E(s) = \frac{G_3 - G_4}{G_3} \tag{7}$$

where G_i is the conductance of the i-th component. If we let $K = 1$, these subsystems produce a system with transfer function:

$$T(s) = \frac{G_1 G_3 - G_1 G_4}{G_1 G_3 + G_2 G_3} \tag{8}$$

From (2) and (8), we know that $C(s) = G_2 G_3$ and $B(s) = G_1 G_4$. To satisfy (4), we choose $G_1 = \frac{1}{R_1}$ (Resistor), $G_2 = sC_2$ (Capacitor), $G_3 = sC_3$ (Capacitor) and $G_4 = \frac{1}{R_4}$ (Resistor). The open loop transfer function of this system is :

$$T(s) = \left(\frac{1}{sR_1C_2 + 1}\right)\left(1 - \frac{1}{sR_4C_3}\right) \tag{9}$$

Characteristic equation of this closed loop system is:

$$C(s) + B(s) = C_2C_3s^2 + \frac{1}{R_1R_4} = 0 \tag{10}$$

and this system produces a sinusoidal output waveform with frequency :

$$\omega_o = \sqrt{\omega_1\omega_2} \tag{11}$$

where $\omega_1 = \frac{1}{R_1C_2}$ and $\omega_2 = \frac{1}{C_3R_4}$.

The oscillator systems in [2] and [7] need a specific oscillation criterion that sometimes difficult to achieve and might be violated when the frequency is changed. On the contrary, this oscillator system theoretically does not have any specific oscillation criterion, thus the output frequency can be changed without violating the oscillation stability. However in practical implementation, the characteristics of op-amps influence the system and change the characteristics of this system. The effect of op-amp characteristics is discussed further in part III.

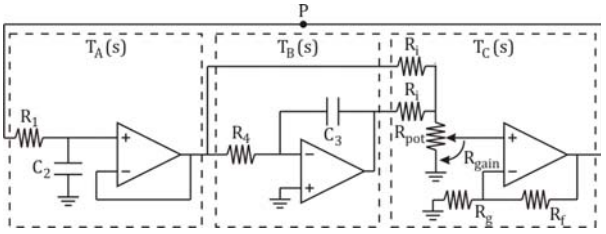


Fig. 2. Oscillator Configuration

C. Circuit Implementation

Fig. 2 shows the implementation of the proposed oscillator system. This circuit consists of 3 sub-circuits, a Buffered Low Pass Filter $T_A(s)$, an Integrator $T_B(s)$ and a Non-Inverting Summing Amplifier $T_C(s)$. If the feedback loop of this circuit is disconnected at P, the open-loop transfer function of this circuit can be stated as :

$$T(s) = T_A(s)[(1 + T_B(s)]T_C(s) \tag{12}$$

The value of R_{pot} is chosen so that $R_{pot} \gg R_i$. It can be seen that if $R_g = R_f$ and using the assumption that the op-amps are ideal, (12) is equivalent to (9).

III. Circuit Analysis

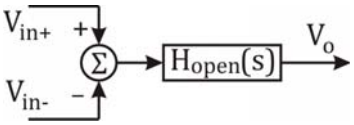


Fig. 3. Model of Operational Amplifier

In this analysis we build a realistic model of this oscillator circuit by considering the characteristics of op-amps to predict its behaviour more accurately. There are some assumptions that are made namely :

1) The op-amps have very high input impedance and very low output impedance.
2) The frequency and peak voltage of its output are constrained, so that the distortion due to their finite slew rate does not occur.
3) The op-amps can be modeled as a system shown in Fig. 3 and the open-loop transfer function of this op-amp is:

$$H_{open}(s) = \frac{A_{ol}\omega_L\omega_H}{(s + \omega_L)(s + \omega_H)} \tag{13}$$

where A_{ol} is the open loop (DC) gain of op-amps and $[-\omega_L, -\omega_H]$ are the poles of op-amps.

Using these assumptions, transfer functions $T_A(s)$, $T_B(s)$, $T_C(s)$ can be determined and we obtain :

$$T_A(s) = \left[\frac{\omega_1}{s + \omega_1}\right]\left[\frac{A_{ol}\omega_L\omega_H}{(s + \omega_L)(s + \omega_H) + (1 + A_{ol})\omega_L\omega_H}\right] \tag{14}$$

$$T_B(s) = -\frac{A_{ol}\omega_L\omega_H\omega_2}{(s + \omega_L)(s + \omega_H)(s + \omega_2) + s(A_{ol}\omega_L\omega_H)} \tag{15}$$

$$T_C(s) = \left[\frac{K}{2\beta}\right]\left[\frac{\beta A_{ol}\omega_L\omega_H}{(s+\omega_L)(s+\omega_H)+(1+\beta A_{ol})\omega_L\omega_H}\right] \quad (16)$$

where $\beta = \frac{R_g}{R_f+R_g}$ and $K = \frac{R_{gain}}{R_{pot}}$. Finally, from (12), (14), (15) and (16) , the open-loop transfer function of this oscillator circuit, $T(s)$, can be determined. Now, we have a complete model of this oscillator circuit. This model can then be used to perform a further analysis of this oscillator circuit.

Based on the realistic model, this closed-loop system has 8 poles (instead of 2 poles in ideal condition). A stable oscillation is generated when a pair of poles is positioned on the imaginary axis and the other poles are in the left half-plane. It can be achieved by adjusting the gain K. In this paper, this value of K is calculated using **rlocfind()** command in MATLAB. Then this gain value is used when performing Multisim simulation to generate a stable oscillation.

Frequency changes can be done by changing the value of resistors (R_1,R_4) and/or capacitors (C_2,C_3). However, this change affects the oscillation stability and the output is either vanished or distorted. So re-adjustment of gain value K is needed whenever there is a change in the value of resistors and/or capacitors.

IV. Simulation Results and Discussion

The simulation is performed virtually using two different softwares, MATLAB (simulates the ideal and realistic model of oscillator) and circuit simulator (simulates circuit implementation of this oscillator). MATLAB simulation is performed to verify that the behaviour of the oscillator basic model (as shown in II-B) can work well as an oscillator.

Then simulation using realistic model is also performed. First, this simulation is performed using MATLAB to calculate gain value that leads to a stable oscillation and its corresponding frequency for various values of ω_1 and ω_2. Then the gain value from MATLAB simulation is used to perform an experimental verification of this circuit in real condition using Multisim as circuit simulator. Before performing those simulations, the open-loop transfer function of op-amp is determined to complete our realistic model.

A. Transfer Function of Operational Amplifier

The op-amps that have a high slew rate (SR) value are chosen to produce a high frequency output signal. TL072 family of op-amp used in this simulation has minimal SR of $8V/\mu s$. This op-amps are supplied with $V_{cc} = \pm 9V$ so it can produce the maximum output voltage (V_{max}) about $7.5V$. This operating condition limits the output frequency that can be generated. The maximum permissible output frequency can be determined using equation :

$$f_{max} = \frac{SR}{2\pi V_{max}} \quad (17)$$

So, the maximum permissible output frequency of the op-amps under this condition is about $170kHz$. The frequency value used in the simulation are adjusted accordingly.

To determine the transfer function of TL072, the AC analysis of this open-loop op-amp is performed. The circuit configuration that used in this simulation is shown in Fig. 4 and its result is shown in Fig. 5.

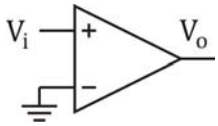


Fig. 4. Circuit Configuration to Determine the Open-Loop Transfer Function of Operational Amplifier

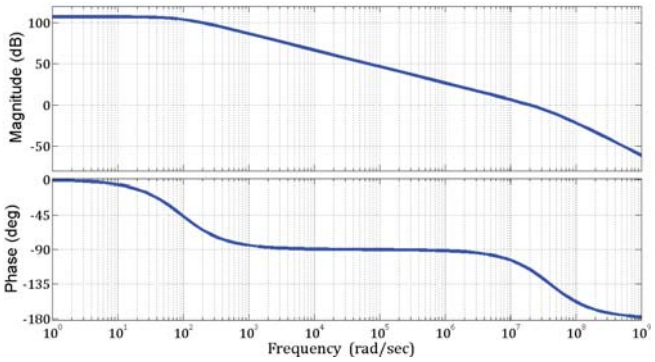


Fig. 5. AC analysis result of the TL072 open-loop gain

Fig. 5 shows that there are two transitions in the phase response, from 0^0 to -90^0 and from -90^0 to -180^0. These transitions imply that this system has two real poles. These transition can also be used to determine the location of poles. The poles' location of this system is at the middle of transitions (in this case when the phase are -45^0 and -135^0). So, it can be concluded the position of poles are at $s = -\omega_L = -9.44 \times 10^1 rad/s$ and $s = -\omega_H = -4.04 \times 10^7 rad/s$. From the magnitude response at very low frequency, it indicates that the open loop (DC) gain of this op-amps is $A_{ol} = 2.29 \times 10^5$.

B. Simulation Result using Ideal Model

This simulation is performed using MATLAB to simulate the ideal model this oscillator. The ideal oscillator, where its open-loop transfer function is defined in (8), can be implemented using a scheme as shown in Fig. 6. The simulation is performed by examining the output signal of this oscillator after an impulse signal is given as the input.

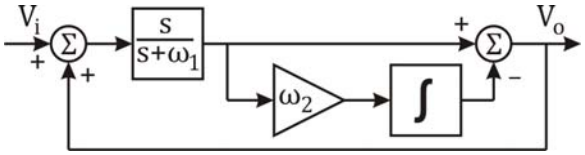


Fig. 6. System Configuration for Simulation using Ideal Model

The result of this simulation are shown in Fig. 7. From this result, it can be verified that the output signal of this oscillator is a sinusoidal signal. In this case, it indicates that an oscillation occurs in this system. Given a change in component's value. As the result, the oscillation occurs in this system with different frequency. This result indicates that the ideal model of this oscillator can work well because a stable oscillation occurs in this oscillator and this condition does not change even when there are some changes in the component's value.

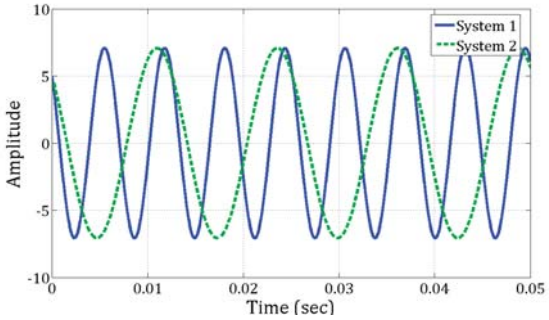


Fig. 7. Simulation Result using Ideal Model

C. Simulation Result using Realistic Model

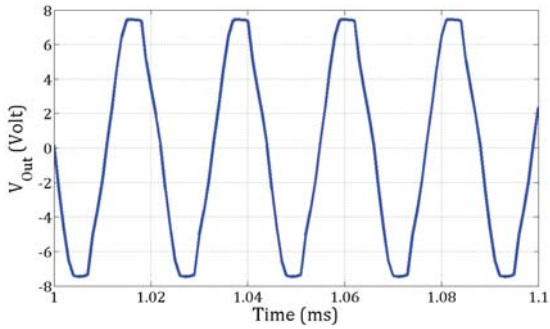


Fig. 8. Simulation Result of Oscillator Circuit with $\omega_1 = 10^6 rad/s$, $\omega_2 = 10^5 rad/s$ and $K = 1$

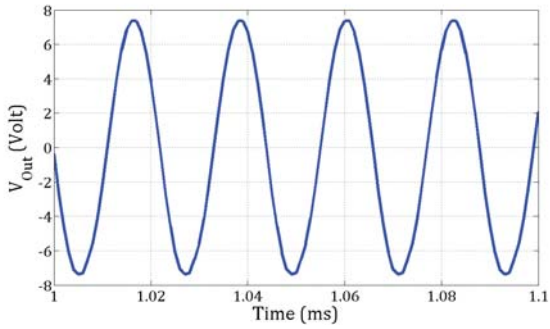


Fig. 9. Simulation Result of Oscillator Circuit with $\omega_1 = 10^6 rad/s$, $\omega_2 = 10^5 rad/s$ and gain K is Adjusted $(K = 0.98)$

After MATLAB simulation, a simulation in a more realistic condition is also performed using circuit simulator. In this simulation, we choose the component's value so that $\omega_1 = 10^6 rad/s$, $\omega_2 = 10^5 rad/s$ and $K = 1$ $\left(\frac{R_{gain}}{R_{pot}} = 1\right)$. The result of this simulation is shown in Fig. 8.

Ideally, in this condition the oscillator circuit produces a sinusoidal waveform as shown in Fig. 7. But in real condition, the oscillator circuit produces a distorted sinusoidal waveform. When the gain value K is adjusted, this circuit can produce a better sinusoidal waveform as shown in Fig. 9. In this case, the gain value $K = 0.98$.

This condition shows that a more realistic model of this

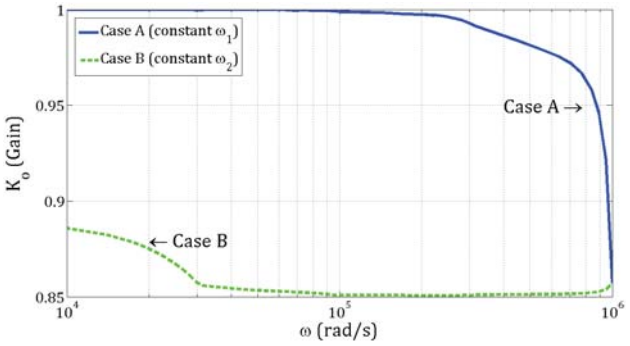


Fig. 10. The value of K_o for various value of ω. In case A, $\omega_1 = 10^6 rad/s$ and in case B, $\omega_2 = 10^6 rad/s$ (Note : $\omega = \sqrt{\omega_1 \omega_2}$)

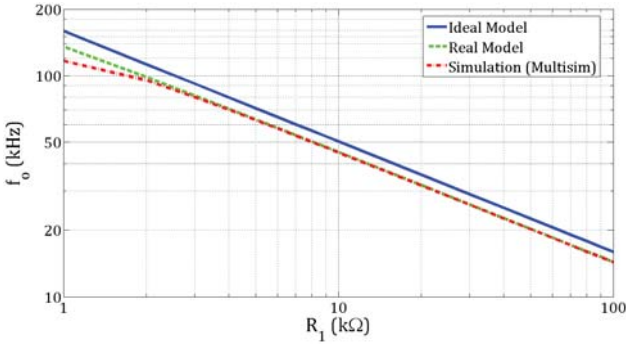


Fig. 11. Oscillation frequency (f_o) as the function of R_1 for $C_2 = 1nF$ and $\omega_2 = 10^6 rad/s$.

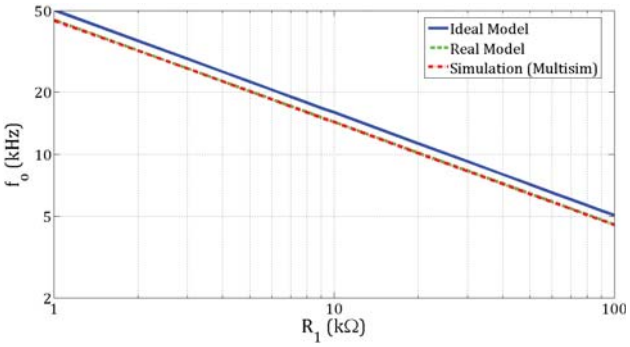


Fig. 12. Oscillation frequency (f_o) as the function of R_1 for $C_2 = 10nF$ and $\omega_2 = 10^6 rad/s$.

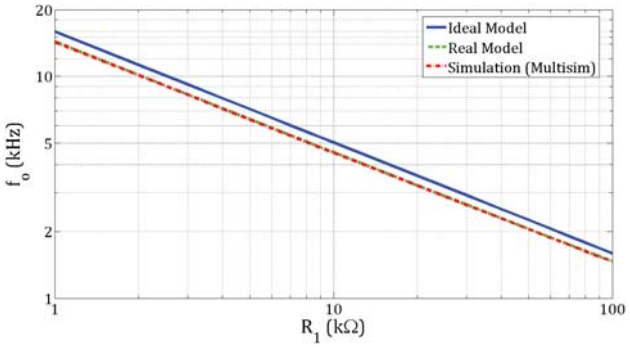


Fig. 13. Oscillation frequency (f_o) as the function of R_1 for $C_2 = 100nF$ and $\omega_2 = 10^6 rad/s$.

oscillator circuit is needed to analyze this circuit. We have to use the realistic model that is explained in part III to predict the behaviour of this circuit more accurately. This model can be used to calculate K_o, the gain value that leads to a stable oscillation, and also its corresponding frequency for various value of ω_1 and ω_2. The value of K_o for various value of ω_1 and ω_2 is shown in Fig. 10. Then this gain value K_o is used when performing simulation using circuit simulator as an experimental verification of this circuit in real condition.

Fig. 11 - Fig. 16 show the relation between oscillation frequency f_o and the value of components (R_1, C_2, C_3 and R_4). In each figure, there are 3 lines that represent 3 different prediction methods. The first one is the theoretical prediction using ideal model. The second one is the prediction of output frequency using realistic model where the frequency output is predicted using **rlocfind()** command in MATLAB. And the last one is the simulation result that is performed using circuit simulator.

Those figures show that the oscillation frequency in real condition (simulated using circuit simulator) differs from the theoretical prediction using ideal model. It shows that the characteristics of op-amp cause a significant oscillation frequency deviation from the ideal condition. However, the theoretical prediction using realistic model can fit perfectly with the oscillation frequency in real condition. So we can conclude that this realistic model can be used to predict the behaviour of this oscillator circuit accurately.

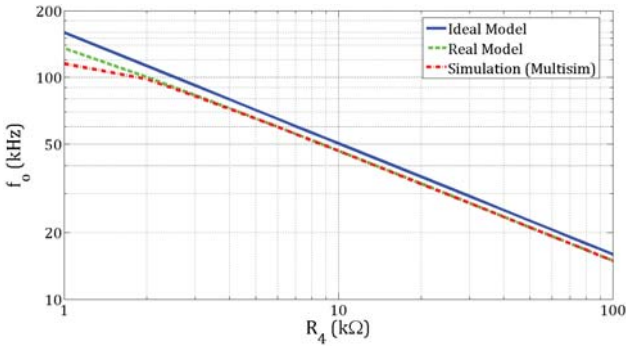


Fig. 14. Oscillation frequency (f_o) as the function of R_4 for $C_3 = 1nF$ and $\omega_1 = 10^6 rad/s$.

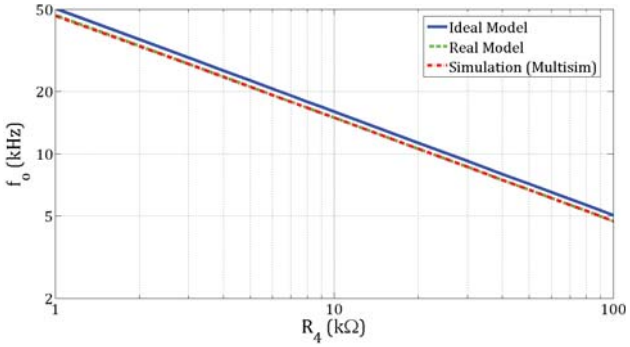


Fig. 15. Oscillation frequency (f_o) as the function of R_4 for $C_3 = 10nF$ and $\omega_1 = 10^6 rad/s$.

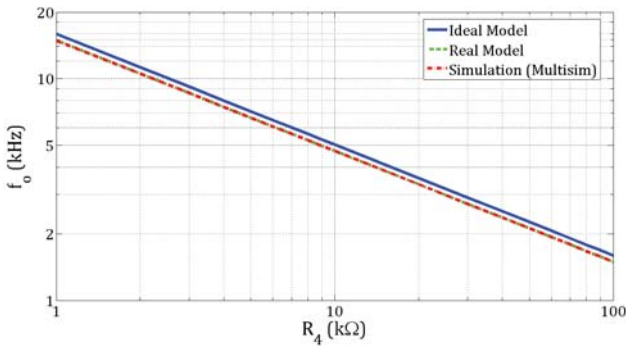


Fig. 16. Oscillation frequency (f_o) as the function of R_4 for $C_3 = 100nF$ and $\omega_1 = 10^6 rad/s$.

V. Conclusion and Further Works

A new sinusoidal oscillator circuit is proposed in this paper. The ideal model of the proposed oscillator can work well to produce a sinusoidal output. This circuit design ensures the oscillation stability. Any changes in this oscillator's frequency does not violate the oscillation stability. Some changes in the characteristics of this oscillator occur when this oscillator is implemented using op-amps and passive components. Different value of gain K is needed for each oscillator frequency to obtain a stable oscillation and its frequency also differs from the prediction of ideal model.

In this paper, it explains that the implementation of this oscillator causes some changes in oscillator's ideal characteristics. So, further investigation of the implementation of this oscillator is needed. The purpose is to obtain the implementation of this oscillator where its characteristics are close to the ideal ones. In the end, we hope we can obtain an oscillator where its oscillation stability is ensured.

Acknowledgment

The authors would like to thank Mr. Reka Inovan for his review and his constructive feedbacks that were really helpful in the preparation of this manuscript.

References

[1] D. R. Bhaskar and R. Senani, "New CFOA-Based Single-Element-Controlled Sinusoidal Oscillators", *IEEE Trans. Instrumentation and Measurement*, vol. 55, no. 6, pp. 2014-2021, Dec. 2006.

[2] A. Budak, and K. Nay, "Operational Amplifier Circuits for Wien-Bridge Oscillator", *IEEE Trans. Circuit Syst.*, vol. CAS-28, no. 9, pp. 930-934, Sep. 1981.

[3] J. L. Huertaz, A. Rodriguez-Vazquez and B. Perez-Verdu, "High-Frequency Design of Sinusoidal Oscillators Realised with Operational Amplifiers", *IEE Proc. G, Electron. Circuits & Syst.*, vol. 131, no. 4, pp. 137-140, Aug. 1984.

[4] R. S. Sidorowicz, "An Abundance of Sinusoidal RC Oscillators", *IEE Proc.*, vol. 119 no. 3, pp. 283-293, Mar. 1972.

[5] V. Prem Praya, S. C. Dutta Roy and S. S. Jamuar, "Identification and Design of Single Amplifier Single Resistance Controlled Oscillators", *IEEE Trans. Circuit Syst.*, vol. CAS-30, no. 3, pp. 176-181, Mar. 1983.

[6] N. Boutin, "On Identification and Design of Single Amplifier Single Resistance Controlled Oscillators", *IEEE Trans. Circuit Syst.*, vol. CAS-31, no. 12, pp. 1046-1048, Dec. 1984.

[7] R. Senani, "New RC-Active Oscillator Configuration Employing Unity-

Gain Amplifiers", *Electronics Letters*, vol. 21, no. 20, pp. 889-891, Sep. 1985.

Asynchronous Delta-Sigma Modulator with Multiple-Valued Output

Arif Abdul Mannan, Hiroki Tamura, and Koichi Tanno
Department of Electrical and Electronic Engineering
University of Miyazaki
Miyazaki, Japan
tanno@cc.miyazaki-u.ac.jp

***Abstract*—In this paper, a continuous-time asynchronous delta-sigma (ΔΣ) modulator (ADSM) is proposed. The proposed circuit is multiple-valued output. The circuit is consists of hysteresis comparator, operational transconductance amplifier, and digital-to-analog circuits. Next, an example circuit of 5-level ADSM is discussed and simulated. An analog-to-digital converter consisted of the 5-level ADMS and multiple-valued logic decoder and digital counter is presented and simulated. All of the proposed circuits are simulated in transistors level through HSPICE with set of 2.0µm CMOS process parameters. Detailed simulation results are shown in this paper.**

Keywords—multiple-valued logic; asynchronous; delta-sigma modulator; time-continuous; analog-to-digital converter

I. INTRODUCTION

The advance technology in digital CMOS processes result the increase of the performance in the intrinsic speed of the transistors, give benefits for analog CMOS implementation to obtain high time resolution in the circuit application. Therefore high accuracy CMOS analog circuit can be obtained by using analog signal in time domain circuits instead of the conventional voltage domain [1]-[2]. One of the time domain analog circuit implementation is a duty-cycle modulation or asynchronous delta-sigma (ΔΣ) modulator [1]-[4]. Multiple-valued logic (MVL) application in analog circuit [5],[6] is one of possible solution to solve the problem such as increasing of circuit complexity, interconnection, and power consumption. In order to overcome this kind of problems, MVL is also applied in asynchronous ΔΣ modulator as multiple-valued logic asynchronous ΔΣ modulator (MADSM).

A MADSM is presented in this paper. The MADSM is expanded from the binary asynchronous ΔΣ modulator (BADSM) by using MVL circuits. The different between the BADSM and the MADSM is hysteresis comparator; the binary hysteresis comparator is used in BADSM, while MVL hysteresis comparator is used in MADSM. In this paper, the concrete circuit schematic of the proposed modulator is shown. The modulator is evaluated through HSPICE simulations with 2-metals, twin-well 2.0µm CMOS process. Furthermore, the analog-to-digital converter (ADC) is proposed as an application of the proposed asynchronous ΔΣ modulator. This ADC consists of the proposed modulator and MVL decoder and digital counter (MDDC). This application circuits is also confirmed through HSPICE.

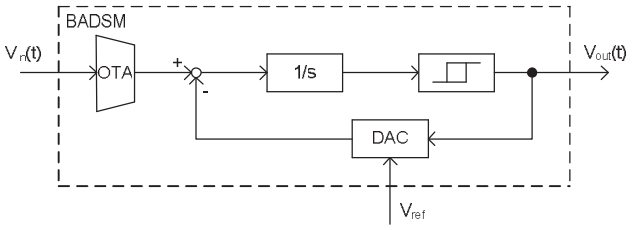


Fig. 1. Block diagram of binary asynchronous ΔΣ modulator.

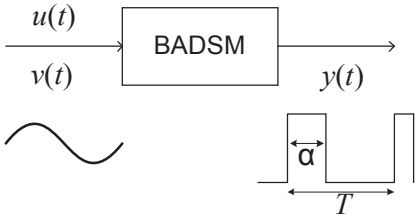


Fig. 2. Input and output waveforms of BADSM.

In section II, the BADSM consisted of an operational transconductance amplifier (OTA), an integrator, a hysteresis comparator and a digital-to-analog converter (DAC) is reported. The application of MVL in asynchronous ΔΣ modulator as MADSM circuit is also discussed. The example circuit of the MADSM is presented and the simulation result is also shown in Section III and Section IV, respectively. As MADSM application, the MVL ΔΣ analog-to-digital converter is discussed in Section V.

II. MVL DELTA SIGMA MODULATOR

A block BADSM is shown in Fig. 1 [1]. The BADSM is consists of an OTA which is convert the analog input voltage into current, an integrator (1/s), a hysteresis comparator, and a DAC which is also convert the reference voltage into current as a feedback. With the hysteresis window in a hysteresis comparator, no clock is needed in BADSM [1],[4].

The input and output waveforms of BADSM are shown in Fig. 2. The $u(t)$ is defined as input voltage (V_{in}) normalized

from the reference voltage (V_{ref}) with $|u(t)| \leq 1$, and $y(t)$ is defined as digital output voltage (V_{out}) variable continuously in time domain. The duty cycle α/T and the frequency $\omega=(2\pi/T)$ of the output $y(t)$ is given by

$$\frac{\alpha}{T} = \frac{u(t)+1}{2} \tag{1}$$

and

$$\frac{\omega}{\omega_c} = 1-u(t)^2 \tag{2}$$

respectively [1],[2]. The frequency ω_c is the free running frequency when the input $u(t)$ is zero.

If mono polar power supply is used (V_{ss} equal to GND), and V_{ref} is also referenced from GND, then the $v(t)$ is defined as input voltage (V_{in}) normalized from the reference voltage (V_{ref}) with $0 \leq v(t) \leq 1$. The duty cycle α/T and the frequency $\omega=2\pi/T$ of the output $y(t)$ is given by

$$\frac{\alpha}{T} = v(t) \tag{3}$$

and

$$\frac{\omega}{\omega_c} = 4\left(v(t)-v(t)^2\right) \tag{4}$$

respectively, and the free running frequency ω_c is occur when $v(t)$ is 0.5. From (1) and (3) the correlation between $u(t)$ and $v(t)$ is given by

$$v(t) = \frac{u(t)+1}{2} \tag{5}$$

From the binary modulator, BADSM, we try to expand the binary logic into MVL. For MVL circuits, the number of "*n*"-valued logic is representation of the number of multi-valued level. In this paper, *n*-valued in MVL circuits means *n* number of values, which consists of V_i, where $i = \{0, 1, 2, \ldots, n\text{-}1\}$.

For proposed MADSM, the input voltage is divided into *n*-1 area, with each area is defined as $v_{j,j+1}(t)$, where $0 \leq v_{j,j+1}(t) \leq 1$ and $j = \{0, 1, 2, \ldots, n\text{-}2\}$. Note that for $v_{j,j+1}(t)$, $v_{0,1}(t)$ is equal to $v(t)$ from V_0 to V_1 area. Therefore the $m(t)$ is also defined as input voltage (V_{in}) divide by the divided reference voltage by number of logic minus 1 ($V_{ref}/(n-1)$) with $0 \leq m(t) \leq n\text{-}1$, and $y_{j,j+1}(t)$ is a MVL output voltage (V_{out}) variable continuously in time domain. The MVL output $y(t)$ has *n* level output logic with *n*-1 number of duty cycle α/T and frequency $\omega=2\pi/T$, with each duty cycle and frequency is represent each area from the divided input voltage. Note that the divided area has only positive value, therefore $u(t)$ from (1) and (2) cannot be used. The block diagram, input, and output waveforms of MADSM (in case of n=3) are shown in Fig. 3.

The duty cycle and the frequency of the MVL output $y(t)$ for $y_{j,j+1}(t)$ is given by

$$\frac{\alpha_{j,j+1}}{T_{j,j+1}} = v_{j,j+1}(t) \tag{6}$$

and

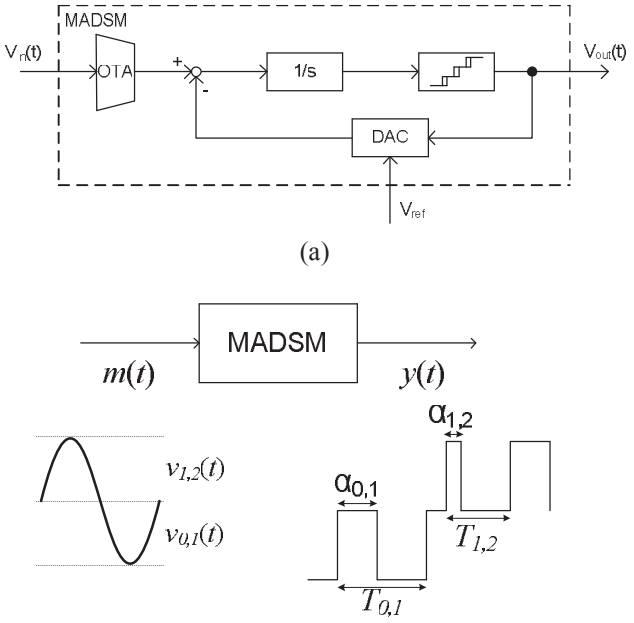


Fig. 3. Block diagram (a) and input and output waveforms (b) of MADSM.

$$\frac{\omega_{j,j+1}}{\omega_{c(j,j+1)}} = 4\left(v_{j,j+1}(t)-v_{j,j+1}(t)^2\right) \tag{7}$$

where

$$v_{j,j+1}(t) = m(t)-i \tag{8}$$

From (8) and the definition of $m(t)$, $v_{j,j+1}(t)$ is given by

$$v_{j,j+1}(t) = \frac{(n-1)V_{in} - i \cdot V_{ref}}{V_{ref}} \tag{9}$$

(9) is base equation for the proposed MADSM circuit.

The correlation between input voltage V_{in}, reference voltage V_{ref}, $m(t)$, $v_{j,j+1}(t)$, and output voltage $y(t)$ and $y_{j,j+1}(t)$ (in case of n=3) is shown in Fig. 4. From (6)-(9) and Fig. 4, it is obvious that the proposed MADSM circuit can be realized by stacking *n*-1 number of BADSM circuits. Each BADSM has its own input area and has its own output area. Therefore there will be no "collision" between one BADSM and another.

For actual circuit implementation, the direct implementation of *n*-1 number of stacked BADSM will be difficult. Therefore the MVL hysteresis comparator is used instead of standard hysteresis comparator, and the direct implementation of equation (9) can simplify the overall circuit. Therefore only one integrator, one multiple-input or weighted-input OTA, one multiple-output or weighted-output or MVL DAC and one MVL hyste*r*esis comparator with *n*-1 number of hysteresis windows are required. In this paper, multiple-input OTA and MVL hysteresis comparator are used.

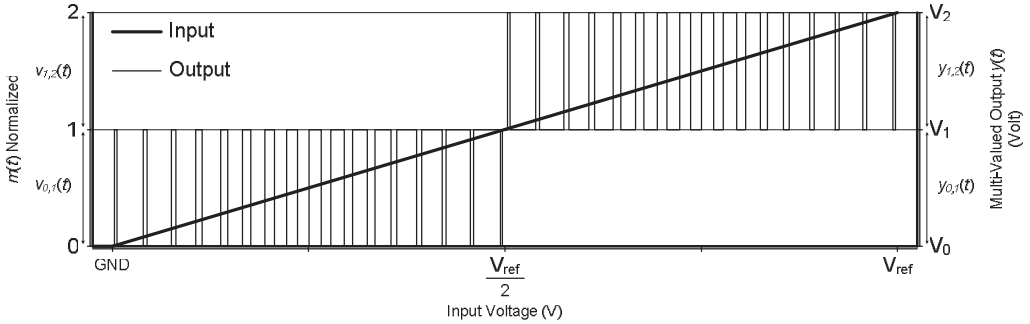


Fig. 4. Graph of input voltage V_{in} (from GND to V_{ref}) with $m(t)$, $v_{i,i+1}(t)$, and output voltage $y(t)$.

The example of the proposed MADSM circuit is shown in Fig. 5. The proposed circuit is 5-level asynchronous ΔΣ modulator. In the proposed circuit, an 8-input fully differential OTA is used. The input voltage is connected to 4-input terminals of OTA to implementing multiplication $(n\text{-}1)V_{in}$, while 4-line output of DAC is connected to the another 4-input terminals of OTA to implementing multiplication $i \cdot V_{in}$ (0 to 3 lines) and feedback (1-line). The 4-line DAC is controlled by 4-bit output of the MVL hysteresis comparator V_i, which $i = \{1, 2,3,4\}$.

For example, in case of n=5, the input voltage is divided into 4 area. If input voltage is some were between GND ≤ V_{in} ≤ ¼ V_{ref}, then i=0. The integral output is operated on hysteresis area 1 (between V_0 and V_1). Therefore from (9), the normalized input $v_{0,1}(t)$ is given by

$$v_{0,1}(t) = \frac{4 \cdot V_{in}}{V_{ref}} \tag{10}$$

and the duty cycle

$$\frac{\alpha_{0,1}}{T_{0,1}} = \frac{4 \cdot V_{in}}{V_{ref}} \tag{11}$$

In this condition, the MADSM circuit is work as duty-cycle modulator with modulated output between V_0 and V_1.

If the input voltage is some were between ½ V_{ref} ≤ V_{in} ≤ ¾ V_{ref}, then i=2. The integral output is operated on hysteresis area 3 (between V_2 and V_3). Therefore $2 \le m(t) \le 3$, and the normalized input $v_{2,3}(t)$ is given by

$$v_{2,3}(t) = \frac{4 \cdot V_{in} - 2 \cdot V_{ref}}{V_{ref}} \tag{12}$$

and the duty cycle

$$\frac{\alpha_{2,3}}{T_{2,3}} = v_{2,3}(t) \tag{13}$$

In this condition, the MADSM circuit is work as duty-cycle modulator with modulated output between V_2 and V_3.

From this point forward, the example circuit will be explained. The multi-input OTA from the example circuit is shown in Fig. 6. The OTA output characteristic is shown in Fig. 7.

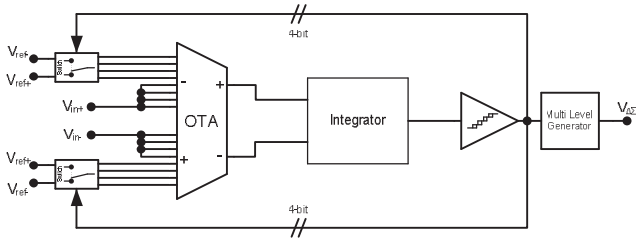


Fig. 5. The example of proposed 5-level asyscronous ΔΣ modulator circuit.
The Example Circuit

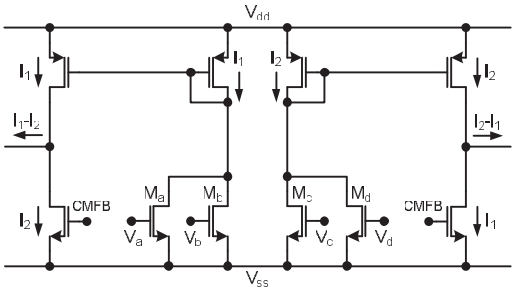


Fig. 6. Circuit schematic of the proposed n-type 2-input OTA.

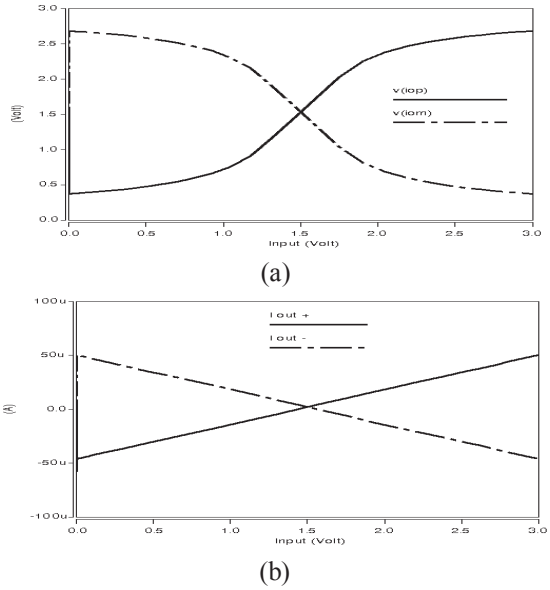


Fig. 7. Example OTA output characteristic in voltage (a) and current (b).

The n-type multi-input OTA is use transistors M_a, M_b, M_c, and M_d as input transistors, works as transconductance amplifier, and use current mirror and common mode feedback circuit to apply fully differential OTA. The number of input transistor can be varied depend on the requirement (8 input transistors for example circuit). Therefore the output of multi-input OTA is given by

$$I_1 - I_2 = (I_a + I_b) - (I_c + I_d) \quad (14)$$

and

$$I_a = K_0 \frac{W_{Ma}}{L_{Ma}} (V_a - V_T)^2 \quad (15)$$

where K_0 is unit transconductance parameter and V_T is the internal MOSFETs threshold voltage. The n-type multi-input OTA is combined with p-type multi-input OTA to produce rail-to-rail input range.

The example of the proposed MADSM circuit in Fig. 5, use an Op-amp as an integrator. The example integrator circuit is a voltage mode integrator circuit shown in Fig. 8(a). Therefore use the OTA output characteristic shown in Fig. 7(a). The integrator circuit can also use current mode shown in Fig. 8(b) as proposed in [1],[7]. For current mode integrator, the OTA output characteristic shown in Fig. 7(b). The example circuit use voltage mode integrator (non linear OTA) for the input range analysis discussed in Section V.

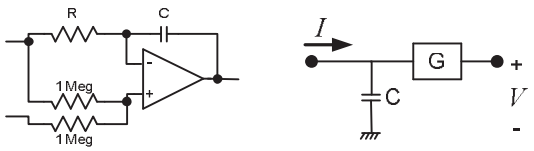
The implementation of MVL hysteresis comparator in MADSM is very important to achieve asynchronous operation, as same as the BADSM use hysteresis comparator [1]-[4]. The basic structure of MVL hysteresis comparator has already proposed in [6], but with improvement in time delay characteristic and hysteresis windows. The multi-level generator circuit is also similar to that proposed in [6], with improvement in time delay characteristic. The MVL hysteresis characteristics use in example circuit is shown in Fig. 9. The MVL hysteresis comparator has 4 hysteresis windows (n-1, with n=5). The hysteresis windows characteristic is ΔV_1=(1.243 - 1.187)=0.056V, ΔV_2=(1.428 - 1.368)=0.060V, ΔV_3=(1.603 - 1.554)=0.049V, ΔV_4=(1.849 - 1.769)=0.080V.

III. Simulation Result

In this paper, the transistor level simulation is performed. The proposed circuits shown in Fig. 5 were evaluated using HSPICE with 1-poly, 2-metal, 2-well 2.0µm standard CMOS process parameters and single 3.0V for V_{dd}. The input signal of this simulation consists of four frequencies of 1.3, 3, 7, and 10 kHz. The transient simulation and FFT results are shown in Fig. 10 and 11, respectively.

From Fig. 10, the proposed circuit is successfully implementing the equation (9), and 4 area of input range and 4 area of duty-cycle modulator output can be observed. At the FFT results in Fig. 11, four dominant output frequencies at 1.3, 3, 7, and 10 kHz with the noise floor level at -60.5dBV are also observed.

From the simulation result in Fig. 10 and Fig. 11, it shown that the proposed architecture shown in Fig. 5 is work well as theory at transistor level simulation.



(a) (b)

Fig. 8. Voltage (a) and current (b) mode integrator for OTA.

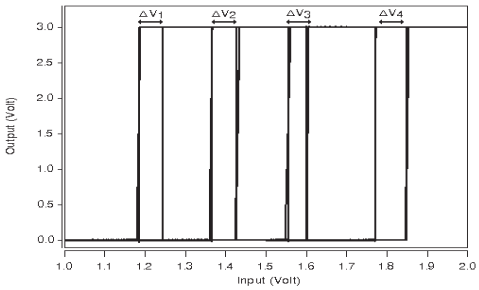


Fig. 9. The MVL hysteresis comparator output, DC characteristic.

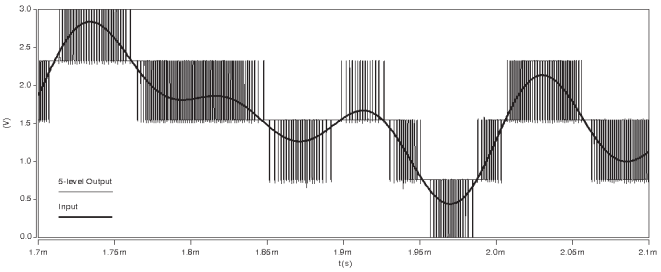


Fig. 10. Simulation result of example circuit with 3.0V V_{dd}, 3.0V differential V_{ref}, 3.0 Vp-p differential input voltage (4 frequencies input: 1.3, 3, 7, and 10 kHz @0.75Vp-p differential input voltage).

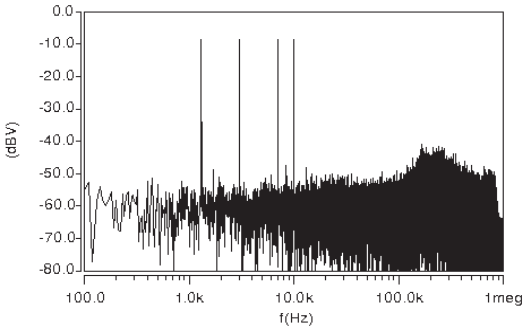


Fig. 11. The FFT result of 5-level output of example circuit.

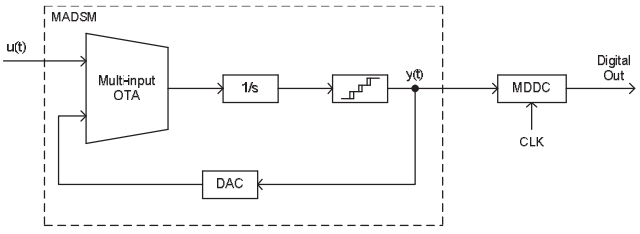


Fig. 12. The block diagram of MVL ΔΣ ADC.

IV. MVL ΔΣ ANALOG TO DIGITAL CONVERTER

In this Section, analog-to-digital converter is selected as MADSM application. The analog-to-digital converter is consists of a MADSM circuit and MDDC. The block diagram of MVL ΔΣ ADC is shown in Fig. 12.

The MDDC circuit is use similar operation as time-to-digital circuit proposed in [1], but in this paper, the MDDC work not only as time-to-digital converter and decimator, but also as MVL to digital decoder. The block diagram of MDDC is shown in Fig. 13.

The MDDC use directly the digitized output of MVL hysteresis comparator (shown in Fig. 9) rather than use the MVL output of multi-level generator. Therefore no another MVL threshold detector is required. The output of MVL hysteresis comparator is decoded and filtered in 3-bit zinc filter, than the MSB of the 3-bit digital output is counted by 8-bit counter. The MDDC operation is illustrated in Fig. 14.

Note that if logic V_4 is "1", than logic V_1 to logic V_3 is also "1", and if logic V_1 is "0", than logic V_2 to logic V_4 is also "0". The MDDC is applied with 2MHz clock input, and the output sampling rate is at 50 kHz with 10-bit output (use only 256 clk, 4 times faster than binary ΔΣ ADC application).

To examine the digital output result, a R2R DAC is used in simulation. R2R DAC has good performance and high accuracy in simulation level. The test setup for MVL ΔΣ ADC is shown in Fig. 15.

The input voltage for MVL ΔΣ ADC is use 3.0Vp-p differential input voltage with 4 dominant frequencies: 3, 5, 7, and 10 kHz. The transient respond of the MVL ΔΣ ADC and FFT result of R2R DAC output is shown in Fig. 16.

From Fig. 16(a) the analog input waveform is successfully reproduced. The FFT result shows the noise floor is -54.5dBV. The ADC integral nonlinearity with 3.0Vp-p input and 1.8Vp-p input is shown in Fig. 17.

Fig. 17(a) shows the effect of OTA nonlinearity, but the output is little different with the characteristic show in Fig. 7(a). This phenomenon is the result of the application equation (9), where the input voltage and V_{ref} is divided by n-1, or the $v_{j,j+1}(t)$ is given by

$$v_{j,j+1}(t) = \frac{V_{in} - i \cdot \dfrac{V_{ref}}{n-1}}{\dfrac{V_{ref}}{n-1}} \qquad (16)$$

giving high quality application in circuit with only small linear area of OTA.

V. CONCLUSION

In this paper, we have been proposed the new MVL asynchronous ΔΣ modulator (MADSM), which is operated as n-1 number of stacked binary asynchronous ΔΣ modulator (BADSM). The stacked BADSM can be simplified by using equation (9). The proposed MADSM is also applied as an analog-to-digital converter (ADC) for further analysis. The applied ADC circuit is 10-bit output digital with 50 kHz sampling frequency and -54.5dBV noise floor performance. From ADC integral nonlinearity analysis, we could confirm that application of MADSM give advantage in the application

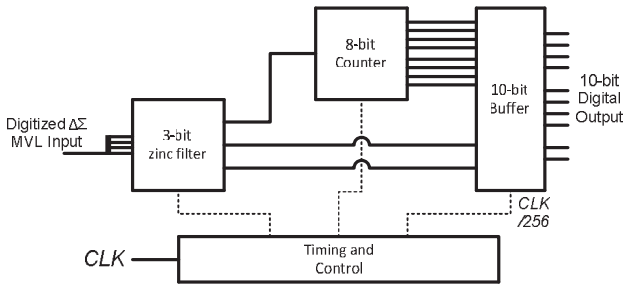


Fig. 13. The block diagram of MDDC.

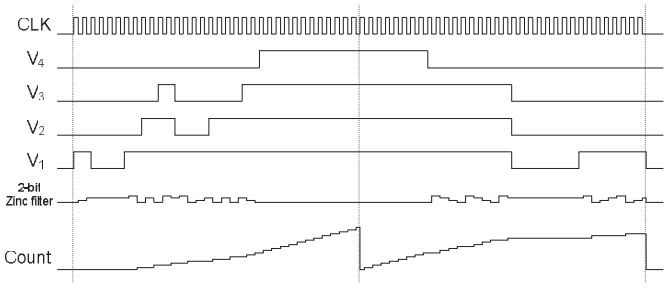


Fig. 14. Operation of the MDDC

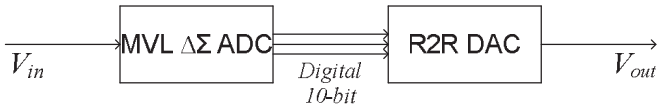
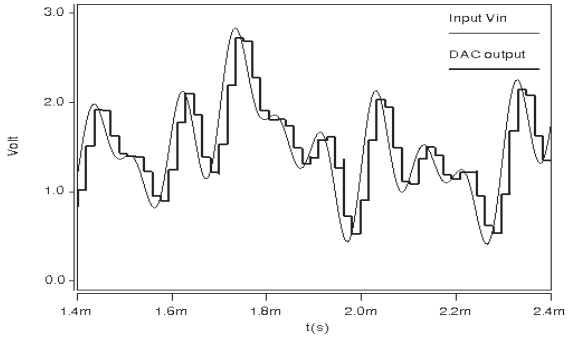
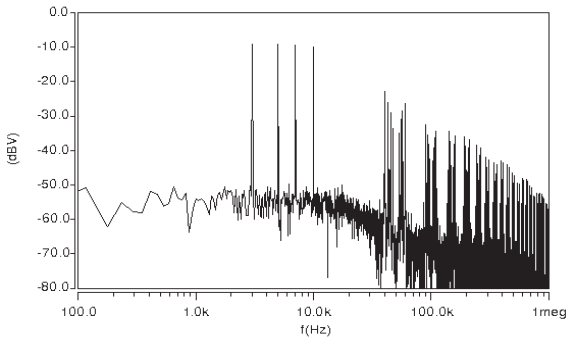


Fig. 15. Test setup for MVL ΔΣ ADC



(a)



(b)

Fig. 16. Transient respond (a) and FFT result of the DAC output

of small linear area of OTA. From simulation results, we could confirm that the proposed circuits work well as theory.

The output analysis of MADSM from the MVL hysteresis comparator characteristic (time propagation delay and the unequal hysteresis windows width and space between hysteresis windows), the OTA nonlinearity and the integrator non perfection is future work.

ACKNOWLEDMENT

This work is supported by VLSI Design and Education Center (VDEC), the University of Tokyo in collaboration with Synopsys, Inc. and Cadence Design Systems, Inc. I would like to thank to all staff of the University of Miyazaki and University of Brawijaya, for giving the chance to study and research at University of Miyazaki.

REFERENCES

[1] K. Kuribayashi, K. Machida, Y. Toyama, and T. Waho, "Time-Domain Multi-bit Delta Sigma Analog-to-Digital Converter," 41st IEEE International Symposium on Multiple-Valued Logic (ISMVL 2011), pp. 254-258, 2011.

[2] E. Roza, "Analog-to-digital conversion via duty-cycle modulation," IEEE Trans. Circuits and Systems II: Analog and Digital Signal Processing, Vol. 44, No.11, pp. 907 - 914, 1997.

[3] J. Daniels, W. Dehaene, M. Steyaert, and A. Wiesbauer, "A/D conversion using an Asynchronous Delta-Sigma Modulator and a time-todigital converter," Proc. IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1648 - 1651, 2008.

[4] J. Daniels, W. Dehaene, M. Steyaert, and A. Wiesbauer, "A/D Conversion Using Asynchronous Delta-Sigma Modulation and Time-to-Digital Conversion," IEEE Trans. Circuits and Systems I: Regular Papers, Vol. 57, No, 9, pp. 2404 - 2412, 2010.

[5] T. Waho, "Multiple-Valued Technique in Analog-to-Digital Converters: A Brief Survey," Proc. 16th International Workshop on Post-Binary ULSI Workshop, pp. 27 - 30, May 13, 2007.

[6] A. A. Mannan, K. Tanno, H. Tamura, T. Toyama, A. Darmawansyah, "Expandable MVL Inverter Compatible with Standard CMOS Process and Its Application to MVL Hysteresis Comparator," IEEE 43rd International Symposium on Multiple-Valued Logic (ISMVL), pp. 260-265, 2013

[7] M. Yano, "Design of Continuous-Time Delta-Sigma Modulator and Its Improvement for Reduction of 1/*f* Noise," Master Tesis of University of Miyazaki, Frebuary 2012.

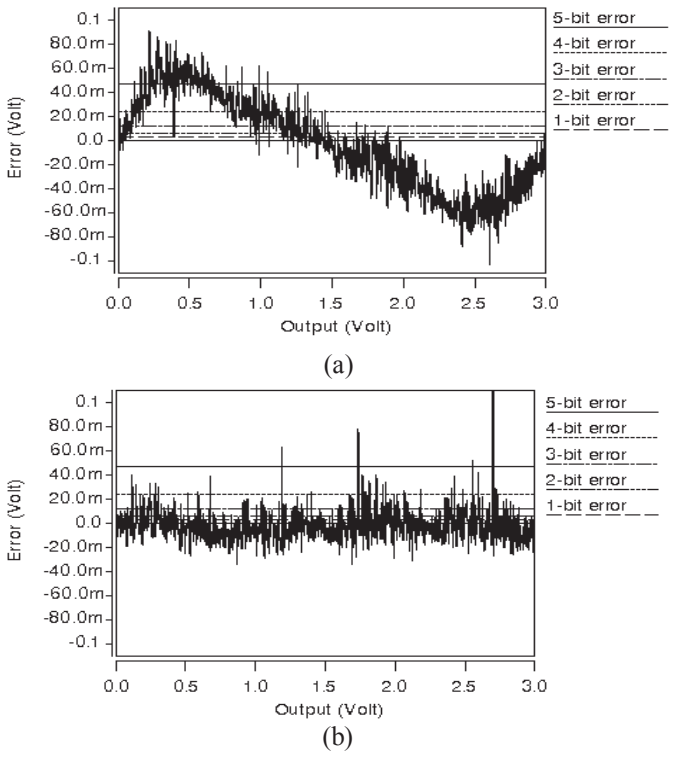


Fig. 17. Integral nonlinearity with 3Vp-p (a) and 1.8Vp-p (b) input range.

Blind and Lighting Control to Maintain Comfort Light Intensity of The Classroom Utilizing Microcontroller ATmega8535

Arthur Silitonga, and I Gusti L. W. Indrawan
Study Program of Electrical Engineering
Faculty of Engineering, President University
Bekasi, Indonesia
email : asilitonga@gmail.com

***Abstract*— A stable light intensity level of a classroom can actually be obtained by integrating automatic lamp brightness controller with automatic blind system. A brightness controller of the lamp, especially incandescent type can be developed by adopting the principle of phase firing angle. It should be implemented in such a way so that the firing angle is set precisely to keep the classroom's light intensity stays in a steady level. Next, the opening and closing process of the blind can also be done using stepper motor which will be controlled automatically by a microcontroller through a hardware driver. This is intended to take the advantage of light contribution from the outside instead of directly switching on or brightening the lamp**.

Keywords*— *light intensity, automatic lamp brightness controller, automatic blind system, phase firing angle, stepper motor, microcontroller

I. INTRODUCTION

Most of the artificial lighting systems are presented and operated base on the day and night only without taking any consideration of light contribution from the outside of the room, especially in the day. This situation can be inconvenient for the student inside a classroom to manually switch on or switch off the lamps several time according to the ambient light intensity level that may increases or drops significantly. Besides that, studying in a low light intensity will affect the eyes become tired faster than studying under a sufficient light intensity. On the other case, if presentation is done in a very bright classroom will not give a good display and if it is too dark it will not good also for the audience who need to make any notes. From this point of view, the idea to develop and to implement this project comes up. The main objective is to design and to model an automatic light intensity controller which will be used to maintain a proper lighting system for the classroom equipped with incandescent AC lamps to provide a comfortable learning atmosphere for the student. A comfortable light intensity can be maintained by integrating automatic brightness controller with automatic blind system. Automatic brightness controller is built by adopting the principle of phase firing angle while blind will be driven by stepper motor. A well planned blind and lighting system in a classroom will give significant effect to the learning atmosphere and concentration of the students or the users.

II. LITERATURE REVIEW

A. Phase Firing Angle

Phase firing angle is a time proportioning method to control the amount of power that goes into a load (either resistive load or inductive load) by cycle basis. By varying the portion of firing delay time in each cycle of the alternating current wave will determine the total voltage being delivered into the load. This means the shorter the firing delay, the more the power is delivered into the load and vice versa.

In order to get a precise firing delay time, a fixed reference point is required. This is called as zero crossing point. Zero crossing is the instantaneous point when alternating current sinusoidal wave crosses zero volt. Zero crossing occurs twice in every full cycle as shown on the Fig. 1[1][5].

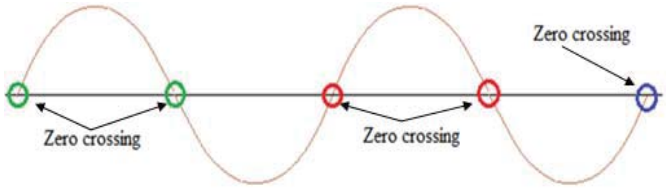


Fig. 1. Zero Crossing

Time interval of zero crossing points is determined by the frequency of AC sinusoidal wave. In Indonesia, the State Electricity Company (PLN) distributes 220 volts alternating current at a fundamental frequency of 50 Hz for each full cycle which gives a period of 20 ms. Since zero crossing point occurs twice in every full cycle, then it will occur every 10 ms

Theoretically, by knowing the firing delay time, then the firing angle can be determined by simply varying it after each zero crossing point between an interval of 0 up to 10 ms. To convert the firing delay into firing angle, this interval can be divided into 180 time slots and each time slot will be equal to 1 degree. From this information, if α is assumed as the firing angle and t_d is the firing delay, then the relation of these variables is simply determined by (1).

$$\alpha = 18\ t_d \quad (1)$$

For alternating current (AC), the input voltage oscillates in a sine wave pattern $v_i(t) = V_p \sin \omega t$ with V_p as the peak voltage and the average voltage can be calculated using (2).

$$V_{av} = \frac{V_p}{\pi}(\cos\alpha + 1) \quad (2)$$

If a non-RMS Voltmeter is used for the load voltage measurement, the result should be equivalent with (3).

$$V_{voltmeter} = V_{av}\, x \left(\frac{\pi}{2\sqrt{2}} \right) \quad (3)$$

III. PROJECT DESIGN

A. Block Diagram

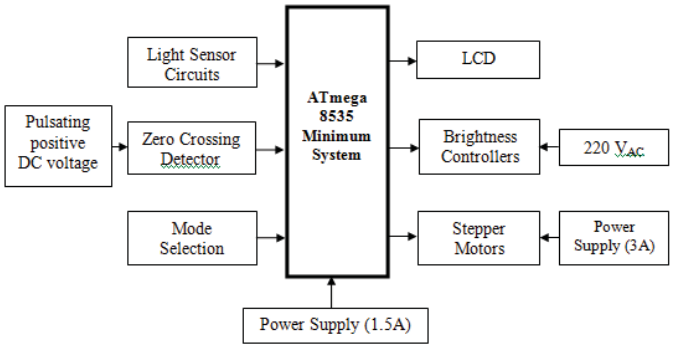


Fig. 2. Organization of the System Controller

Complete block diagram showing how hardware aspect of the system controller is organized depicted in Fig. 2.

There are two power supply units that provide 5 volts DC with different current ratings. A lower current rating power supply is used to supply all of the logic circuits while the higher one is specially used to supply the stepper motors only. Next, the mode selection button is used to choose either normal mode or presentation mode.

B. Light Sensor

Light dependent resistor (LDR) is used as the illumination measurement unit. It is built in combination with 10 kΩ resistor in a voltage divider configuration as shown on the Fig. 3.

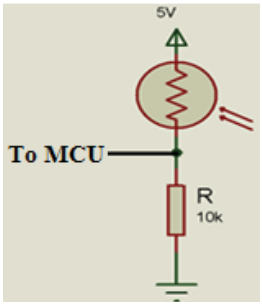


Fig. 3. LDR Configuration

The output of this measurement circuit is an analogue signal. Since the project utilizes ATmega8535 which is completed already with analogue to digital converter (ADC), then external ADC IC is not required. By setting up the components into voltage divider configuration, the circuit will directly obtain a voltage that varies with the measured values of light intensity, instead of a resistance change. By adapting the concept, then the output voltage of this configuration can be calculated using (4).

$$V_{out} = 5x \left(\frac{10x10^3}{LDR + (10x10^3)} \right) \quad (4)$$

C. Zero Crossing Detector

The zero crossing detector circuit is utilizing a step down transformer, five diodes 1N5392GP, optocoupler NPN, and two resistors. Practically, this circuit can be connected directly into the power supply circuit so it utilizes the same transformer and bridge rectifier circuit as the power supply used. Complete configuration of these components is shown on the Fig. 4.

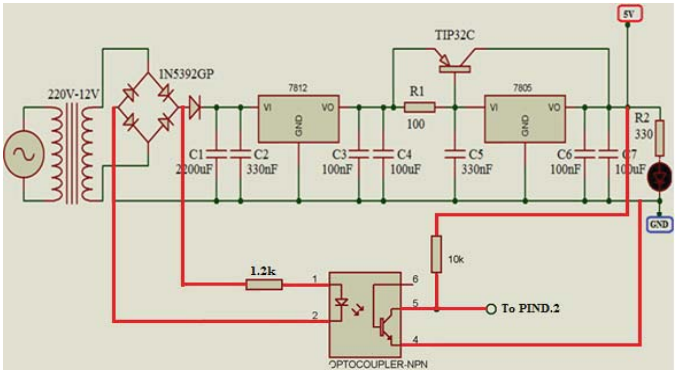


Fig. 4. Zero Crossing Detector Circuit

Firstly, the 12V alternating current is rectified into direct current by the bridge rectifier circuit. The output of this

rectifier is not pure direct current, but it is a pulsating positive DC voltage as shown on the waveform in the Fig. 5[3].

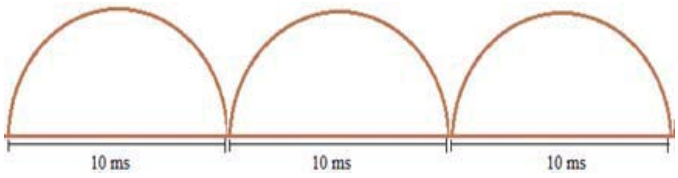


Fig. 5. Pulsating Positive Voltage

This pulsating positive voltage is used to drive the infrared LED inside the NPN optocoupler and a 1.2 kΩ resistor is used to limit the current through it. By doing this, the optocoupler LED will stay ON for most of the cycle except at the time when the AC sinusoidal wave crosses 0 V (or about to cross 0 V). When this LED is glowing, it will pull the external interrupt pin of the microcontroller low, in this case external interrupt 0 PIND.2 (EXT INT0). In contrary, when the LED is OFF, then the external interrupt PIND.2 will be pulled high. Theoretically, based on the working principle above, output of this zero crossing detector should look like the red waveform in the Fig. 6.

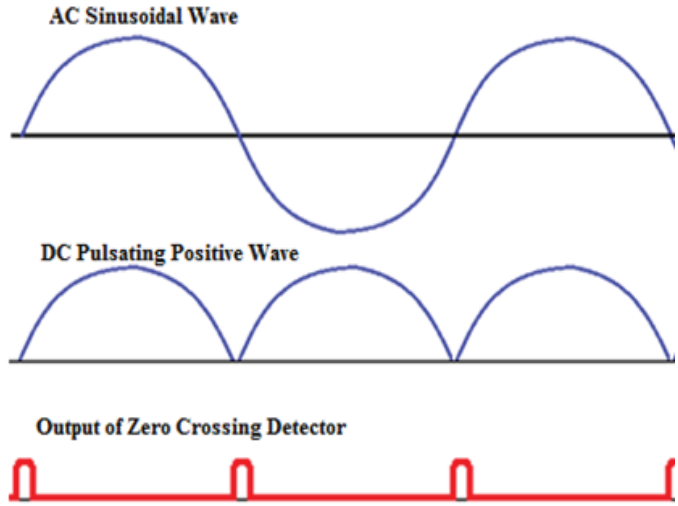


Fig. 6. Zero Crossing Output Signal (Theoretically)

By the presence of this interrupt signal into PIND.2, continuously the microcontroller will be interrupted every 10 ms and it will be able to recognize when the wave pass through the zero line. Next, this point will be used as reference point to start the triac firing delay precisely.

D. Brightness Controller Circuit

Brightness controller circuit has the responsibility to maintain the brightness of the lamp. This circuit consists of triac BT139 [2][7], optocoupler MOC3020 as a triac driver, and several current limiting resistors. Here MOC3020 is used also as a protection for the whole logic circuits if there is unwanted high voltage or high current occur in the load. By using this optocoupler, all of the logic circuit is completely separated from the AC load (lamp). A complete circuit of the brightness controller is shown on the Fig. 7.

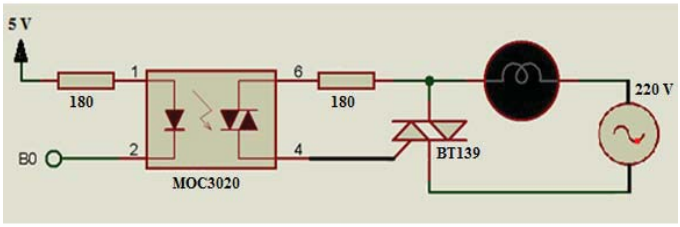


Fig. 7. Brightness Controller Circuit

Triac gating through the MOC3020 is done by giving a low pulse into optocoupler pin 2. Active low is chosen so that the microcontroller does not need to provide voltage for the circuit. This condition is generally known as sinking where the digital I/O only provides ground needed in the circuit. Sinking is chosen instead of sourcing because of the optocoupler current rating consideration. When a low pulse is given to optocoupler pin 2, then the bilateral switch inside MOC3020 will directly ON and triac is latched until it reaches the zero crossing point.

E. Stepper Motor Driver

Stepper motor driver is used as an interface between microcontroller with the stepper motor itself. The driver being used is dual H-bridge, high current, 16-pins L293D. In this final project, since the motor requires quite high current to operate than it is better to separate this motor driver with the other logic circuits. For this purpose TLP521 optocoupler is used to separate the motor driver with the microcontroller depicted in Fig. 8. Related with this, the power supply for the motor and the driver are also separated so that in case the load draws high current, then the other circuits will not be disturbed.

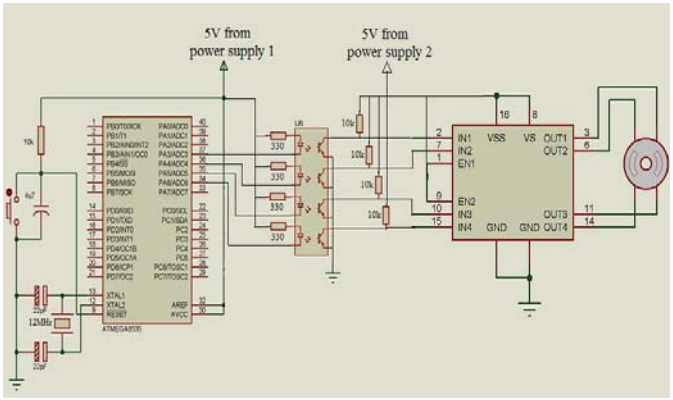


Fig. 8. Stepper Motor Configuration

The input signal of the driver can be simply set by following the stepper table[4].

TABLE I
STEPPER TABLE

	Pin 2	Pin 7	Pin 10	Pin 15
STEP 1	ON	OFF	ON	OFF
STEP 2	OFF	ON	ON	OFF
STEP 3	OFF	ON	OFF	ON
STEP 4	ON	OFF	OFF	ON

In order to change the direction of the rotation can be done by simply reverse the step[4]. It means step 4 become step 1, step 3 become step 2, step 2 become step 3, and step 1 become step 4. This reversed step order is shown in the table below.

TABLE II
REVERSED STEPPER TABLE

	Pin 2	Pin 7	Pin 10	Pin 15
STEP 1	ON	OFF	OFF	ON
STEP 2	OFF	ON	OFF	ON
STEP 3	OFF	ON	ON	OFF
STEP 4	ON	OFF	ON	OFF

IV. RESULT

A. Hardware Configuration

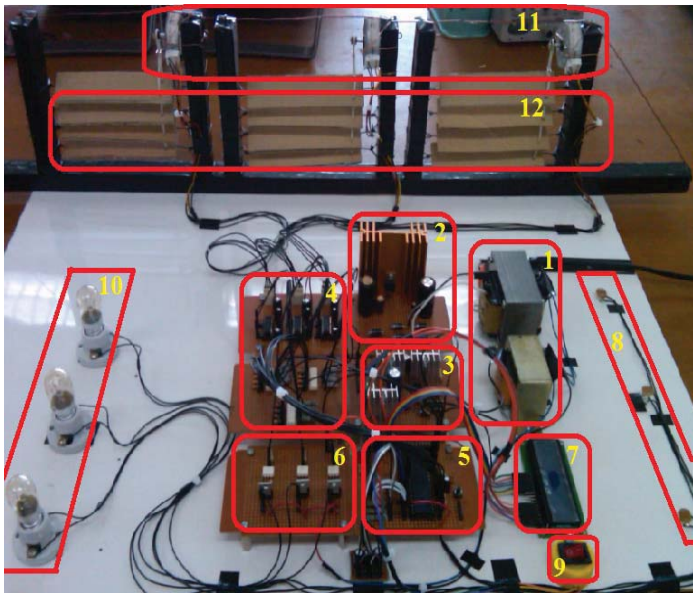


Fig. 9. Top View of System's Miniature

Miniaturization of the classroom, and the light intensity controller of the classroom are shown on Fig. 9. Depicting in the Fig. 9, the component of the miniature can be mentioned as follows.

1. *Step down transformers*
2. *High current power supply for the three stepper motors only*
3. *Power supply for all of the logic circuit*
4. *Stepper motor drivers configuration*
5. *ATmega8535 minimum system[6]*
6. *AC lamps brightness controller*
7. *Liquid Crystal Display (LCD)*
8. *Light Dependent Resistor (LDR)*
9. *Mode selection button*
10. *AC lamps*
11. *Bipolar stepper motors*
12. *Blinds*

The device has successfully operated to keep the light intensity in a constant level. In the normal mode, if the light intensity drops below the predetermined setpoint then the stepper motor opens the blind while the lamp stays OFF. If the blind has already opened and the light intensity still below the setpoint, the lamp will ON and get brighter until the setpoint is reached.

On the other hand, if light intensity is greater than the predetermined setpoint, the lamp will be dimmed until the setpoint is reached (in this case the lamp might be completely OFF). If the light intensity is still greater than the setpoint and the lamp is OFF, then the stepper motor will close the blind. Next, in the presentation mode the operating way is quite similar. The difference is only on the setpoint for the first lamp (the lamp that will be placed in front side of the room) while the rest are exactly the same.

B. Zero Crossing Detector

Output signal of the zero crossing detector looks like a repeated positive pulse with a fix time interval as shown on the Fig. 10.

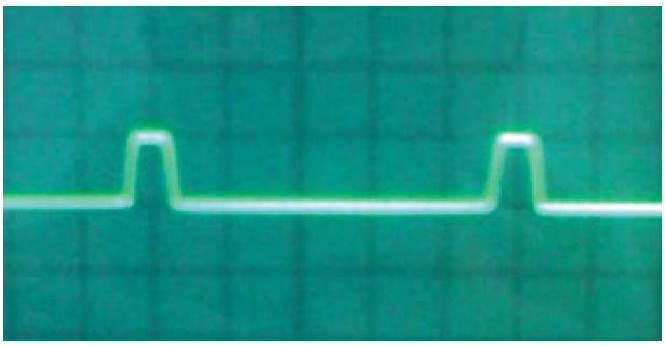


Fig. 10. Zero Crossing Output Signal (Practically)

Theoretically, zero crossing point must occur every 10 ms as explained in the previous section. However this is not that accurate in practical. The infrared LED of NPN optocoupler will goes off a moment before and after the AC sinusoidal wave crosses zero because the power is too small at this time interval. Because of this, the NPN optocoupler's output signal

will consume a small amount of time at every zero crossing point. Base on estimation, it is around ± 0.5 ms on each edge of the zero crossing point. By knowing it, then the firing pulse must not be generated 0.5 ms before and after the zero crossing point.

C. Brightness Controller Circuit

Brightness level of the lamp can be determined by chopping the AC sinusoidal wave that is fed into it. In order to know how much voltage is exactly delivered into the lamp, several testing are performed by varying the firing delay of the triac. Some of the results are shown on the Fig. 11, Fig. 12, Fig. 13, and Fig. 14.

1. Firing delay: 0 ms

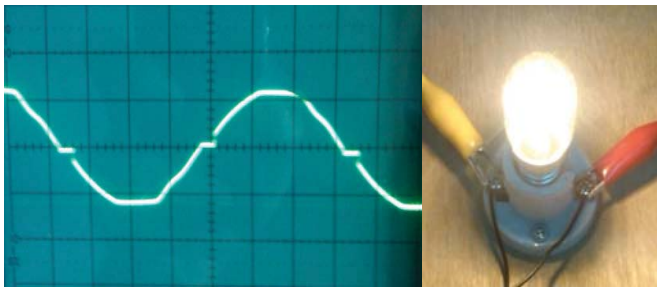


Fig. 11. Output of 0 ms firing Delay

The conduction time of the phase firing angle is much longer than the firing delay time. Hence, the total voltage delivered to the lamp is also much high which can be depicted in Fig. 11. When we decreased the conduction time of the phase firing angle which also means that the firing delay was being increased, the total voltage delivered to the lamp for one complete cycle of a sinusoidal wave became smaller shown on Fig. 12.

2. Firing delay: 2 ms

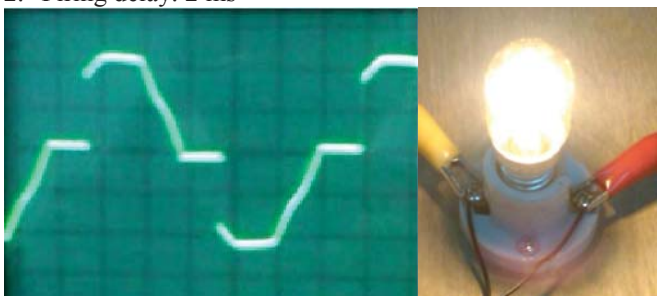


Fig. 12. Output of 2 ms firing Delay

Varying the conduction time of the phase firing angle was also possible when we decreased again the conduction time. Depicted in Fig. 13, firing delay increased as the conduction time of the phase firing angle had decreased, and this configuration causes the total voltage measured at the lamp to become much smaller.

We described also that period time of one complete cycle of AC signal generated by the State Electricity Company is 20 ms. Indeed, the period of a half cycle of the AC signal is 10 ms. When we increased the firing delay near to 10 ms, i.e. 7 ms, the conduction time of phase firing angle became smaller. In Fig. 14, we could depict that the total voltage delivered to the lamp was very small. This condition also means the intensity of the light emitted by the lamp became very low.

3. Firing delay: 5 ms

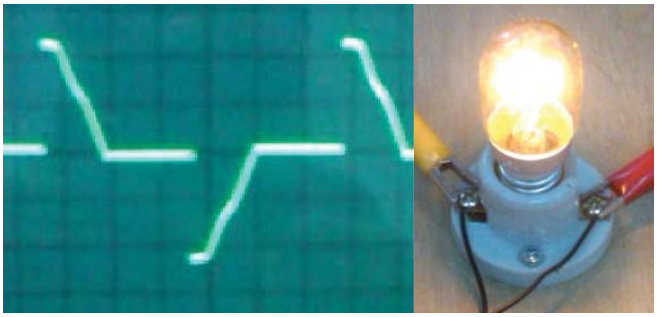


Fig. 13. Output of 5 ms firing Delay

4. Firing delay: 7 ms

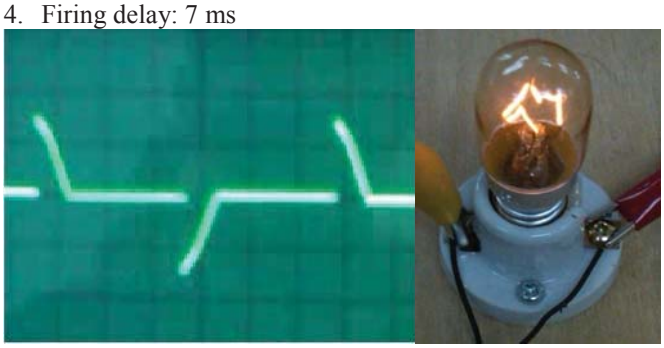


Fig. 14. Output of 7 ms firing Delay

The comparison between calculated load voltage (at the lamp) and measured load voltage (at the lamp) are shown in the table and on the graph below. Note that the load voltage measurements are done ten times for each conduction time. Therefore the measured values listed below are the average of the ten times measurements.

TABLE III
CALCULATED VS. MEASURED LOAD VOLTAGE

Conduction Time (ms)	Calculated Load Voltage (V)	Measured Load Voltage (V)	Error (%)
1	5.38	4.83	10.22
2	21	20.11	4.24
3	45.34	42.63	5.98
4	76	73.48	3.32
5	110	106.63	3.06
6	143.99	138.62	3.73
7	174.66	167.21	4.27
8	198.99	189.85	4.59
9	214.62	205.1	4.44
10	220	209.47	4.79

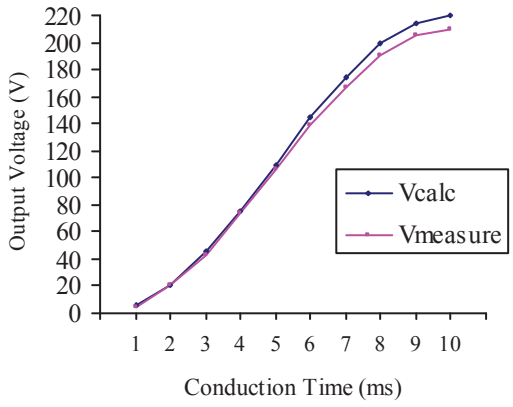


Fig. 15. Measured Voltage vs. Calculated Voltage

There are some insignificant errors in practical result. Several analysis have been done and found that the width of zero crossing pulses significantly affect the error of phase firing angle. As explained previously, the zero crossing pulses take about ± 0.5ms before and after zero crossing point. However this condition cannot be eliminated since it is belong to one of the zero crossing detector properties. By the existence of these pulses will prevent the triac firing to start exactly at the zero crossing points, but a moment after zero crossing points. Beside zero crossing pulses, there are also other factors that contribute in the error and they are:

1. Unstable or fluctuating source voltage
2. Misreading of the measurement tools
3. Precision of the measurement tools
4. Component tolerance

V. CONCLUSIONS

In conclusion, this work entitled "Blind and Lighting Control to Maintain Comfort Light Intensity of The Classroom Utilizing Microcontroller ATmega8535 has already successfully done its function as follows.

1. The light intensity sensing circuits has successfully detected the light intensity level and the microcontroller display it on the LCD in the unit of Lux.
2. The zero crossing detector circuit has successfully sent output signal that indicate the points when alternating current sinusoidal wave crosses zero (0) volt into the microcontroller.
3. The brightness controller circuit has successfully control the brightness level of the lamps.
4. All of the bipolar stepper motors have operated smoothly to open and close the blind and they are integrated well with the other circuit.

Implementing this device will automatically maintain a comfortable light intensity level inside the classroom both in normal mode and presentation mode. In addition, the student or the user can study conveniently without need to care of the outside light intensity level.

The development of this controller is one of development series using microcontrollers or microcomputers which have been done in purpose to solve several practical problems found in our daily life or occured in several industrial institutions by approaching with appropriate solutions based on hardware/software co-design-based applications. In summary, we are able to design and implement one of the solutions as answers of problems which we encountered, i.e. by designing "Blind and Lighting Control to Maintain Comfort Light Intensity of The Classroom Utilizing Microcontroller ATmega8535".

REFERENCES

[1] M. H. Rashid, Power Electronics Handbook (3rd edition). Oxford, UK : Elsevier, 2011.
[2] Thyristor Theory and Design Considerations. On Semiconductor™, 2006.
[3] T. Igoe, and D. O'Sullivan, Physical Computing: Sensing and Controlling the Physical World with Computers. Boston, MA: Thomson Course Technology PTR, 2004.
[4] http://www.seattlerobotics.org/encoder/may98/steppers.html.
[5] A. A. Birjandi, and Z. Ameli, "Three phase controlled rectifier study in terms of firing angle variations", ACEEE International Journal on Electrical and Power Engineering, Vol. 02, No. 02, May 2012, pp. 25-28.
[6] http://www.atmel.com/images/doc2502.pdf, retrieved on January 2013.
[7] F. Swainston. A Systems Approach to Programmable Controllers. Albany, New York: Delmar Publishers Inc., 1992, pp. 23.

Design of Capacitance Measurement Circuit for Data Acquisition System ECVT

Arbai Yusuf, Wahyu Widada, Warsito P. Taruno
CTECH Labs, the Center for Tomography Research Edwar Technology Co.
Tangerang, Indonesia
Email: {arbaibrwlpn, w_widada, wsito}@yahoo.com

***Abstract*— In this research, we proposed a capacitance measurement circuit Electrical Capacitance Volume Tomography (ECVT) to perform three-dimensional image visualization. The ECVT system is consists of three main parts i.e. sensor, data acquisition system, and computer. Data acquisition system is composed of capacitance measurement circuit and microcontroller to measure an unknown capacitance inside the sensor, collect data and send it to the computer. Further, these data is used to reconstruct 3D image. The design of the circuit used a sine wave 14.6 $V_{p-p}$and 2.5 MHz of frequency injected to the electrode pair to measure an unknown capacitance inside the sensor. An experiment is performed using simulated phantom using sensor having the form of a half-sphere with combined triangular and rectangular shapes. The system is able to measure a capacitance value as low as four femto-Farads with 0.34% margin error.**

Keywords—ECVT, Differential circuit, Capacitance to voltage circuit, Programmable gain amplifier, Standing capacitance.

I. INTRODUCTION

Electrical capacitance volume tomography (ECVT) is a volumetric tomography technique based on electrical capacitance measurement. This new system developed by Warsito et al, (IEEE Sensor Journal, 2007) has revolutionized and replaced the classical two-dimensional slicing technique of tomography imaging [1]. ECVT has been widely applied for real time volumetric imaging in industry such as imaging water and bubble, water and oil, soil and water. In the recent application, ECVT has been used in breast cancer scanning. In our previous research have developed a 32-channel data acquisition system ECVT with 5 volt square wave and 2.5 MHz frequency pulse injection. The system is capable to measure an unknown capacitance down to femto-farad with 8.3% of error and the accuracy is far above 5 femto farad. In the previous work, the system error is relatively high, so it cannot detect a small phantom, and is relatively less sensitive. This work is to modify the capacitance measuring circuit from the previous research, by changing the excitation voltage become sine wave 14.6 volts and 2.5 MHz.

This paper described a design of capacitance measurement circuit and calibration process to measure an unknown capacitance inside the sensor. The capacitances are being measured is very small (in the range of femto-Farads), hence it is necessary to use a reliable design of the capacitance measuring circuit. To do that, the capacitance to voltage circuit has to be modified. The aim of this excitation voltage change is to increase sensitivity and accuracy. Moreover, a digital to analog converter and programmable gain amplifier was added. Therefore, with the new circuit is expected to detect a phantom with relatively low permittivity value such as a small wet tissue, a small rubber ball inside paraffin, small plastic ball, etc. To validate the results is by comparing the shape of the phantom.

II. CAPACITANCE MEASUREMENT CIRCUIT

The capacitance measurement circuit is based on an AC-Based-circuit architecture where, it has been use widely many years on ECT systems and proved to be stray-immune. Like other capacitance tomography system, ECVT system is easily influenced by stray capacitance that can interfere the capacitance measurement process. The stray capacitances are mainly generated from (1) coaxial cable that connected from sensor to the data acquisition system (100 pF for one meter long); (2) CMOS switches that used to activate an electrode pair (8 pF); (3) The shielding of the sensor in the form of screen that surrounds the sensor (depend on the dimension of the sensor). Therefore, total of the stray capacitance are about 150 pF [2].

A. Capacitance to Voltage Circuit

The most importance of the data acquisition system is capacitance to voltage (C-V) circuit. The excitation voltage (V_{in}) is 14.6 V_{p-p} and 2.5 MHz of frequency injected to the excitation electrode. An operational amplifier (Op-Amp) with feedback R_f and C_f are used to convert it into current flow of sine wave output (V_o). With the current approach, the incoming current flow to the Op-Amp is equal to current flow out of the Op-Amp ($i_1 = i_2$), so following equation holds:

$$V_o = -{}^{C_x}/_{C_f} V_{in} - \omega C_x R_f V_{in} \quad (1)$$

Where ω is an angular frequency from the excitation voltage (V_{in}) that generated by sine wave signal generator. C_x is the

unknown capacitance to be measured. Figure 1 show the capacitance to voltage circuit.

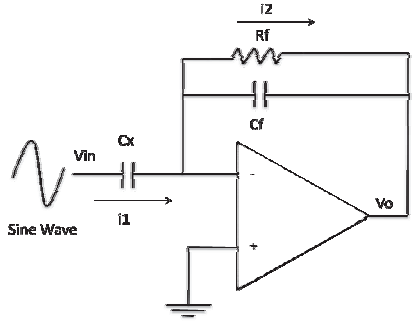


Fig 1. Capacitance to voltage circuit

The Op-Amp selections are based on the Op-Amp abilities in forwarding and amplify the signal, i.e. slew rate (S_r) and frequency-unity gain ($f_{unitygain}$) and can be calculated with the following equation:

$$Sr = 2\pi f V_o \quad (2)$$
$$= 2 * 3.14 * 2.5MHz * 0.5V$$
$$= 7.85\, V/\mu s$$

$$f_{unitygain} = f * gain \quad (3)$$
$$= 2.5\, MHz * 1$$
$$= 2.5\, MHz$$

With a 2.5 MHz of frequency and the desired output of the voltage for example is 0.5 volt, the minimum slew rate is calculated as 7.85 V/μs and the minimum $f_{unitygain}$ is 2.5 MHz. Therefore, the AD827 operational amplifier from analog device is selected, where the Op-Amp has 300 $V/\mu s$ of slew rate and 50 MHz of bandwidth.

As mentioned above, the stray capacitance is about 150 pF dependent on coaxial cable, switch, and sensor screening. With a good coaxial cable selections, good BNC connectors, as well as CMOS switches with low capacitances, and a good sensor design will be reduce the stray capacitance. Hence, the total of the stray capacitance for this circuit is not exceeding more 100 pF. Therefore, C_f of 100 pF was selected. The minimum resistance (R_f) can be calculated with the following equation:

$$R_f = {}^{1}/_{2\pi f C_f} \quad (4)$$
$$= \frac{1}{2 * 3.14 * 2.5 * 0.0001}$$
$$= 636.62\ \Omega$$

To obtain a good performance, the minimum resistance (R_f) is selected as 636.62 Ω. For this design, a 750 Ω of resistance R_f was selected.

B. Signal Generator

The capacitance measuring technique is using AC-Based circuit. The 14.6 V_{p-p} sine wave signal was made from a square wave signal with low pass filter and then amplified it. It uses a microcontroller Attiny2313 to produce a square wave signal. Having a '*clockout'* facility in one of port. Then the fuse bit namely '*ckout'* inside the microcontroller registers is activated. With a 2.5 MHz of crystal frequency it can generate a 5 volts square wave and 2.5 MHz of frequency at '*clockout*' pin [3].

Figure 2 shows a signal converter circuit from square wave to sine wave. We use a second order of low pass filter to produce a much smoother of sine wave signal. Using filterpro software from Texas Instrument, it can be determined the value of resistors and capacitors. With R_1=1K5, R_2=15K, C_1=27p, C_2=5p can produce -84.2° of phase, -0.02db of gain, and 100ns of time response.

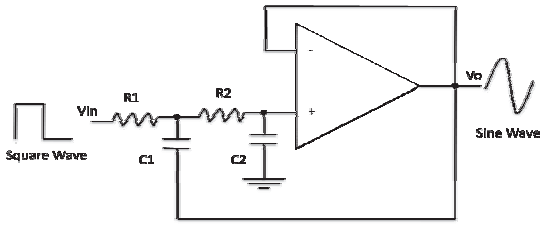


Fig 2. Square wave to sine wave circuit

To produce much higher voltage of sine wave up to 14.6 V_{p-p} we used an Op-Amp with gain 4x (Figure 3). By using an equation two and three, we obtained the slew rate minimum of 114.66 V/ μs and 10 MHz of frequency unity gain. The AD817 from analog device was chosen having 350 V/ μs of slew rate and 50 MHz of bandwidth.

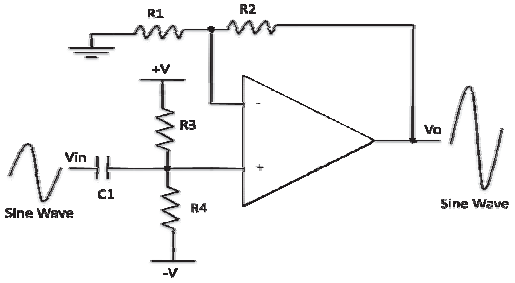


Fig 3. Gain for sine wave circuit

III. MEASUREMENT PROCEDURE

A. ECVT System Configuration

The ECVT system consists of three main parts shown in figure 4, i.e. (a) sensor, (b) data acquisition system, (c)

computer. The sensor has function to measure the unknown capacitance inside the medium. The data acquisition system has function to activate an electrode pair, collecting data, and send to the computer. Inside the computer was embedded a software and algorithm to reconstruct an image based on the value of capacitance. Inside the data acquisition system was installed two fan to reduce the heat generated by power supply and the signal generator.

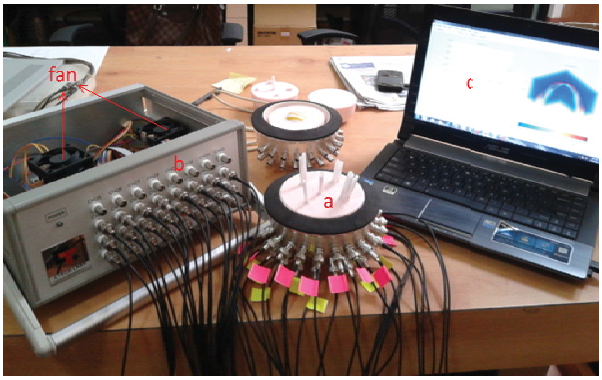


Fig 4. System Configuration of ECVT

B. Sensor Design

In this paper, we used a sensor for imaging of cancerous development inside breast. The 3D capacitance sensors constitute a half-sphere of sensing region to measure a whole permittivity of the breast. The sensor consists of 24-electrodes combine triangular and rectangular shape (Figure 5) [4]. The sensor divided into three levels, each level containing eight electrodes. The first and second levels are containing rectangular shape of electrode and the third level is containing triangular shape of electrode.

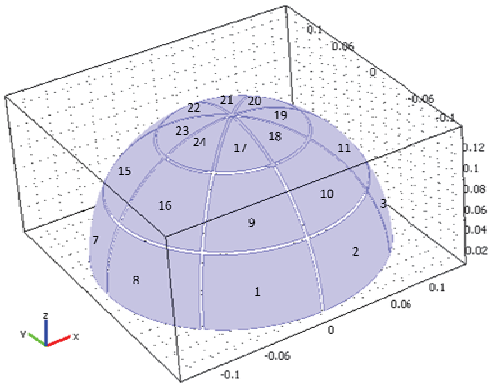


Fig 5. A proposed novel sensor for real time volumetric imaging of cancerous mass inside the breast (unit dimensions are in m) [Warsito et all].

C. Calculated the Standing Capacitance

Capacitance Cx was influenced by stray capacitance Cq, hence, capacitance to be measured is Cx+Cq, namely standing capacitance Cs [5]. The standing capacitance can be measured by connecting the first electrode to the ground so as the sine wave signal can't be injected to the electrode. Afterwards, the second electrode is activated to measure the capacitance Cq_1, and so on until the last electrode so as we earn all of stray capacitance Cq_n.

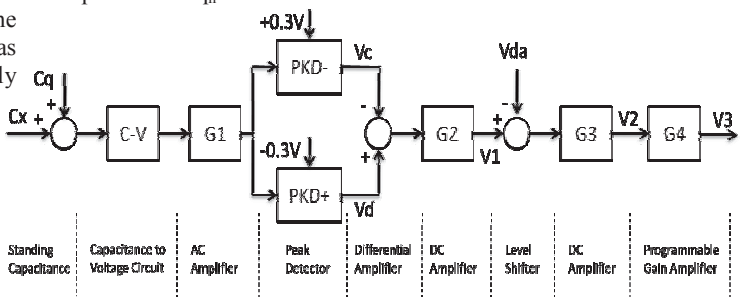


Fig 6. Blok diagram of capacitance measurement

Figure 6 is a block diagram of capacitance measurement procedure for one electrode pair. The unknown capacitance is converted by capacitance to voltage (C-V) circuit as mentioned above. Then, the result was amplified by G_1 to produce much bigger signal. The sine wave signal that has been injected to the electrode pair to produces two kind of signals, namely charge signal (Vc) and discharge signal (Vd). The two of peak detector circuits has function to capture signals amplitude. Output of the peak detector produces a dc signal that has voltage drop of 0.3 volt; that must be considered to calculate the capacitance value. Furthermore, Vc and Vd was summed by a differential amplifier to produce a positive signal, then amplified by G_2. Level shifter, G_3, and G_4 are used for adjust the dc signal into the region of analog to digital converter (adc) readings, where the region of adc readings is 0 to 4 volt. G_4 is a programmable gain amplifier where the amplify scale is 0 to 1. Equation 5 is used for calculated the standing capacitance.

$$C_x + C_q = \left({}^{V_3}/_{G_3G_4} + V_{da} + 0.6G_2\right)\left({}^{C_f}/_{2G_1G_2V_{in}} + {}^{1}/_{2G_1G_2\omega R_f V_{in}}\right) \quad (5)$$

Where V_{da} is dac voltage, C_f=100pf, R_f=750 Ω, V_{in} =14.6$V_{p\text{-}p}$, $\omega = 2\pi f$ where f=2.5MHz, and G_1=2, G_2=4, G_3=4, G_4 is gain of each data and has value between 0 to 1.

The ECVT system must be calibrated before used. The procedures of calibration are following this step:

(1). Power On, and wait until 30 minutes to warming up system.
(2). Empty calibration; set the empty voltage of 2000 mV. It used to calculate the dac voltages for each data.
(3). Full calibration; set the full voltage of 3000 mV. It used to calculate the gains and used to recalculate the empty voltages for each data.
(4). Running.

IV. EXPERIMENTAL RESULTS

Calibration procedure and calculating the standing capacitance are performed as described in section 3. Measurement results of the capacitances shows the values are far above Pico Farad, whereas the measured object is down to femto farad. Hence, the capacitance measuring data must be balanced.

Figure 7 show the two sets of 276 data of standing capacitance in a state empty and full with water. The reading of capacitances are obtained from the electrode pair 1-2, 1-3, .., 1-24, 2-3, 2-4, .., 2-24, 3-4, 3-5, 3-6, .., 3-24, etc up to 23-24. The red chart is value of standing capacitances in the empty state, which contains air with ε_r =1, and the reading of capacitances are from 7.1476 pf to 7.8926 pF. While the blue chart is value of standing capacitances in the full state, which contains water with ε_r =80, the readings of capacitances are from 7.3232 pf to 8.995 pf). In the chart shows the irregular values because the data is not yet in balance. One way is by making en empty value to zero; hence, we obtain a normal value as seen in figure 8. In this research, stray capacitance has not been calculated; hence, the results of measuring capacitances looks much higher. However, this results does not affect the tomography imaging, because the tomography process only need an absolute capacitance.

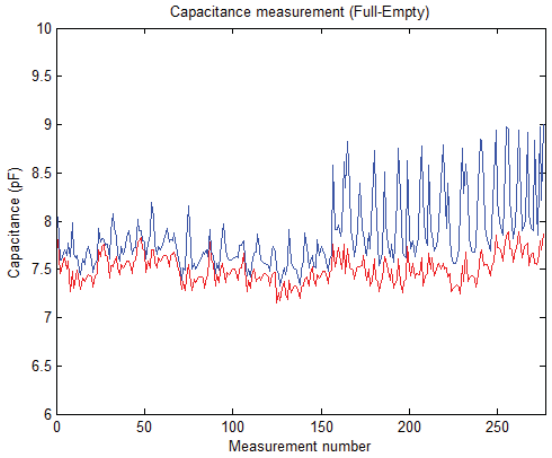


Fig 7. The 276 data of standing capacitance measurement; The red graph is empty ε_r =1; And the blue graph is fulfill with water ε_r=80

Figure 8 show the result of measuring capacitances after balanced with the sensor fulfill with water ε_r=80. The chart shows the regular values of capacitances in the range from 0.0407 pf to 1.2664 pf. The first electrode pair was produced a low capacitance, whereas, the third electrode pair was produce high capacitance. To reconstruct all of the data to become an image, it is necessary a 3D ECVT algorithm in other paper [1].

The next experiment is using a phantom made from silicon (ε_r=11). The straws, which contain some bone powders (ε_r=3) were inserted into the phantom (Figure 9). The results of capacitance measurement are -0.0038 pF to 0.1483 pF. Another experiment used a phantom made from paraffin (ε_r=2.2) and plastic ball (ε_r=80) which contains some water (Figure 10). The results of capacitance measurement are 0.0160 pF to 0.0984 pF. Figure 11 shows photograph of phantom used in the experiment.

The examples of images reconstruction from capacitance distribution using a human breast phantom can be seen in Figure 12. The charts in the left side shows both voltage measurement and absolute capacitance. While, the charts in the right side are shows result of the corresponding reconstruction images of the capacitances data. From the images shows that the object near the sensor is easyly to detect, whereas, the object in the middle of the sensor is not so. The red color shows the higher capacitance, while the blue color shows the lower capacitance. Small ball containing water can be reconstructed more easily than bone powder inside the straws. The measurement result can be validated by comparing the images reconstruction with the actual phantom.

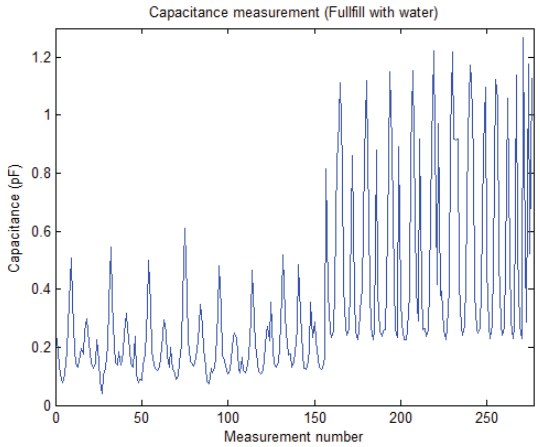


Fig 8. Capacitance measurement changes from empty ε_r =1 to fulfill with water ε_r=80

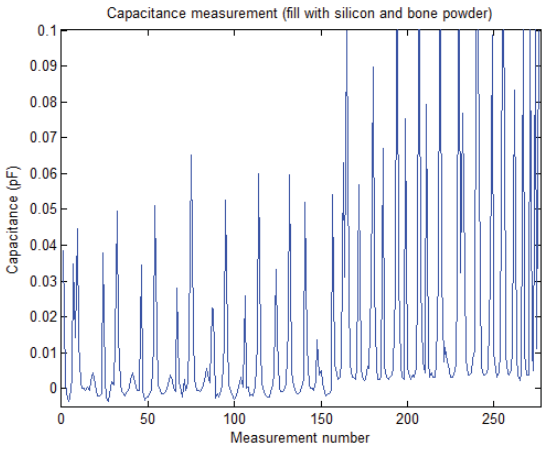


Fig 9. Capacitance measurement, fill with phantom made from silicon ε_r =11 and plastic straw fill with bone powder ε_r=3

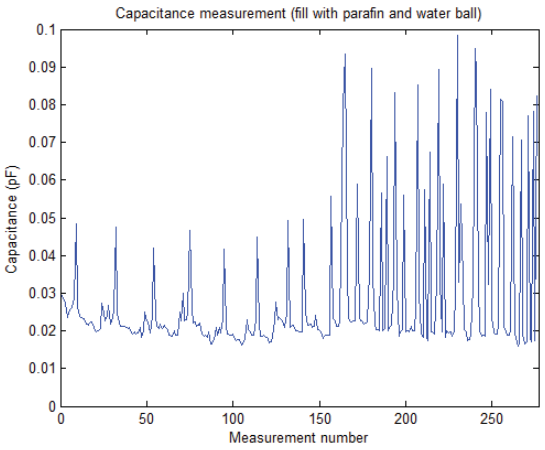
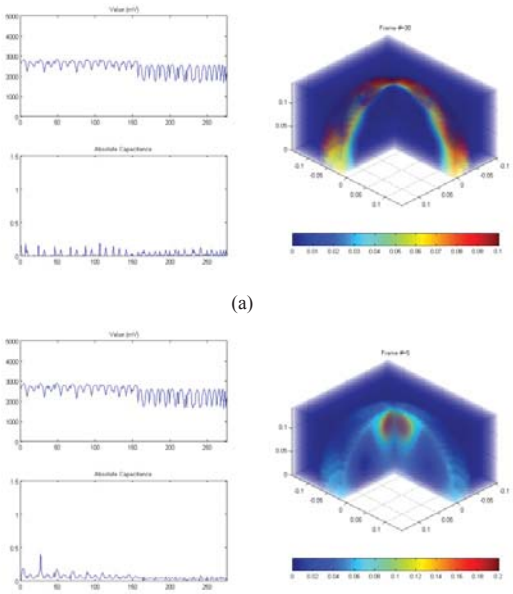


Fig 10. Capacitance measurement, fill with phantom made from paraffin ε_r =2.2 and plastic ball fill with water ε_r=80

Fig 11. Phantoms are used in experiment



(a)

(b)

Fig 12. The example of reconstruction image of phantom of human breast; (a) Phantom silicon and plastic straw fill with bone powder, (b) Phantom paraffin and plastic ball fill with water

V. CONCLUSION

The capacitance measurement circuit for electrical capacitance volume tomography (ECVT) has been designed and manufactured. The 14.6 V_{p-p} of sine wave signal and 2.5 MHz of frequency were used for injection to an electrode pair. Measurement results of the capacitance showed the values which were far above a Pico Farad to as low as few femto-Farads. The data capacitance must be balanced. By balancing the data, the system can be measure down to femto farad. The measurement result can be validated by comparing the images reconstruction with an actual phantom. This system has noise of 5 mV and the error of 0.34%. By considering the noise of the circuit, the system is capable of measuring capacitance with accuracy of 2.3 femto farad. Reconstruction results, showed that the images quality was improved with the new design.

ACKNOWLEDGMENT

Author thanks a lot to all researchers in CTECH Laboratories, Edwar Technology Co that have helped in data collections. (http://edwartechnology.com).

REFERENCES

[1] Warsito, Qussai Marashdeh, Liang-Shih Fan, "Electrical Capacitance Volume Tomography," IEEE Sensor Journal, Vol. 7, No. 4, pp. 525 – 535, April 2007.

[2] W Q Yang, "Hardware Design Of Electrical Capacitance Tomography Systems", *Meas. Sci. Technol*., vol. 7, pp. 225–232, 1996.

[3] Yusuf, Arbai., Widada, Wahyu., Warsito W., "Development of the Data Acquisition System for the 32-Channel Electrical Capacitance Volume Tomography", IEEE Tencon, pp: 1425 – 1428, November 2011, Bali, Indonesia.

[4] Taruno, Warsito P., Baidillah, Marlin R., Sulaiman, Rommy I., Yusuf, A., Widada, W., Alzufri, H., Aljohani, M., "A novel sensor design for breast cancer scanner based on electrical capacitance volume tomography (ECVT)". IEEE Sensor 2012 Conference, pp: 96 – 99. October 2012, Taipei, Taiwan.

[5] W Q Yang, "Charge Injection Compensation For Charge/Discharge Capacitance Measuring Circuits Used In Tomography Systems", *Meas. Sci. Technol*., vol. 7, pp. 1073–1078, 1996.

Novel Soft-switching Forward Converter based on Coupled Inductor

En-hui Chu
Dept.of Information Sci. and Eng.,
Northeastern University
Shenyang, China
chuenhui@mai.neu.edu.cn

Hong Wang
Dept.of Information Sci. and Eng.,
Northeastern University
Shenyang, China
wanghong6585@outlook.com

Yin-yin Wang
Dept.of Information Sci. and Eng
Northeastern University
Shenyang, China
wangyinyin988@163.com

***Abstract*—A interleaved two-transistor forward pulse-width modulation (PWM) DC-DC converter based on coupled inductor is presented. Compared with the conventional interleaved two-transistor forward PWM DC-DC converter, this converter can realize zero-voltage switching for the leading switches and approximately zero-current switching for the lagging switches by the secondary side auxiliary circuit, lossless snubber capacitors and transformer leakage inductance. So the voltage and current peak of the main switches and circulating current loss of the circuit can be effectively reduced, increasing the overall efficiency. The converter employs a simple auxiliary circuit that consists of neither lossy components nor active switches. The circulating current of auxiliary circuit can be adaptively regulated according to the load conditions. The voltage amplitude of holding capacitor, which can adjusts the commutation time, can be continuously regulated by varying the turns ratio of the coupled inductor. The change of the turns ratio not increases voltage stress of main switches and secondary rectifier diodes. This paper analyses the operation principle and soft switching implementation condition of new type converter according to the equivalent circuits in different operation modes. The effectiveness of the proposed converter was illustrated by the simulation results.**

***Index terms*—interleaved two-transistor forward converter; coupled inductor; pulse-width modulation; soft-switching; auxiliary circuit .**

I. INTRODUCTION

Forward converters have simple structure and high efficiency, so switching power supplies based on forward converter have been widely used in industry. However, the converters have an inherent defect that transformer must be magnetic reset, so it is necessary to use magnetic reset circuit. Commonly used magnetic reset methods [1]-[2] are: a reset winding, RCD clamp technique, LCD clamp circuit, and active clamp technique.

Compared with the forward converter, the two-transistor forward converter not only significantly reduces switch voltage stress, but also does not require special reset circuit to complete a reliable magnetic reset of the transformer. Two-transistor forward converter eliminates the shoot-through of full bridge converter [3] from the structure, and improves the reliability of the converter as well. In order to improve the converter efficiency, a novel zero voltage switching (ZVS) two-transistor forward converter [4] is proposed. The converter can achieve zero voltage switching of the main switches, reducing switching losses. Since the MOSFET has large conduction loss. In recent years, IGBT is widely used. However, because of the current tail characteristics of IGBT, it is more suitable used in zero current switching state. Thus, a novel zero current switching (ZCS) two-transistor forward converter [5] is proposed. The main switches achieve zero current switching, improving overall efficiency. However, the converter requires two auxiliary circuit control two main switches, so the circuit is complicated. Although the improved circuits reduce the switching losses, the existence of diodes in the transformer magnetic reset circuit makes the converter duty cycle is less than 0.5. The small duty cycle is an important limiting factor for converters' widely used.

Combination technology [6] is defined by the appropriate combination of converters. The combination technology can reserve converter advantages, meanwhile, overcome the converter disadvantages. The interleaved two-transistor forward converter reserves the advantages of the two-transistor forward converter, such as low switch voltage stress, no shoot-through and high reliability. The topology itself has the following advantages ① the frequency of the output filter inductor has doubled, reducing the volume of the filter inductor. ② Equivalent duty cycle is doubled, reducing the output voltage and current ripple. ③ the rectifier diodes voltage is reduced to half the peak voltage in the case of same output voltage. ④ each of the parallel branch flows through the smaller power. Not only eliminate converter "hot spots", but also reduce the difficulty of the thermal design. However, the combination technology does not solve the switch turn-off voltage and current spikes. A novel ZVZCS PWM two-transistor forward converter [7]-[8] is proposed. It can achieve soft switching and inhibit the voltage and current spikes of main switches by the tapped-inductor output filter, lossless snubber capacitor and transformer leakage inductance. However, the setting range of tapped ratio is narrow. When the tapped ratio exceeds a certain range will have a huge impact on the voltage and current stress of the switches. The linear relationship between the output voltage and the duty-cycle is getting worse as the increasing of the tapped ratio. So the complexity of closed loop control strategies will increase.

A interleaved two-transistor forward converter based on the output coupled inductor is proposed in this paper. This converter can realize zero voltage turn-on of leading leg switches and approximately zero current turn-off of lagging leg switches by the secondary side auxiliary circuit, lossless snubber capacitors and transformer leakage inductance. The overall efficiency of the converter is improved. The auxiliary circuit is simple and the circulating current can be adaptively regulated according to the load conditions, decreasing the conduction loss. The voltage amplitude of holding capacitor,

The work was supported by the Fundamental Research Funds for the Center Universities (N120404004).

which can adjusts the commutation time, can be continuously regulated by varying the turns ratio of the coupled inductor.

II. CIRCUIT TOPOLOGY AND OPERATION PRINCIPLE

A. Circuit description

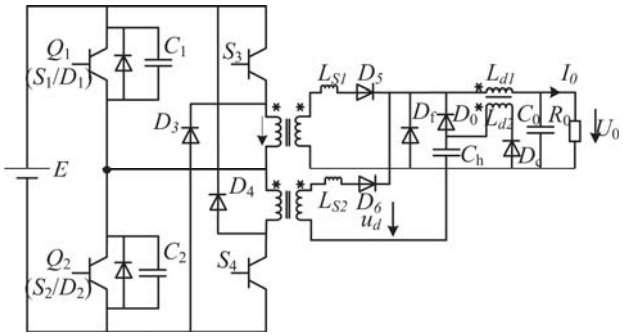


Fig.1. ZVZCS interleaved two-transistor forward converter

Fig 1 shows the circuit diagram of the proposed converter. One two-transistor forward converter consists of T_2、 S_1、 S_4、D_2 and D_4, the other consists of T_1、S_2、S_3、D_1 and D_3. The diode D_5、D_6 is rectifier diode for the two converters respectively. The diode D_1、D_2 and the snubber capacitors C_1 、C_2 parallel with S_1、S_2 respectively. The auxiliary circuit on the secondary side consists of holding capacitor C_h, diode D_f 、D_c 、D_0 and the output filter inductance coupled windings and output filter capacitance C_0.

B. Circuit operation

In Fig 2.v_{g1}~v_{g4} are respectively the driving waveforms of S_1~S_4, t_u is the phase-shift time between S_1 (S_2) and S_4 (S_3). S_1 (S_3) and S_2 (S_4) are complementary work with a short blanking time t_d.

In order to simplify the analysis of the converter, it is assumed that all the devices are ideal. The output filter inductor is large enough to be considered as a current source. The value of magnetizing inductances of transform T_1, T_2 is L_p (L_{P1}=L_{P2}=L_P). The value of leakage inductances of transform T_1, T_2, which is converted to the secondary side, is L_S (L_{S1}=L_{S2}=L_S).

The proposed converter has 10 modes in each half of a switching cycle. The operation modes are changed slightly at light load conditions, since the circulating current is changed according to the load conditions. As mentioned before, the converter employs the phase-shift control that all switches operate with duty-cycle of D, D=2(t_{on}-t_u)/T. The phase-shift time is t_u. The key waveforms and the equivalent circuit of each operation mode are respectively shown in Fig. 2 and 3.

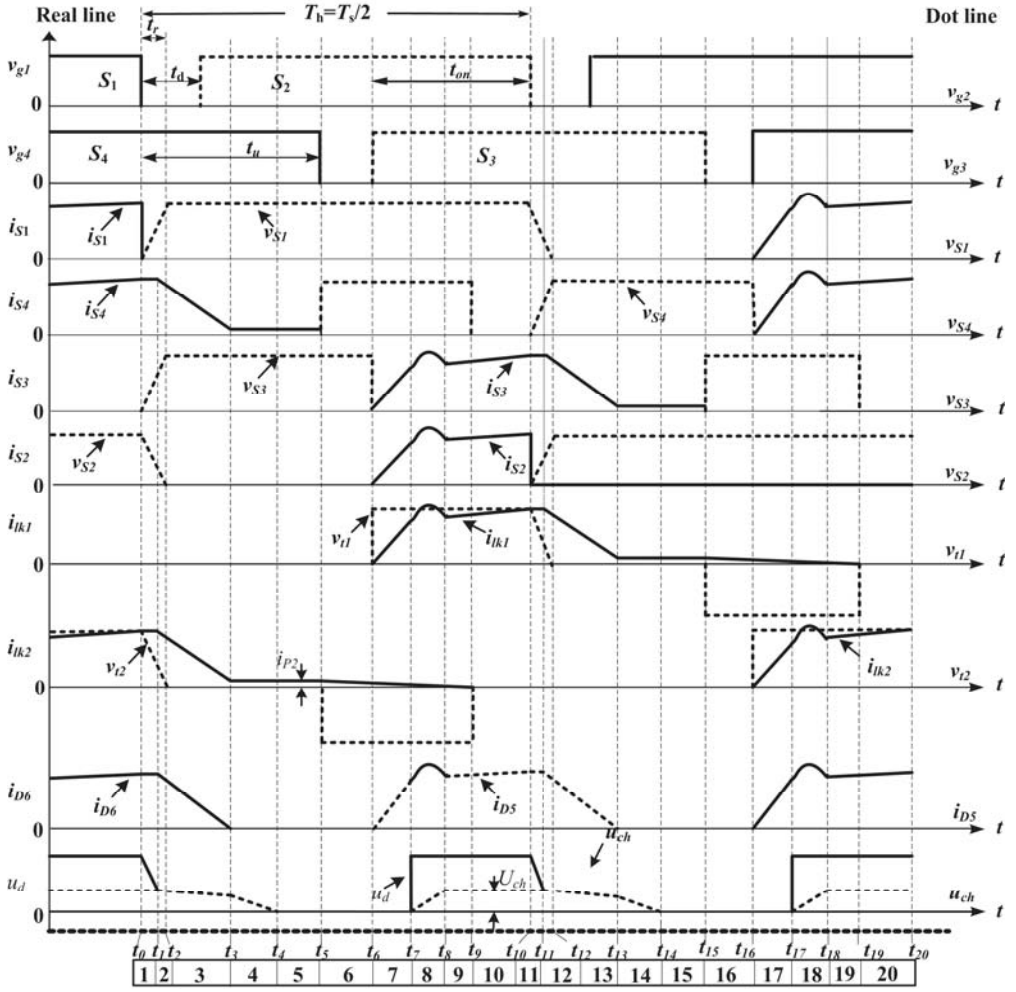


Fig.2. Key waveforms of the proposed converter

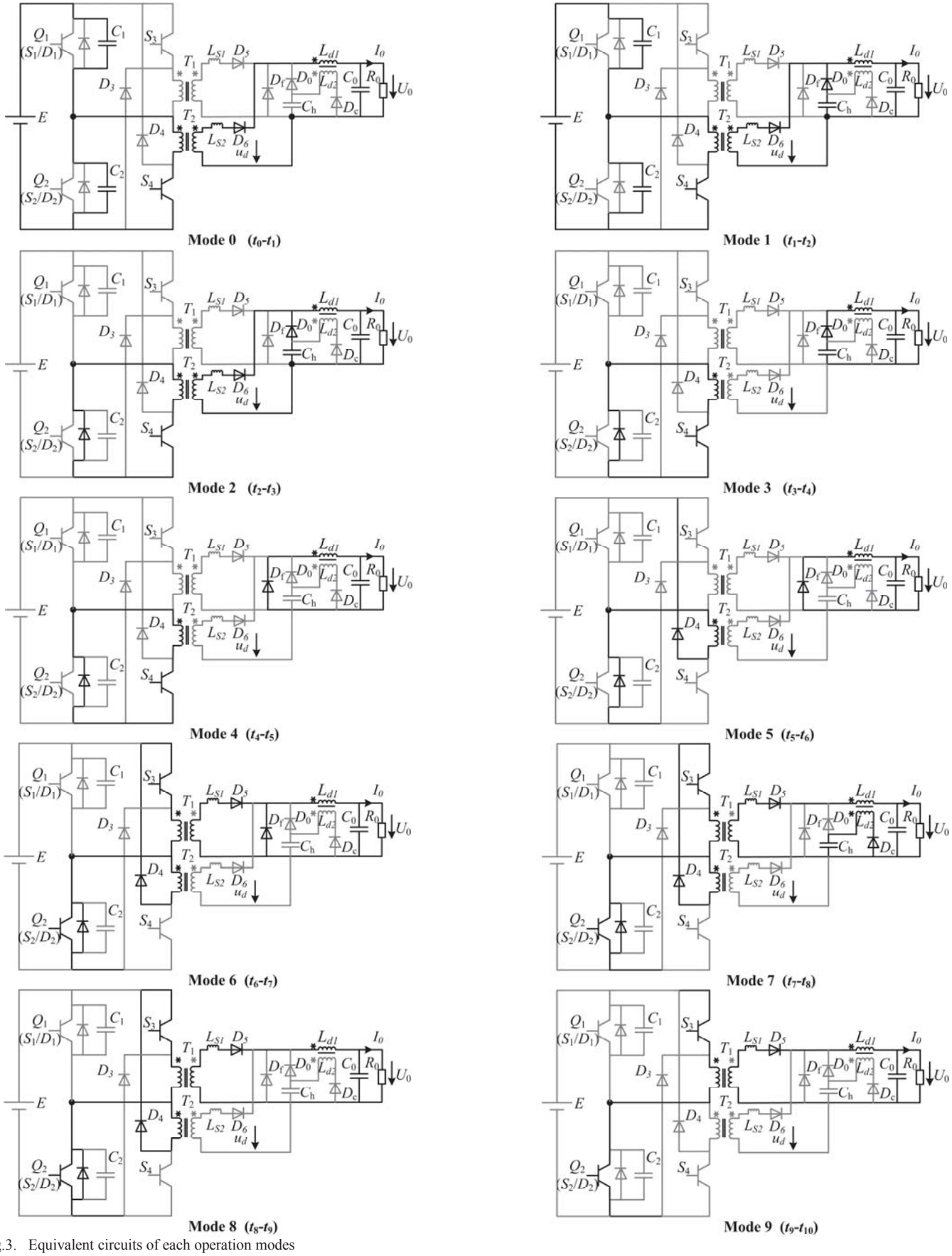


Fig.3. Equivalent circuits of each operation modes

1) Mode 0 [t_0～t_1]: At t_0, S_1 turns off. The primary current charges the snubber capacitor C_1 and discharges subber capacitor C_2. Due to the magnetizing inductor of transformer and the output filter inductor L_{d1} are large enough, i_{d6} and i_{p2} are considered as constant. The primary side voltage of the transformer T_2 decreases linearity as

$$u_{t2}(t) = E - \frac{i_{lk2}(t_0)}{2C} \tag{1}$$

Where $i_{lk2}(t_0) = i_{p2}(t_0) + i_{d6}(t_0)/N_T$ (2)

When the secondary side voltage of the transformer T_2, u_d reduces to U_{ch}, the mode ends.

2) Mode 1 [t_1～t_2]: At t_1, the diode D_0 turns on and u_d is clamped at U_{ch}. $i_{p2}(t_0)$ that is the maximum magnetizing current of the transformer remains constant in this mode. The voltage difference between the primary and the secondary of transformer is applied to the leakage inductance. The current flow through the rectifier diode and the voltage of the snubber capacitor C_2 decrease at the same time. The mode ends when the snubber capacitor reduces to zero.

3) Mode 2 [t_2～t_3]: At t_2, the voltage of C_1 rises to E and the C_2 is completely discharged. The anti-parallel diode D_2 begins to conduct and clamps the voltage of S_2 at zero. S_2 can realize ZVS in the period of D_2 conduction. The interaction of holding capacitor and leakage inductance of transform decreases the current through D_6 more quickly. At t_3, the current through D_6 decreases to zero. Only small magnetizing current of the transformer T_2 circulates through S_4 and D_2. Because i_{p2} is very small, only a little circulating current loss is produced.

4) Mode 3 [t_3～t_4]: At t_3, the current through rectifier diode decreases to zero. The holding capacitor is discharged through D_0, supplying the whole load current. The voltage of transformer secondary side decreases until it reaches zero as

$$u_d = U_{ch} \cdot \cos w_s (t_3 - t_2) - \frac{I_0}{c_h}(t - t_3) \tag{3}$$

Where $w_s = 1/\sqrt{L_S C_h}$ (4)

5) Mode 4 [t_4 ～ t_5]: At t_4, the holding capacitor is completely discharged and the whole load current freewheels through the freewheeling diode, D_f. At t_5, only small magnetizing current of the transformer T_2 circulates through S_4 and D_2, which permits the lagging leg switch S_4 to turn off with approximately zero current.

6) Mode 5 [t_5～t_6]: At t_5, S_4 turns off and the diode D_4 conducts. The small magnetizing current of the primary of transformer feedback to the power supply through diode D_4 and D_2, beginning the magnetic reset of the transformer T_2.

7) Mode 6 [t_6～t_7]: At t_6, S_3 turns on and meanwhile D_5 turns on. The current through S_3 increases linearity, while the current through freewheeling diode decreases until it reduces to zero.

8) Mode 7 [t_7～t_8]: Since the freewheeling diode D_f turns off naturally, the whole load current through the rectifier diode D_5. In this mode, S_2 and S_3 are conducting and the energy of primary side is transferred to the secondary. The holding capacitor is charged by resonance between the holding capacitor and the leakage inductance of the auxiliary winding.

When the voltage of holding capacitor rises to and remains in U_{ch}, which must be lower than power supply conversion to the secondary. The resonating current decreases to zero. The diode D_c turns off naturally and the mode ends.

9) Mode 8 [t_8～t_9]: The energy of primary side is still transferred to the secondary. The current i_{p2} decreases to zero at t_8, turning off D_4 and completing the magnetic reset of transformer T_2.

10) Mode 9 [t_9~t_{10}]: In this mode, S_2, S_3 and D_5 conduction and the energy is transferred to the load through T_1. At t_{10}, S_2 turns off and the half working stages of a switching cycle are completed. Due to the symmetrical configuration of the proposed converter, the analysis about the next half working stages is omitted.

C. ZVS conditions for leading leg switches

The time for snubber capacitor C_1 was charged from zero to E is t_r, and it must be shorter than the dead time between S_1 and S2, ensuring the ZVS for the leading leg switches.

$$t_d > t_r = t_{m1} + t_{m2} \tag{5}$$

$$\text{Where } t_{m1} = t_1 - t_0 = \frac{2C(E - U_{ch}N_T)}{i_{lk2}(t_1)N_T} \tag{6}$$

$$t_{m2} = t_2 - t_1 = \frac{1}{w_a}\sin^{-1}\frac{U_{ch} \cdot N_T}{i_{lk2}(t_1)\sqrt{\frac{L_S}{2C}}} \tag{7}$$

D. ZCS conditions for the lagging leg switches

In order to reduce the current flow through rectifier diode to zero rapidly, which permits the lagging leg switches to turnoff with approximately zero current, the energy of the holding capacitor should be large enough. The primary current should reduce to magnetizing current of transformer before the turn-off of the lagging leg switches. The relationship is obtained as

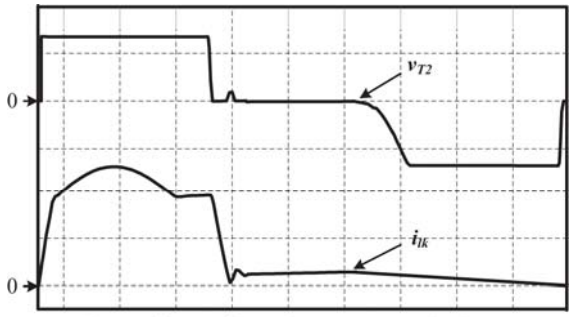
$$\frac{1}{w_s}\sin^{-1}\frac{N_T Z_S\left[i_{lk2}\left(\frac{t_{m2}}{N_T\sqrt{2L_S C}}\right) - i_{p2}(t_1)\right]}{U_{ch}} < (1-D)\frac{T_S}{2} \tag{8}$$

We can draw a conclusion from above that the turn off of the lagging leg switches can be realized by increasing the value of C_h and U_{ch}, which reduces the loss of energy. The U_{ch} must be lower than E/N_T, meanwhile, larger C_h increases the circulating energy of the secondary side. The selection of component parameter should according to actual need.

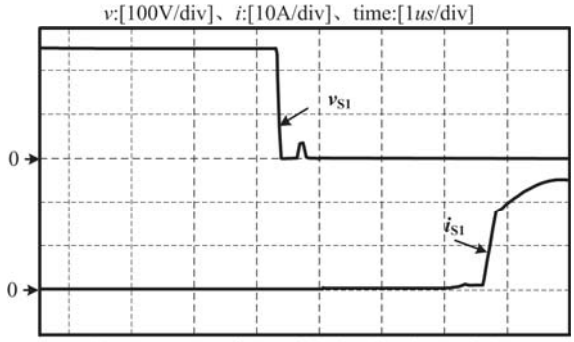
E. The couple output inductor

In order to ensure the soft turn-off of the diode D_c of the auxiliary circuit, the leakage inductance of the coupling inductor should be as

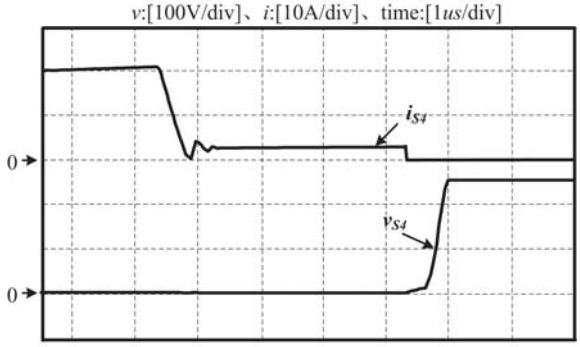
$$L_{ks} < \left(\frac{T_S D_{\min}}{4\pi}\right)^2 \frac{1}{C_h} \tag{9}$$



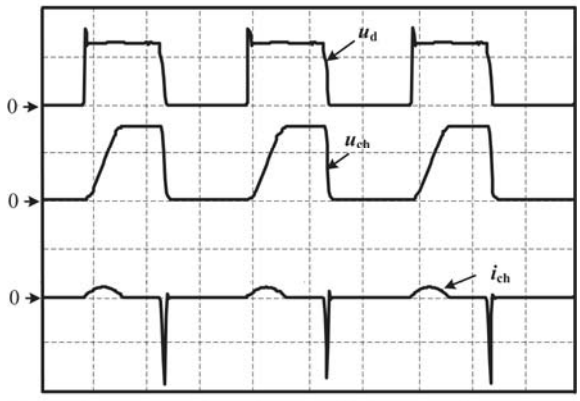
(a)Waveforms of the primary current and voltage.
v:[100V/div]、*i*:[10A/div]、time:[1*us*/div]



(b)Turn-on waveforms of the leading-leg switch, S_1.
v:[100V/div]、*i*:[10A/div]、time:[1*us*/div]



(c)Turn-off waveforms of the lagging-leg switch, S_4.
i:[10A/div]、*v*:[100V/div]、time:[0.5*us*/div]



(d)Transformer secondary voltage and holding capacitor voltage
v:[100V/div]、*v*:[50V/div]、*i*:[10A/div]、time:[1.5*us*/div]

Fig .4. Key waveforms of the simulation

III. SIMULATION AND EXPERIMENTAL RESULT

To verify the effectiveness of the proposed converter, analyze the simulation results of the proposed converter. Rated output power of the converter is 500w. Input voltage is 260V and switching frequency is 100 *kHz*. The turns ratio of transformer is N_T=2. The magnetizing inductance and leakage inductance converted to the secondary side of the transforms are respectively 3*mH* and 0.5μH. The turns ratio of the output coupling inductor is m=2 and the coupling factor is 0.95. The value of the output filter inductor is 150μH. The value of the holding capacitor is 0.47μF. t_d=0.5μS is the dead-time between S_1 and S_2 and t_u=0.7μS is the phase-shift time between S_1 and S_4. The load resistance is 1.8 *ohm*. The main simulation waveforms of the proposed converter are shown in Figs.4. The Fig.4 (a) shows the primary voltage and current waveforms of the transformer T_2. S_1 turns off and the primary current of transformer decreases quickly to small magnetizing current, permitting the lagging leg switch S_4 to turn off with approximately zero current. After S_4 turns off**,** the small magnetizing current of the primary of transformer feedback to the power supply through diode D_4 and D_2, completing the magnetic reset of the transformer.

The Fig.4 (b) shows the voltage and current waveforms of the leading leg switch S_1. After S_1 turns off, because the output inductor L_{d1} is large enough, so the current i_{d6} is regarded as constant. The current i_{d6} charge C_1 and discharge C_2, realizing zero voltage switching of the leading switches.

The Fig.4 (c) shows the voltage and current waveforms of the lagging leg switch S_4. When the voltage of the primary of transformer is reduced to U_{ch}/N_T, the diode D_0 turns on. The difference between the transformer secondary voltage and the primary voltage is applied to the leakage inductance. So the transformer primary current is rapidly reduced to magnetizing current. S_4 turns off with approximately zero current.

The Fig.4 (d) shows the voltage and current waveforms of the holding capacitor and the voltage waveform of the transformer secondary. When the voltage of holding capacitor rises to and remains in U_{ch}, which is kept below E/N_T, the resonating current decreases to zero, *D*c is softly commutated.

When u_d decreases to U_{ch}, D_0 turns on and the holding capacitor is discharged. The current of the holding capacitor ensure that the current through the primary of the transformer reduced to magnetizing current during the freewheeling period.

IV. CONCLUSION

This paper proposed a novel ZVZCS interleaved two-transistor forward converter. This converter can realize zero-voltage switching for the leading switches and approximately zero-current switching for the lagging switches by the secondary side auxiliary circuit, lossless snubber capacitors and transformer leakage inductance. In theory, if the transformers are assumed ideal, the lagging leg switches turn off with zero current, but in actually using of the transformer the ideal transformer not exist. The design of the transformer has an important influence on the performance of the converter, therefore, this paper consider the non-ideal condition. The auxiliary circuit is simple and consists of neither lossy components nor active switches, not increasing voltage stress of the rectifier diode. During the freewheeling period, the holding capacitor supply the load current and enable the lagging leg switch turn off with approximately zero current. During auxiliary circuit resonant charging process, when the holding capacitor is charged to U_{ch}, the diode D_c is soft turn-off.

REFERENCES

[1] F. D. Tan, "The forward converter: from the classic to the contemporary," Proc. IEEE Applied Power Electronics conf (APEC). 2002. vol. 1, pp. 857-863.

[2] T. B. René, and B. Ivo, "A Double ZVS-PWM Active-Clamping Forward Converter: Analysis, Design, and Experimentation" IEEE Trans, Power Electronics, Vol.16, No.6, pp.745-751, Nov.2001.

[3] J. G. Cho , J. W. Baek, D. W. Yoo, H. S. Lee and G. H. Rim, "Novel zero-voltage and zero-current-switching (ZVZCS) full bridge PWM converter using transformer auxiliary winding," IEEE Trans. Power Electron. vol. 15, pp. 250–257, Mar. 2000.

[4] K. B. Park, G. W. Moon, M. J. Youn, "Two-switch active-clamp forward converter with one clamp diode and delayed turnoff gate Signal," IEEE Trans, Industrial Electronics, Vol.58, No.10, pp.4768-4772, Oct.2011.

[5] J. D. S. Oliveira, C. A. Jr. Bissochi, J. B. Jr. Vieira, V.J. Farias, L.C. de Freitas, "A lossless commutation PWM two level forward converter operating like a full-bridge," APEC 2000,Vol.1, pp.334-339,2000.

[6] M. T. Zhang, M. M. Jvanovic, F. C. Lee, "Analysis and evalution of interleaving techniques in forward converters" IEEE, Trans, Power Electron. vol.13, No.4, pp.690–698, 1998.

[7] S. Hamada, S. Moisseev, O. Kondriavtsev and M. Nakaoka, "Double two switch forward soft switching PWM DC-DC converter." Transaction of the institute of Electrical Engineer of Japan, Part. D, Vol.123, No.2 pp.21-30, Feb. 2003.

[8] S. Moisseev, S. Hamada and M. Nakaoka, "Double two-switch forward transformer linked soft-switching PWM DC-DC power converter using IGBTs." IEEE Pro, Electric Power application, Vol.150, No.1 pp.31-38,Jan 2003.

Performance Comparison of Asymmetric Drain/Source Topology in Nanoscale Double Gate Vertical MOSFET

Munawar A Riyadi
Department of Electrical Engineering
Diponegoro University
Semarang 50275, Indonesia
munawar.riyadi@ieee.org

***Abstract*—Double Gate MOSFET structure is a promising architecture for advanced devices in nanometer regime. This paper elaborates the asymmetric topology of Vertical Double Gate MOSFET (VDGM) with ORI method as source/drain fabricating technique using numerical analysis approach. The electrical characteristics of the drain-on-top (DOT) and source-on-top (SOT) topology were analyzed, especially in the sub-threshold performance, to observe the short channel effect (SCE) of the device. The result shows that silicon pillar thickness reduction enhance the DIBL performance, while the threshold voltage roll-off change in nearly the same degree with the thickness variation. The floating body effect will likely occur for thicker silicon pillar in SOT, as the drain's depletion layer creates deeper barrier between substrate and pillar region. The performance comparison of sub-threshold slope revealed better SCE control for DOT topology in the lower silicon thickness for short channel length up to 30 nm.**

Keywords— vertical MOSFET, double gate, asymmetric source/drain, nanoscale device

I. INTRODUCTION

The recent development of micro/nanoelectronic devices shows a tendency for adoption of novel structures and new materials in the commercial chip products by the major players [1-2]. Due to the concerns over several limitations possessed by conventional planar structure for further device downscaling, multiple gate MOSFET has especially gained popularity for scaling the device into nanometer regime. This is also supported by the International Technology Roadmap for Semiconductor (ITRS) [3]. Multiple gate structure is able to maintain further scaling by flexible threshold voltage adjustment, with more stringent control on the channel, added by the possibility of lower doping profile [4-6].

The concern of increasingly complicated lithography process for device resolution in nanometer regime leads to the development of Vertical MOSFET structure as a viable alternative of double/multiple gate architecture [6-12]. The vertical structure introduces an opportunity in making lithography-relaxed gates and its respective channel for nanoscale dimension, thus enabling the prolonging of device scaling with the compatibility to existing technology. It is easier to fabricate double/surrounding gate at the sidewall of the silicon pillar which results in increasing current drive [13]. Moreover, this vertical structure also offers gate-to-gate self-alignment feature in double gate processing which is difficult to obtain in its planar structure counterpart. Furthermore, the introduction of oblique rotating ion-implanted (ORI) method in vertical MOSFET has become a great opportunity to improve short channel effect (SCE) control [14-15]. With this method, it is possible to have direct vertical current direction in the channel area from drain to source as the result of the bottom region reaching the corner side of pillar's bottom. This feature is of a great advantage in applying ORI method. With conventional implantation technique it is likely that the channel area become L-shape (with vertical and horizontal part near the bottom corner of pillar), as a consequence that the implanted dopant in source region cannot reach the pillar's corner at the bottom. This structure suffers from mechanical stress at the corner and may decrease the mobility of carrier, thus reducing the performance of the device.

The vertical MOSFET structure implies that the drain and source may not lie in the same horizontal plane, as does in planar structure. Instead, the drain and source are located in the top and bottom directions, although it is not necessarily sandwiched to each other. The Drain on Top (DOT) topology is more popular in the vertical double gate structures instead of source-on-top (SOT), due to the common practice of connecting the source and body terminals ($V_{SB} = 0$). However, given that the ORI method may result in the floating body of channel for smaller pillar thickness [16], it is less necessary to maintain the DOT topology. In addition, the use of SOT may be required for some applications due to reducing wiring design complexity reason.

This paper elaborates the SOT topology of Vertical Dual Gate MOSFET (VDGM) with ORI method as source/drain fabricating technique using numerical analysis approach. The analysis of the SOT topology compared to the DOT was based on the short channel effect, especially on the various pillar thicknesses and on the reduction of channel length. The range of the channel length (L_{ch}) was taken to be 30-100nm, whilst L_{ch} < 30nm was not selected because of the concern over the quantum effect that will eventually takes prominence in that

dimension. The device's electrical characteristic and its respective subthreshold behaviour are also elaborated to understand the device performances, especially for the short channel effect.

II. Device Simulation

The vertical double gate MOSFET structure with symmetrical self-aligned source/drain region was simulated using Silvaco TCAD tools. The process flow for the device is shown in Fig. 1. A <100> silicon wafer with uniform boron doping of 1.10^{19} cm-3 was selected as the base substrate. This relatively high substrate doping has the benefit of suppressing short channel effect [17].

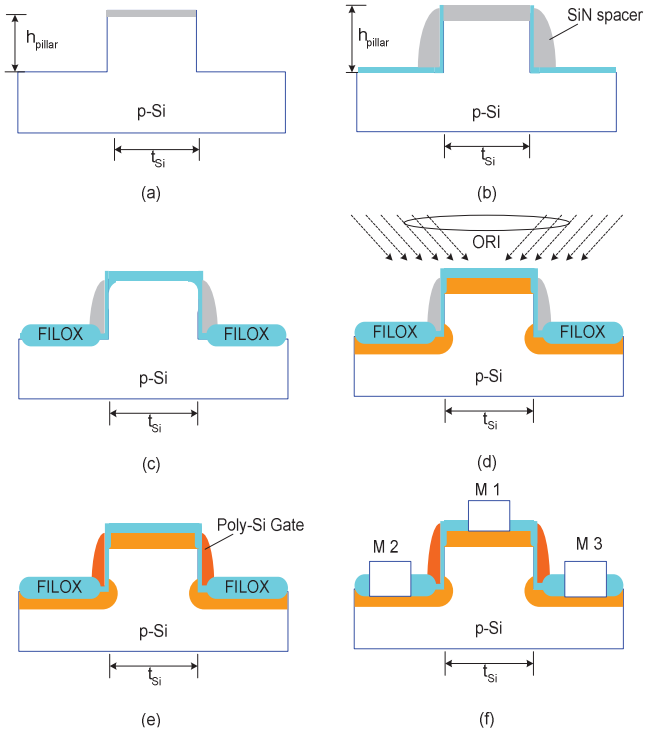


Fig. 1. Process flow for the vertical double gate fabrication: (a) pillar definition, (b) nitride spacer, (c) FILOX formation, (d) source/drain implant using ORI (45^0, 180^0 rotation), (e) poly gate formation, and (f) metal contact.

The silicon pillar was formed by dry etching the substrate with pillar thickness t_{si} and height hpillar which later determines the channel length definition, as shown in Fig. 1 (a). Stress relief oxide was thermally grown followed by the deposition of nitride layer, which was subsequently dry-etched anisotropically to define the active area (Fig. 1 (b)). After that the fillet local oxidation (FILOX) process [18] was done in the whole active area and on the top of the pillar for reducing overlap capacitance (Fig. 1 (c)). The self-aligned source and drain region were constructed by arsenic implantation (6.10^{15} cm^2, 150 keV) using Oblique Rotating Implantation (ORI) method [14] (Fig. 1 (d)). Later, a 3-nm silicon oxide layer was grown as a gate dielectric on the pillar sidewall, followed by polysilicon deposition (in-situ As doping) for gate electrode. Polysilicon spacer was then dry-etched to form double gate structure with self-aligned features (Fig. 1 (e)). After deposition of LTO for isolation, rapid thermal annealing (RTA, 1100 ^{0}C, 10 s) was carried out for dopant activation, followed with the metal contact process at gate, source and drain (Fig. 1 (f)). All the above process steps are similar for DOT and SOT, except for the contact wiring at the electrodes. In DOT, the drain terminal is M1 and the source is M2 and M3, while in SOT, the terminals are interchanged.

The electrical characteristics of the device were obtained by simulating the final structure using Silvaco's Atlas software package[19]. The Drift-Diffusion (DD) transport model was used, as it is able to predict the I-V characteristic of DG MOSFET [20]. It was also combined with Lombardi CVT model [21] where its semi-empirical equation gave the complete correlation between carrier concentration, carrier mobility, electric field and temperature for non-planar device, one of which is the vertical device. Moreover, the combination of Gummel and Newton numerical methods was employed for obtaining a convergence in the numerical solution.

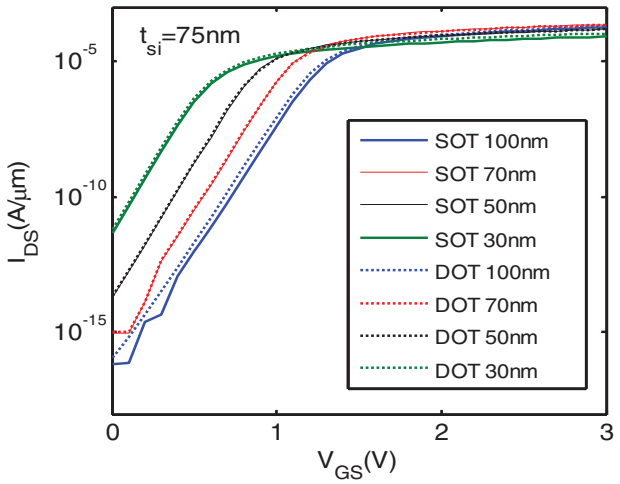


Fig. 2 The transfer characteristic comparison between SOT and DOT (V_d = 0.1V for both cases) for various channel lengths L_{ch} at t_{si} =75 nm

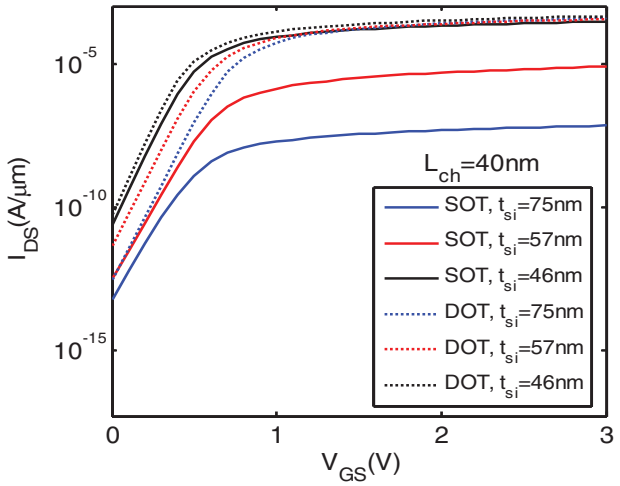


Fig. 3 The transfer characteristic comparison between SOT and DOT (V_d = 0.1V for both cases) for some pillar thicknesses t_{si} at L_{ch} = 40 nm

III. Result And Discussion

Basically, there is slight difference between SOT and DOT topologies due to the asymmetry in the structure of both top and bottom pillar's implanted regions. In the case of thinner t_{si}, the implanted region at the bottom parts of pillar may eventually merge together and forming single, large region in the thinning of pillar. Thus the pillar's centre region or the channel area become floating [16]. For SOT configuration, the joining of drain regions (left and right bottom sides of pillar, as shown in Fig. 1(f)) occurs for thicker pillar thickness, as the drain's depletion (in the bottom, for SOT) is wider than that of source (for DOT case). For device with t_{si} = 75nm and above, it was found that for DOT, the pillar region is connected to and have the same potential with substrate, similar to bulk mode. For smaller pillar thickness, this region is separated from substrate by the source region and is not tied to any potential, thus creating the floating body effect. On the other hand, the same channel area is already floating for SOT at t_{si} = 75nm and below. This floating body effect could be a disadvantage for the device performance, as it could reduce the output resistance and creating parasitic transistor on it [22].

The transfer characteristic of SOT configuration versus DOT is shown for fixed pillar thickness (Fig. 2) and for fixed channel length (Fig. 3), both taken for V_D = 0.1 V. The graphs show a slight difference in the subthreshold characteristics of I_D-V_G between SOT and DOT for channel length variation in thick pillar, t_{si} = 75 nm, while presenting considerable asymmetry in the change of pillar thickness in 40 nm channel length, especially for the on-state current I_{ON}. However, the various pillar thickness reveals that the DOT configuration has smaller curve variations than the SOT.

The threshold voltage comparison (Fig. 4) shows the tendency to decrease for lower t_{si} and L_{ch}, which is found for both SOT and DOT. However, the SOT has higher threshold voltage than DOT which comes from the additional depletion barrier created in the bottom pillar for V_D > 0. This additional barrier also explains the finding of worse DIBL for SOT, as the channel area get into floating faster when the drain voltage is increased. For higher drain voltage, the threshold voltage of SOT topology drops more than the DOT may experience, thus the DIBL value is increased, as evident in Fig. 5. The DIBL rises when the channel length L_{ch} is scaled further as the V_T roll-off is bigger at higher V_D (1.0V) due to the floating body effect which introduces majority carriers across the S/D junction and resides on the depletion region. This pattern is evident for all pillar thickness at t_{si} = 57 and 46 nm while for t_{si} = 75 nm the DIBL variation is smaller (due to the connection of centre region to bulk) and also for t_{si} = 36 nm, where the centre region is near fully depleted at which only small quantity of charge are available to be affected by higher V_D. The DIBL reduction for lower pillar thickness confirms the reduced effect of channel scaling on threshold voltage. It also shows the dependence of V_T roll-off with the silicon thickness, as previously stated. On the other hand, the larger pillar thickness in DOT gives the possibility of direct connection of channel area to the ground potential, and it results in better DIBL for t_{si} = 75 nm. With the pillar going thinner, floating body occurred resulting in larger threshold voltage drop thus increasing DIBL.

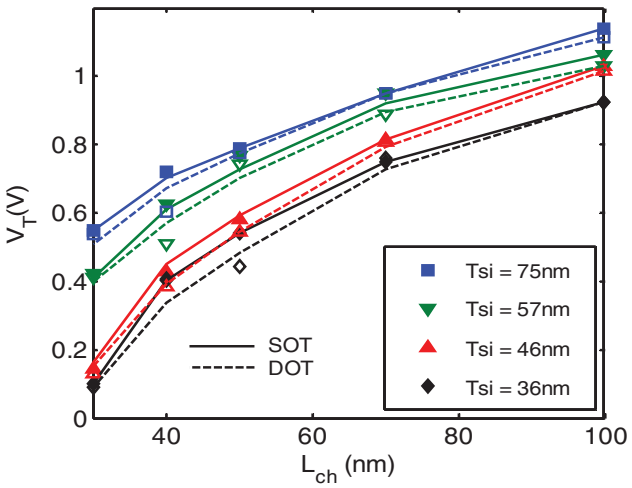


Fig. 4 Comparison of threshold voltage for SOT and DOT at several channel length in different pillar thickness (taken for V_D = 0.1 V)

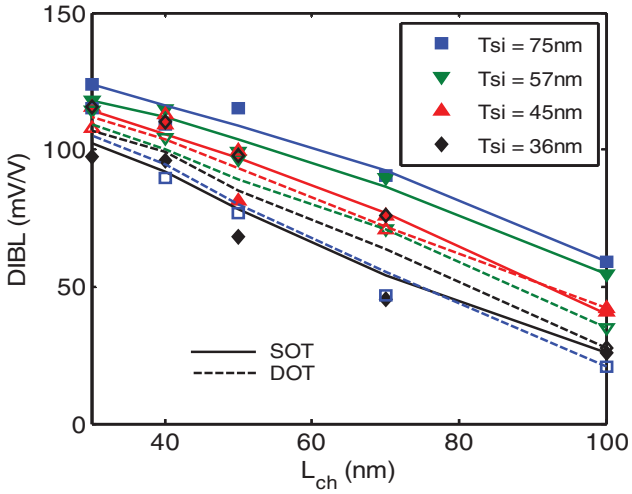


Fig. 5 DIBL of both SOT and DOT configuration for several channel lengths in different pillar thickness (for V_D = 0.1 V and V_D = 1.0 V)

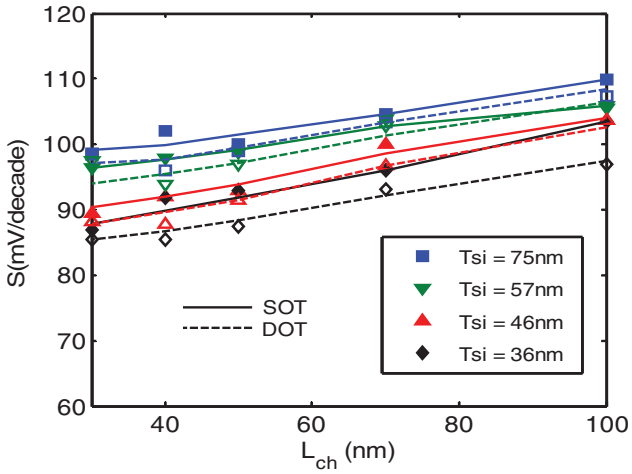


Fig. 6 Subthreshold swing characteristic of several channel length in different pillar thickness, for SOT and DOT configurations

Fig. 6 reveals the trend of subthreshold slope (S) with variations of channel length at several pillar thicknesses. The slope have a tendency towards lower value with the reduction of pillar thickness but still far from ideal value of 60 mV/decade, possibly due to high body effect coefficient [23]. Low swing reflects a good transition time. Thus lower S value is preferable. The decreased swing indicates an increase in gate–to-gate charge coupling with a decreasing pillar thickness, as explained in [24]. The subthreshold slope also tends to decrease for reduced channel length. This phenomenon could be explained by the low gate capacitance when L_{ch} gets shorter while near 40 nm the charge sharing take place. The wider depletion of drain at the bottom region is believed to be the cause of larger variation. Moreover, the subthreshold swing in DOT configuration is less than that in SOT due to the effect of bigger depletion charge between gate and drain which converts the channel to inversion faster at lower gate voltage.

Reducing the silicon pillar thickness appeared to improve the performance of this vertical double gate MOSFET. Generally the device performance deteriorates at $L_{ch} < 30$ nm with higher leakage current, lower threshold voltage and suffer from higher subthreshold swing, which are the indication of prominence of SCE. This could lead to tremendous rise in power consumption and overheating as the power needed in 'off' state will be significantly high.

IV. CONCLUSION

The electrical characteristics of the SOT and DOT topology have been analyzed especially at the sub-threshold region in order to observe the short channel effect of the device. The transfer characteristics of both topologies shows that asymmetry occurred by channel length reduction and pillar thinning. The result shows that silicon pillar thickness reduction (t_{si} = 36 – 100 nm) enhanced the DIBL and subthreshold switching performance, while the threshold voltage roll-off changed in nearly the same degree with the thickness variation. While the asymmetry in the performance of sub-threshold slope between DOT and SOT topologies is seen for several metrics, the acceptable SCE control is obtained in the lower silicon thickness for short channel length up to 30 nm with some concern on its power consumption. In more complex circuits with both topologies used, a careful design is needed to balance the performance between topologies and to compensate the asymmetry from each type of devices used.

ACKNOWLEDGMENT

The author would like to thank Prof. Dr. Razali Ismail of Universiti Teknologi Malaysia for his assistance in collecting the simulation data.

REFERENCES

[1] K. Mistry, *et al.*, "A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging," in *IEEE International Electron Devices Meeting (IEDM)*, 10-12 Dec. 2007, pp. 247-250.

[2] K. J. Kuhn, "CMOS transistor scaling past 32nm and implications on variation," in *IEEE/SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 11-13 July 2010, pp. 241-246.

[3] ITRS. The International Technology Roadmap for Semiconductors [Online]. Available: http://www.itrs.net/reports.html

[4] T. Poiroux, *et al.*, "Multiple gate devices: advantages and challenges," *Microelectronic Engineering,* vol. 80, pp. 378-385, 2005.

[5] T. J. K. Liu and L. Chang, "Transistor Scaling to the Limit," in *Into the Nano Era*, H. R. Huff, Ed.: Springer Berlin, 2009, pp. 191-223.

[6] E. J. Nowak, *et al.*, "Turning silicon on its edge [double gate CMOS/FinFET technology]," *IEEE Circuits and Devices Magazine,* vol. 20, pp. 20-31, 2004.

[7] M. A. Riyadi, J. E. Suseno, and R. Ismail, "The Future of Non-planar Nanoelectronics MOSFET Devices: A Review," *Journal of Applied Sciences,* vol. 10, pp. 2136-2146, 2010.

[8] G. M. Ertosun, *et al.*, "A nanoscale vertical double-gate single-transistor capacitorless DRAM," *IEEE Electron Device Letters,* vol. 29, pp. 615-617, 2008.

[9] M. Masahara, *et al.*, "Vertical Double-Gate MOSFET Device Technology," *Electronics and Communications in Japan,* vol. 91, pp. 46-51, 2008.

[10] H. Cho, *et al.*, "A low power, highly scalable, vertical double gate MOSFET using novel processes," in *65th DRC Device Research Conference*, 2007, pp. 173-174.

[11] E. Gili, *et al.*, "Shallow junctions on pillar sidewalls for sub-100-nm vertical MOSFETs," *IEEE Electron Device Letters,* vol. 27, pp. 692-695, 2006.

[12] J. Moers, "Turning the world vertical: MOSFETs with current flow perpendicular to the wafer surface," *Applied Physics A: Materials Science & Processing,* vol. 87, pp. 531-537, 2007.

[13] T. Schulz, *et al.*, "Planar and vertical double gate concepts," *Solid State Electronics,* vol. 46, pp. 985-989, 2002.

[14] I. Saad and R. Ismail, "Self-aligned vertical double-gate MOSFET (VDGM) with the oblique rotating ion implantation (ORI) method," *Microelectronics Journal,* vol. 39, pp. 1538-1541, 2008.

[15] M. A. Riyadi, *et al.*, "Vertical Double Gate MOSFET For Nanoscale Device With Fully Depleted Feature," *AIP Conference Proceedings,* vol. 1136, pp. 248-252, 2009.

[16] M. Riyadi, I. Saad, and R. Ismail, "Investigation of pillar thickness variation effect on oblique rotating implantation (ORI)-based vertical double gate MOSFET," *Microelectronics Journal,* vol. 41, pp. 827-833, 2010.

[17] S. Oda and D. Ferry, *Silicon Nanoelectronics*. Boca Raton, FL: CRC Press, 2006.

[18] E. Gili, *et al.*, "Asymmetric gate-induced drain leakage and body leakage in vertical MOSFETs with reduced parasitic capacitance," *IEEE Trans. Electron Devices,* vol. 53, pp. 1080-1087, 2006.

[19] *Atlas and Athena User Manual*: Silvaco International, 2005.

[20] N. D. Jankovic and G. A. Armstrong, "Comparative analysis of the DC performance of DG MOSFETs on highly-doped and near-intrinsic silicon layers," *Microelectronics Journal,* vol. 35, pp. 647-653, 2004.

[21] C. Lombardi, *et al.*, "A physically based mobility model for numerical simulation ofnonplanar devices," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems,* vol. 7, pp. 1164-1171, 1988.

[22] E. Gili, *et al.*, "Single, double and surround gate vertical MOSFETs with reduced parasitic capacitance," *Solid State Electronics,* vol. 48, pp. 511-519, 2004.

[23] Y. Taur and T. H. Ning, *Fundamentals of modern VLSI devices*: Cambridge University Press, 1998.

[24] M. M. A. Hakim, *et al.*, "Effect of transition from PD to FD operation on the depletion isolation effect in vertical MOSFETs," in *Proc. 6th International Conference on ULtimate Integration of Silicon (ULIS)*, Bologna, Italy, 2005, pp. 131-134

Session 5

Control Systems

Controlling of Non-Minimum Phase Micro Hydro Power Plant Based on Adaptive B-Spline Neural Network

Iwan Setiawan[1,2]
[1]Department of Electrical Engineering
Diponegoro University
Semarang, Indonesia
[2]Department of Electrical Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
iwan12@mhs.ee.its.ac.id

Ardyono Priyadi[3], Mauridhi Hery Purnomo[4]
Department of Electrical Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
[3]priyadi@ee.its.ac.id
[4]hery@ee.its.ac.id

***Abstract*—Hydro power plant is a power generation system that have non-minimum phase model showing initial inverse response characteristic. For span of broad electrical load regulation, conventional non adaptive control techniques, such as PI and PID control would degrade the performance of this power generation system. To ensure the stability of Hydro power plant for severe load variations, we need a kind of controller that has adaptive capability. On the other hand, the utilization of conventional adaptive techniques such as Self Tuning Regulator and Model Reference Adaptive Controller will be diverge to control plants showing non-minimum phase mode. In this paper, the implementation of adaptive intelligence control based on B-Spline neural network along with forward controller for controlling micro hydro power plant will be presented. Based on its characteristic, this adaptive control technique could be implemented directly without any prior training phase. From the simulation studies, the proposed scheme results fast transient response to load variations compared to traditional PI control and also very stable in responding to severe disturbance.**

Keywords—Adaptive Control; B-Spline Neural Network; Micro Hydro Power; Non-minimum Phase model

I. INTRODUCTION

Micro hydro power plant (MHPP) is a type of green and renewable energy resource that has potential to be developed in rural or remote areas that have no access to main power grid. One major problem encountered in MHPP is regulating the rotational speed of the hydro turbine at a predetermined reference. To ensure the quality of electric power, the speed of turbine rotation should ideally be independent from electrical loading variations. Without proper control method, severe changes of the electrical load will directly affect transient stability and final output of turbine rotation speed.

In the large scale of hydro power plant, the component which has the responsibility to control turbine rotation speed is hydraulic-mechanical or electro-hydraulic governor [1][2]. The main function of this component is to regulate the water gate or valve continuously in response to electrical load variations so mechanical input and electrical output power balance are maintained. However, these governors are not economical and practically very complicated to implement. The more viable alternative to regulate the water gate or valve in small or micro scale of hydro power plant is using servo motor [3][4].

Regarding to control methods utilized in MHPP, until now, the control strategy based on conventional and non-adaptive techniques such as PI, PID and Fuzzy control has been proposed in the literatures [1][2][5][6][7]. To get optimal response, the parameters of these controller should be tuned in a certain nominal load condition, in the event of large load variations, these techniques would generally provide less satisfactory results. To ensure the stability of the control system in response to severe load variations, we need controller that has adaptive capability. On the other hand, the implementation of conventional adaptive techniques such as Self Tuning Regulator (STR) and Model Reference Adaptive Controller (MRAC) that are very common found in adaptive control field will be diverge to control plants showing non-minimum phase model [8].

To regulate the turbine rotation speed which essentially has non-minimum phase model in an adaptive manner, intelligence method is one of a very promising control technique to be utilized. In this paper, the implementation of adaptive intelligence control based on B-Spline neural network and its scheme for controlling the rotational speed of hydro turbine in MHPP will be presented. From the perspective of control designer, the implementation of this intelligence technique has several advantages compared to conventional control techniques: (1) the designer does not need to know the exact model and its parameters of the plant. (2) this adaptive techniques can be implemented directly without any prior training phase. The main objective of this paper is to study the implementation of B-Spline based controller and the effect of controller parameter variation against MHPP stability.

The remainder of this paper is organized as follows. Section 2 describes the components composing of MHPP and the short theories of B-Spline neural networks. Moreover, the proposed control scheme will be discussed in Section 3. Next, Section 4 shows the simulation results and discuss the performance of the

intelligence control technique that have been implemented. Finally, the conclusions are drawn in Section 5.

II. Micro Hydro Power Plant Model and Adaptive Neural Network Based Controller

A. Synchronous Generator Model

Synchronous generator is a common type of generator used in power generation system. The generated frequency of the electrical voltage is sync with the speed of rotor rotation. Generator steady state operation will be achieved if active load electrical power consumption equals with mechanical power input driving the generator. The imbalance between the mechanical power input and electrical power output will make the rotor speed to change until a new steady condition is reached again. Equation (1) below shows rotation model of synchronous generator in Laplace domain.

$$\Delta f = \frac{K}{1+sT}(\Delta P_M - \Delta P_L) \quad (1)$$

where K (Hz/pu) and T (s) respectively are static gain and time constant of generator, whereas Δf (Hz), ΔP_M (pu), dan ΔP_L (pu) respectively are deviations of frequency, mechanical power input (turbine mechanical power output) and real electrical power consumption.

B. Hydro Turbine Model

The equation (2) shows linear relationship betwen deviation of turbine mechanical power output against valve position change in the Laplace transfer function form.

$$\frac{\Delta P_M}{\Delta V} = \frac{1-T_w s}{1+0.5T_w s} \quad (2)$$

In this case, ΔV (%) is valve opening deviation from the nominal operating point, whereas Tw (s) is known as water starting time that has varying value between 1-4s [2], the more high of head, the bigger value of Tw.

From equation (2), it is clear that matematically, hydro turbine is a non-minimum phase model (the zero of that model reside in the right half plane). The output of this model would indicate undershoot (overshoot) at the initial response against positive (negative) input changes. These characteristics practically will make control desain is very difficult [11].

In relation to adaptive control techniques, the utilization of elementary self-tuners and most MRAC algorithms to control a non-minimum phase model will give instability, so there is a misapprehension that non-minimum phase systems can pose insoluble problems with the application of adaptive control [8]. So until recently direct conventional adaptive control was limited to minimum-phase linear plants with unknown parameters [11].

C. Servo Motor as Governor Model

Judging from the economic aspect, utilization of mechanical-hydraulic or electro-hydraulic governor that is very commonly found in large scale power generation systems is not economical for MHPP. For the micro-scale power system, it would be more feasible if that kind of governor was replaced by servo motor. Transfer function of this servo is shown by equation (3) below [3][4]:

$$\frac{\Delta V}{\Delta U} = \frac{1}{(1+T_1 s)(1+T_2 s)+1} \quad (3)$$

where T_1 (s) and T_2 (s) respectively are electrical and mechanical time constants of servomotor, whereas ΔU is deviation of control signal.

D. Basic Topology of Adaptive Neural Networks based Controller

One of the pioneer work in the utilization of Neural Networks (NN) as an online controller is proposed by Kraft with the basic topology shown in Fig. 1. In the adaptive control field literature, that topology known as fixed stabilizing controller. As indicated by the fig.1, There are fundamentally two main control loop which have different functionality: (1) normal feedback loop with a simple proportional gain that used for stability maintain purpose, and (2) adaptive NN loop which play rule as integrator controller with on line weights updating algorithm. Due to the lack of available explicit target signals to train on line NN weights, in this topology, signal used as a source of training is the output of the controller itself[13].

E. B-Spline Neural Network

B-spline neural network (BSNN) is a type of neural network (NN) with in some senses have Fuzzy logic characteristic. In the field of artificial neural network, BSNN is categorized as Associative Memory Networks (AMN) which have local learning properties [13].

Fig. 2 depicts schematic diagram of BSNN. It is showed that for every single input vector, there are only a certain number of activated basis functions (the number of activated basis functions depend on the order of that function). The output of BSNN is a linear combination of all active basis function outputs which weighted by the associated adaptive weights. Mathematically, the input-output relationship of BSNN is represented by equation (4).

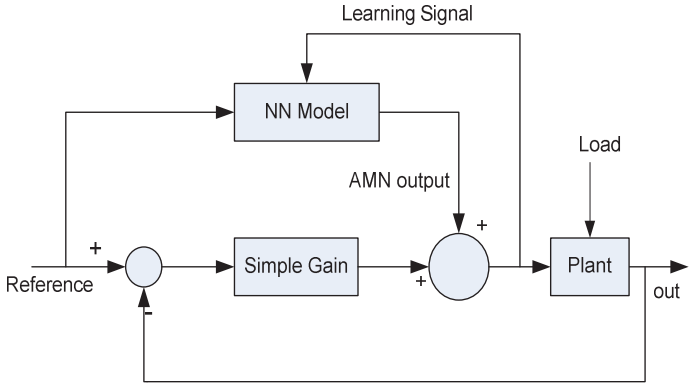


Fig. 1. Basic Adaptive Neural networks based control topology

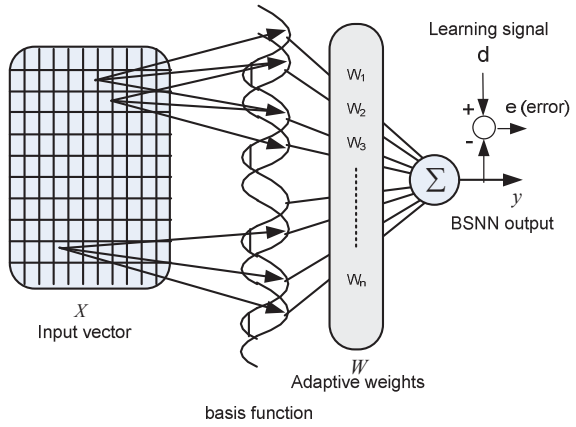


Fig. 2. Schematic diagram of B-Spline Neural Network

$$y(k) = \sum_{i=1}^{p} a_i(k) w_i(k) \quad (4)$$

where $w_i(k)$ is weight that associated with *i-th* basis function, a_i is *i-th* basis function output*: i*=(1,2,…. ρ).

Weight updating in B-spline neural network can be computed easily using least mean square algorithm that showed in equation (5) below:

$$w(k+1) = w(k) + \alpha e(k) a(k) \quad (5)$$

where:

$w(k)$: neural network weight at k-th sample, α : convergention rate ($\alpha > 0$), $e(k) = t(k)-y(k)$: error of output neural netowork at k-th sample, and a(k) : output of basis function at *k-th* sample.

In BSNN, higher orde of basis functions can be computed recursively from the lower ones. The recursive relation of certain basis function represented in equation (6). The non linear modeling capability of BSNN basicaly depend on chosen basis function order. Fig. 3 show three lower order of univariate basis functions.

$$N_k^j(x) = \left(\frac{x - \lambda_{j-k}}{\lambda_{j-1} - \lambda_{j-k}} \right) N_{k-1}^{j-1}(x) + \left(\frac{\lambda_j - x}{\lambda_j - \lambda_{j-k+1}} \right) N_{k-1}^{j}(x)$$

$$N_1^j(x) = 1 \quad \text{if } x \in Ij \ (\lambda j\text{-}1, \lambda j)$$

$$= 0 \text{ , others} \quad (6)$$

In this case, λ_j is knot of *j-th* basis function, $I_j = (\lambda_{j-1}, \lambda_j)$ is *j-th* interval, whereas *k* is basis function order.

III. Propose B-Spline Neural Network Based Control Topology for MHPP

Fig. 4 show the complete diagram block of MHPP control using a light modified of fix stabilizing controller architecture. For the parameter values of MHPP model, in this study we refer to [3][5]. The feed-forward control block $H_{FF}(s)$ is basically just an amplifier that is intended to speed up the response of the control system in response to changes in electrical load.

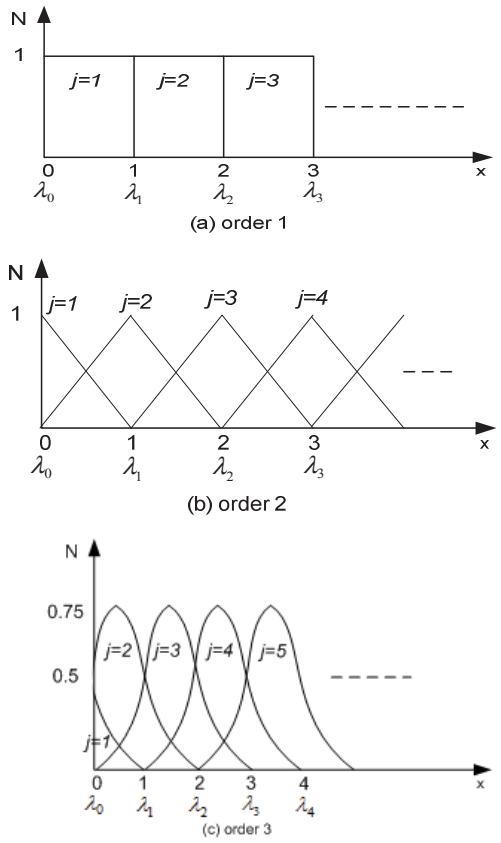


Fig. 3.Three lower order of basis functions

$H_{FF}(s)$ can be found simply by using the following formulation:

$$H_{FF}(0) = \frac{1}{H_1(0) H_2(0)} \quad (7)$$

To maintain the stability of the feedback control system as well as the stability of the neural network training, the output of the control system of a conventional feedback control block should be limited by the saturation block as seen in Figure 4. The role of saturation block is very important especially in the transient time to maintain the stability of the learning process and to constraint the learning signal fed to BSNN block.

To verify the performance and stability of B-spline neural network based-control topology, we have done extensive simulation study using Matlab Simulink software. In this study we investigated the influence of simple proportional gain variation and control system stability under electrical load variation. We also compare this results with PI control strategy with optimum parameter derived from papers[3][5].

IV. Simulation Results and Discussion

Based on simulation result (fig. 5, fig.6 and tabel 1), the performance of BSNN control in response to step change of lectrical load is faster than traditional PI control strategy with

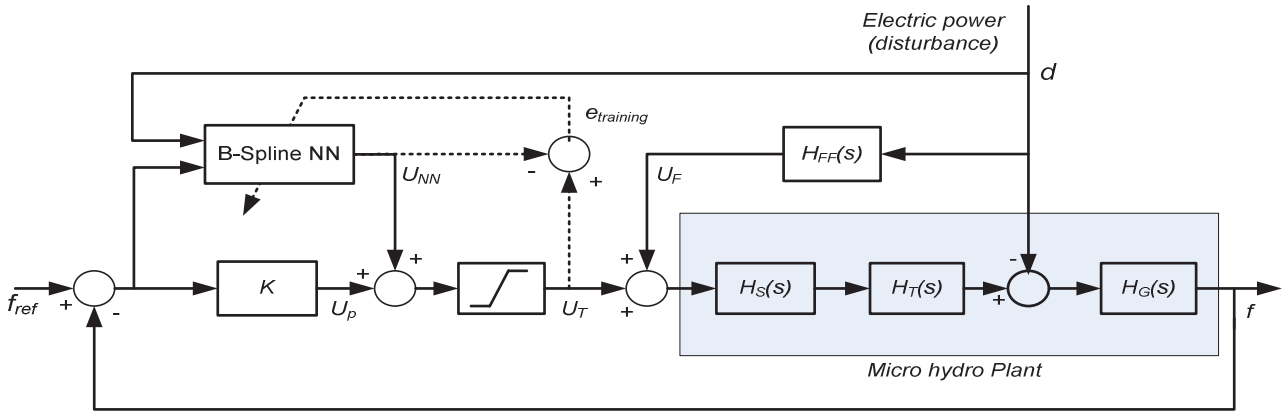


Fig. 4. Proposed intelligence control topology for Small Hydro Turbin

optimum parameter, with initial undershoot response just about 0.6 Hz and settling time less than 50 s. The fast recovery time actually is achived based on feedfoward controller and knowladge of controller about the target signal that should be injected to servo motor to maintain the balance of mechanical and electrical load power. In the steady state, control output that drive valve solely come from BSNN output and feedfoward controller.

The stability of this control topology basically is depend on the simple proportional gain, variation of learning rate and control saturation block that exist in feedback loop of control system. The main rule of this saturation block is to restrict output control that used as learning signal in the definite range so the feedback control will be stable. In this simulation, the saturation of control is set in the range -0.2:0.2. Fig. 7 and fig. 8 respectively show that the transient performance of electrical power frequency output of MHPP is very sensitive to proportional gain (learning rate) value. The bigger proportional gain (learning rate), the fast transient response will be achieved, but however, there should be gain margin in such a way that the control system would be stable. The gain margin and learning rate margin that guaranted the control stability generally depend on the characteristic of plant to be controlled and this is should be done empirically with simulation study using realistic plant model.

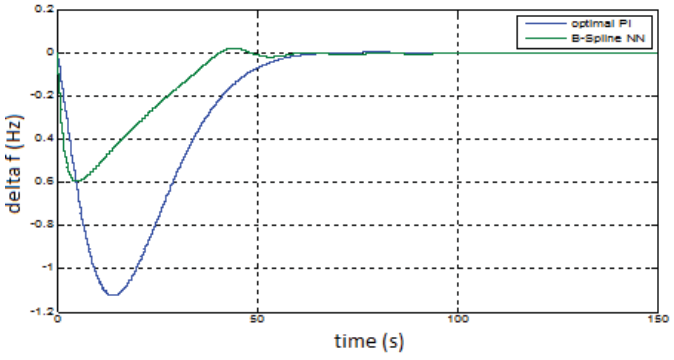


Fig. 5. Response of PI vs Adaptive BSNN control for the step loading variation of 0.03 pu (K=0.0001, α=0.1)

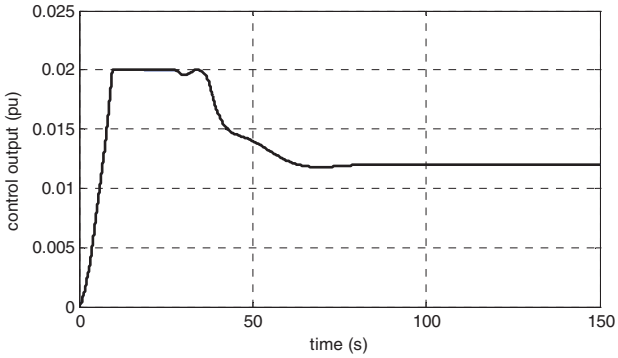


Fig. 6. Output control of Adaptive BSNN for the step loading variation of 0.03 pu (K=0.0001, α=0.1)

This designed BSNN control topology also have good stability in response to changing of electrical load. For the proper simple proportional gain value, fig. 9 show response control system to electrical load variation.

Tabel 1. Comparison of control system performance

	Undershoot (Hz)	Overshoot (Hz)	Time Settling (s)	MSE
Optimal PI	-1.124	-	60	0.0015
B-Spline NN	-0.597	0.021	50	0.000688

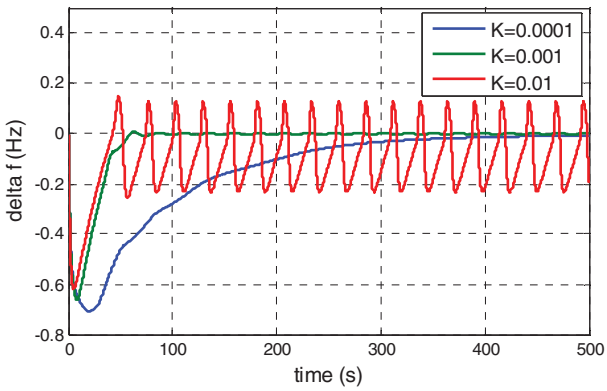


Fig. 7. Effect of simple gain variation

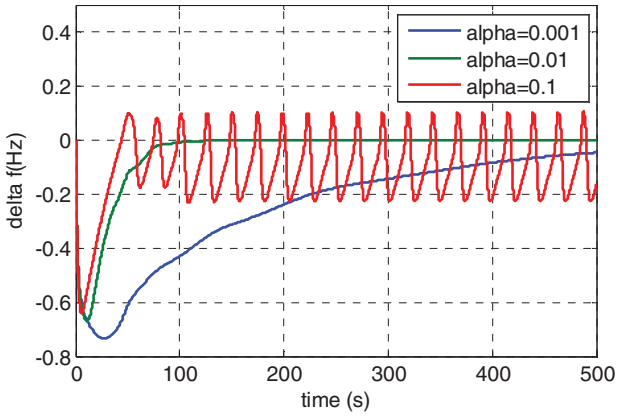


Fig. 8. Effect of learning rate

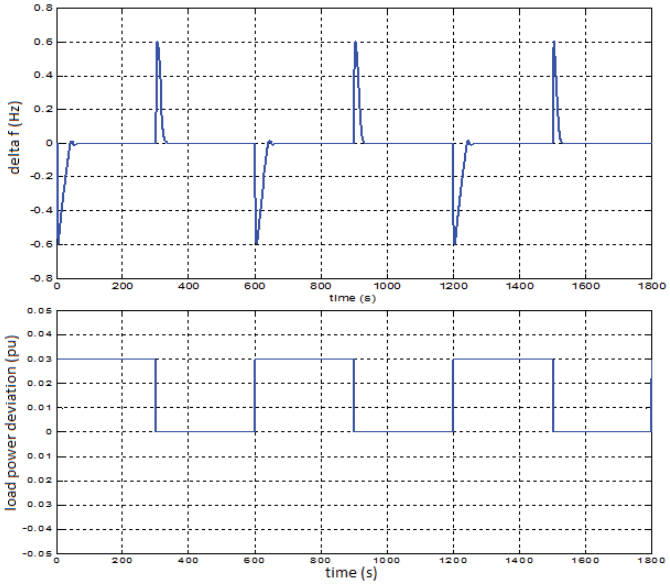


Fig. 9. Frequency respon to variation of load

V. CONCLUSSIONS

In this paper, Adaptive B-spline neural network with light modified fixed stabilizing controller topology for control of MHPP has been proposed. From the exhaustive simulation study, the BSNN controller have fast transient response to variation of electrical load, but however the transient stability of this control topology is very influenced by the simple proportional gain and learning rate value.

In general, this works also has prove the suitablity of adaptive-B-spline neural networks for control of non-minimum phase models.

ACKNOWLEDGMENT

First author would to tanks to Indonesian Directorate General of Higher Education for giving the full financial support to study in Doctor program at ITS, surabaya.

REFERENCES

[1] N. Kishor, R.P Saini, S.P.Singh, "A Review on Hydropower Plant Models and Control", ScienceDirect, Renewable and Sustainable Energy Reviews 11 (2007), p776-796

[2] P.kundur, "Power Stability and Control", Tata-McGraw Hill Co. 1221, Avenue of America, New York, NY, 1994

[3] H.Goyal, M. Hanmandlu, D.P.Kothari, "A Three Gate Optimal Flow Control Approach for Automatic Control of Small Hydro Power Plants Using Neural Networks", International Energy Jounal; vol 7, no 2, June 2006

[4] M. Hanmandlu, H. Goyal, "Proposing a new advanced control technique for micro hydro power plants", ScienceDirect, Electrical Power and Energy Systems 30 (2008) 272–282

[5] M. Hanmandlu, H. Goyal, D.P. Kothari, "An Advanced Control Scheme for Micro Hydro Power Plants", International Conference on Power Eletronis, Drives and Energy Systems, 2006. PEDES 06. 12-15 Dec2006 Page(s):1-7

[6] 1.Nanaware R.A.,S.R. Sawant, B.T. Jadhav, "Fuzzy Based Turbine Governor for Hydro Power Plant", International Journal of Electrical Engineering. ISSN 0974-2158 Volume 5, Number 4 (2012), pp. 383-392

[7] M. Abbas, M. Saleem Khan, N. Ali, "Fuzzy Logic Based Hydro-Electric Power Dam Control System, International Journal of Scientific & Engineering Research Volume 2, Issue 6, June-2011 1 ISSN 2229-5518

[8] D. W. CLARKE, "Self-tuning Control of Non minimum-phase Systems", Automatica, Vol. 20, No. 5, pp. 501 517, 1984

[9] S. Wang, Z. Zhang, "Application of Soft Computing in the Control of Hydroelectric Generating Unit", International Workshop on Intelligent Systems and Applications, 2009. ISA 2009.

[10] B.Satpati, I.Bandyopadhyay, G.Das, C. Koley, "Robust Controller Design for Load Frequency Control ofNon-minimum Phase Hydro Power Plant using PSG Enabled Automated Quantitative Feedback Theory", India Conference, INDICON 2008. Annual IEEE, p: 349 – 354

[11] C.R. Johnson, "On Direct Adaptive for The Control of Nonminimum-Phase Piants", Information Sciences, 20, pp 165- 179(1980)

[12] E. Ozbay, M. Tunay," Self-Tuning Fuzzy PI Controlled System Model for Small Hydro Power Plants", 10th International Conference on Clean Energy (ICCE-2010)

[13] B.M..Harris, "Neurofuzzy Adaptive Modelling and Control", Prentice Hall (UK), 1994.

[14] H.L.Kraft, P.D. Campagna, "A Comparison Between CMAC Neural Network Control and Two Traditional Adaptive Control Systems". IEEE Contr. Sys. Mag. Apr. pp 36-43 1990.

OCP Based Decentralized Data Fusion For Autonomous Underwater Vehicles

Nanang Syahroni
Electrical Department
Politeknik Elektronika Negeri Surabaya
Surabaya, Indonesia
nanang@eepis-its.edu

Abstract --- In this paper, an online decentralized multisensor data fusion algorithm facilitated by middleware networked using CORBA event channel is proposed, in order to deal with simplifying problem in sensor registration and fusion for vehicle state estimation. The networked based navigation concept for Autonomous Underwater Vehicle (AUV) using several sensors is presented. A simulation of various application scenarios are considered by choosing several parameters of UKF, i.e. weighting constant for sigma points and square root matrix. Normalized mean-square error (MSE) of Monte Carlo simulations are computed and reported in the simulation results. Furthermore, the middleware infrastructure based on Open Control Platform (OCP) to support the interconnection between the whole filter structures also reported.

Keywords --- Autonomous Underwater Vehicle, Open Control Platform, Decentralized Data fusion, Unscented Kalman Filter.

I. INTRODUCTION

The software based controls for robotic and autonomous vehicle have been dominated in recent years. An OCP is an object-oriented software infrastructure implemented that allows seamless integration of cross-platform software and hardware components in any control system architecture. An OCP is a middleware that is based on the real-time common object request broker architecture (RT-CORBA). Middleware is connectivity software that consists of a set of services, allowing multiple processes running on one or more machines to interact across a network [1].

The multisensor data fusion using Kalman filter (KF) has been widely applied in integrated navigation system for many applications [2-3]. Estimation of navigation system in nonlinear system approach to use the extended Kalman filter (EKF) which simply linearizes all nonlinear models are reported in [4-5], so that the traditional KF can be used. An alternative approach, the Unscented Kalman filter (UKF) where the random variable, Gaussian distributions is linearized while the nonlinear model equations are directly used in the calculations [6]. The centralized filter where all measured sensor data are communicated to the central site for processing [7], and distributed filter [8] where the local estimators from all sensor can yield the global optimal or sub optimal state estimator according to certain information fusion criterion.

In this paper, an online decentralized multisensor data fusion of two stage federated UKF algorithms connected by RT-CORBA middleware network is proposed. We assumed that the problem solution of fault detection and isolation in the Autonomous Underwater Vehicle (AUV) will made easily.

II. FILTERING MODEL

In this chapter, the design and implementation of online decentralized sensor fusion is discussed. Sensor fusion is the combining of sensory data or data derived from sensory data from disparate sources such that the resulting information is in some sense better than would be possible when these sources were used individually. The term better in that case can mean more accurate, more complete, or more dependable. The data sources for a fusion process are not specified to originate from identical sensors. One can distinguish direct fusion, indirect fusion and fusion of the outputs of the former two. Direct fusion is the fusion of sensor data from a set of heterogeneous or homogeneous sensors, soft sensors, and history values of sensor data, while indirect fusion uses information sources like a priori knowledge about the environment and interupt input. A sensor fusion technique should be able to estimates a state of AUV through time due to its dynamics, which experiences a set of complicated and highly nonlinear forces.

The time evolution is describe using general continuous-time model:

$$x_T(t) = f_T\left[x_T(t), u_T(t), v_T(t), t\right] \tag{6}$$

Subscript T denotes the fact that this is the state of true system, in practice the structure and form of true system is unknown and approximations must be used. The current position and velocity of the AUV as well as on certain characteristics are not precisely known.

In general, the nonlinear system dynamics and observation equations of AUV actuator j in decentralized discrete form of UKF scheme in [8] are given as:

$$\begin{aligned} x_{j,k} &= f(x_{j,k-1}, u_{j,k-1}, v_{j,k-1}, k-1) + G_{k-1} w_{k-1} \\ z_{j,k} &= h_{j,k}(x_{j,k}) + \vartheta_{j,k} \end{aligned} \tag{7}$$

where $f_k(\cdot) \in \mathbb{R}^{n \times n}$ is the process model, $x_k \in \mathbb{R}^n$ is the state vector, $u_k \in \mathbb{R}^n$ is the control input, $v_k \in \mathbb{R}^n$ is the process noise, G_k is the system noise matrix, $w_k \in \mathbb{R}^n$ is the system noise, $z_k \in \mathbb{R}^p$ is the measurement vector, $h_k(\cdot) \in \mathbb{R}^{p \times n}$ is the measurement model, and $\vartheta_k \in \mathbb{R}^p$ is the measurement noise.

A. Initialization

$$\begin{aligned} \chi_{k,k-1} &= f(\chi_{k-1,}k-1) \\ Z_{k,k-1} &= h(\chi_{k-1}) \end{aligned} \tag{8}$$

We assume that noise is uncorrelated Gaussian white noise sequences with mean and covariance as follows: $E\{w_i\} = 0$, $E\{w_i w_j\} = Q\delta_{ij}$, $E\{\vartheta_i\} = 0$, $E\{\vartheta_i \vartheta_j\} = R\delta_{ij}$, $E\{w_i \vartheta_j\} = 0$, for all i, j . Where $E\{\cdot\}$ denotes the expectation, and δ_{ij} is the Kronecker delta function. Q and R are bound positive definitive matrices ($Q>0, R>0$). Initial state x_0 is normally distributed with zero mean and covariance P_0 .

B. Updating

B.1. Computing Sigma Point

$$\begin{aligned} \chi_{0,k-1} &= \hat{x}_{k-1} \\ \chi_{i,k-1} &= \hat{x}_{k-1} + (\sqrt{(n_s+\lambda)P_{k-1}})i, i = 1,...,n_s \\ \chi_{i,k-1} &= \hat{x}_{k-1} - (\sqrt{(n_s+\lambda)P_{k-1}})i, i = n_s+1,...,2n_s \end{aligned} \tag{9}$$

where χ is a sigma points of the augmented vector, n_s is a number of the states in the augmented state vector, λ is a scaling parameter $= \alpha^2(n_s + \kappa) - n_s$, α is to determines how the sigma point are spread, typical value is 10^{-3} , and κ is a scaling parameter which can be used to incorporate up to fourth order precision in the transformation, usually set to zero.

The prediction of the state and measurement vector as well covariance of the state vector is:

$$\begin{aligned} \hat{x}_k^- &= \sum_{i=0}^{2n_s} W_i^s \chi_{i,k|k-1} \\ \hat{z}_k^- &= \sum_{i=0}^{2n_s} W_i^m Z_{i,k|k-1} \\ P_k^- &= \sum_{i=0}^{2n_s} W_i^c (\chi_{i,k|k-1} - \hat{x}_k^-)(\chi_{i,k|k-1} - \hat{x}_k^-)^T + Q \end{aligned} \tag{10}$$

and weights for the states and covariance matrices is:

$$\begin{aligned} W_0^s &= \lambda/(n_s+\lambda) \\ W_0^c &= (1-\alpha^2+\beta) + 0.5\lambda/(n_s+\lambda) \\ W_i^s &= W_i^c = 0.5\lambda/(n_s+\lambda), i = 1,...,2n_s \end{aligned} \tag{11}$$

where β is used to incorporate knowledge of the distribution of state, optimal value for Gaussian distribution is 2, s, and c is a state and covariance respectively.

B.2. Time Update

$$\begin{aligned} \hat{x}_{j,k}^- &= \left(\sum_{i=0}^{2n_s} W_i^s X_{i,k|k-1} \right)_j \\ p_{j,k}^- &= \left(\sum_{i=0}^{2n_s} W_i^c \left(X_{i,k|k-1} - \hat{x}_k^- \right) \left(X_{i,k|k-1} - \hat{x}_k^- \right)^T + Q \right)_j \\ \hat{z}_{j,k} &= \left(\sum_{i=0}^{2n_s} W_i^m Z_{i,k|k-1} \right)_j \\ &\text{where: } j = 1,2,...,L,M \end{aligned} \tag{12}$$

B.3. Measurement Update

Update the measurement prediction covariance, the cross covariance between the state and measurement, the Kalman gain, the state estimate and the state covariance as in [8]:

$$\begin{aligned} P_{\hat{z}_k \hat{z}_k} &= \sum_{i=0}^{2n_s} W_i^c (Z_{i,k|k-1} - \hat{z}_k)(Z_{i,k|k-1} - \hat{z}_k)^T + R \\ P_{\hat{z}_k^- \hat{z}_k} &= \sum_{i=0}^{2n_s} W_i^c (Z_{i,k|k-1} - \hat{z}_k)(Z_{i,k|k-1} - \hat{z}_k)^T \\ K_k &= P_{\hat{z}_k^- \hat{z}_k} P_{\hat{z}_k \hat{z}_k}^{-1} \\ \hat{x}_k &= \hat{x}_k^- + K_k(z_k - \hat{z}_k) \\ \hat{P}_k &= P_k^- - K_k P_{\hat{z}_k \hat{z}_k} K_k^T \end{aligned} \tag{13}$$

When the several identical sensors are used, combining the observations will result an improved estimation. (A statistical advantage is gained by adding the N independent observations is improved by a factor proportional to $N^{1/2}$). To improve the observation process, two sensors that measure angular directions on AUV can be coordinated to determine the position using two sensors, one moving in a known way with respect to another, it can be used to measure instantaneously a position and velocity, with respect to the observing sensors. We employs a decentralize filters in our sensor fusion technique, that will makes a fault detection and isolation easier. Furthermore, the requirement of memory space to the fusion center is broadband, and the parallel structures can increase the input data rates.

In the decentralized form of UKF scheme as in [8], local filters are generally based on the models as:

$$\begin{aligned} x_k &= f(x_{k-1}, u_{k-1}, v_{k-1}, k-1) + G_{k-1} w_{k-1} \\ z_k &= h_k(x_k) + \vartheta_k \end{aligned} \quad (14)$$

As all the local UKFs estimate same state variables, these models that have the same dynamics are appropriate.

The decentralized UKF can obtain the globally optimal estimate by using the information strategy to each local filter and than fusing the estimates of the local filter. For the system with local UKF, equation for the time and measurement update is mentioned as:

Time update equations:

$$\hat{x}_{j,k}^{-} = \left(\sum_{i=0}^{2n_s} W_i^s X_{i,k|k-1} \right)_j \quad (15)$$

$$p_{j,k}^{-} = \left(\sum_{i=0}^{2n_s} W_i^c \left(X_{i,k|k-1} - \hat{x}_k^{-} \right) \left(X_{i,k|k-1} - \hat{x}_k^{-} \right)^T + Q \right)_j \quad (16)$$

where $j=1,2...,L,M$

$$\hat{z}_{j,k} = \left(\sum_{i=0}^{2n_s} W_i^m Z_{i,k|k-1} \right)_j, j=1,2...,L$$

Measurement update:

$$p_{j,\hat{z}_k \hat{z}_k} = \left(\sum_{i=0}^{2n_s} W_i^c \left(Z_{i,k|k-1} - \hat{z}_k \right) \left(Z_{i,k|k-1} - \hat{z}_k \right)^T + R \right)_j \quad (17)$$

$$p_{j,x_k^{-} k \hat{z}_k}^{-} = \left(\sum_{i=0}^{2n_s} W_i^c \left(X_{i,k|k-1} - \hat{x}_k \right) \left(Z_{i,k|k-1} - \hat{z}_k \right)^T \right)_j \quad (18)$$

where $j=1,2,...,L$

$$K_{j,k} = \left(P_{\hat{x}_k \hat{z}_k} P_{\hat{x}_k^{-} \hat{z}_k}^{-} \right)_j$$

$$\hat{x}_{j,k} = \left(\hat{x}_k^{-} + K_k \left(z_k - \hat{z}_k \right) \right)_j$$

$$\hat{P}_{j,k} = \left(P_k^{-} - K_k P_{\hat{z}_k \hat{z}_k} K_k^T \right)_j$$

where $\hat{x}_k^{-} \in \mathbb{R}^{nj}$ is the priori estimate of x_k, $Q \in \mathbb{R}^{nj \times nj}$ is the covariance matrix of system noise, $\hat{x}_k \in \mathbb{R}^{nj}$ is the posteriori estimate of x_k, $P_k^{-} \in \mathbb{R}^{nj \times nj}$ is the priori covariance matrix of estimation error, $\hat{P}_k \in \mathbb{R}^{nj \times nj}$ is the posteriori covariance matrix of estimation errors. In the [8], master filter are generally modeled as:

$$P_{f,k}^{-1} = P_{M,k}^{-1} + \sum_{j=1}^{N} P_{j,k}^{-1} \quad (19)$$

where $P_{M,k}^{-1} = \beta_{M,k} P_{f,k|k-1}^{-1}$.

$$P_{f,k}^{-1} \hat{x}_{f,k} = P_{M,k}^{-1} \hat{x}_{M,k} + \sum_{j=1}^{N} P_{j,k}^{-1} \hat{x}_{j,k} \quad (20)$$

where $\hat{x}_{M,k} = \hat{x}_{f,k|k-1}$

$P_{f,k}^{-1} \in \Re^{nf}$ is the inverse of the fused covariance, $\hat{x}_{f,k} \in \Re^{nf}$ is the fused state estimate.

Once the global solution is obtained, it can be feedback to the local UKF, this operation called the reset operation, by the following algorithm:

$$\left. \begin{aligned} P_{j,k} &= \beta_j^{-1} P_{f,k}, \\ \hat{x}_{j,k} &= \hat{x}_{f,k} \end{aligned} \right\}, j=1,2,...,M \quad (21)$$

where β_j is the information sharing coefficient and must satisfy the following conservation of information principle:

$$\beta_M + \sum_{j=1}^{N} \beta_j = 1 \quad (22)$$

III. SYSTEM DEVELOPMENT

In fig 1, shows the decentralized implementation of th e proposed method, the state vectors of the local and master filters are same. If the fusion and reset operation are performed after every measurement cycle, the decentralized UKF solution is the same as centralized UKF. To remain optimal, the filter must combine the local estimates into a single every cycle. After the combination step, at the start of the next cycle, the estimates can feedback to the local information. Furthermore, the total process information, represented by the matrix Q shared among the local filters, must sum up to the true net process information.

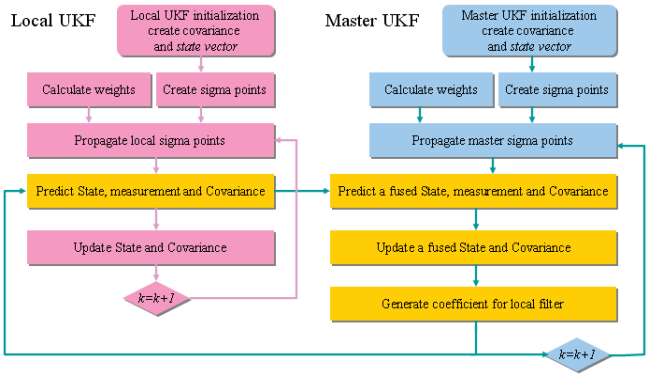


Fig. 1. The UKF implementation in a Decentralized Form

For the application of the proposed online decentralized multisensor fusion algorithm, there is no sensor acts as a fundamental sensor in the system, and its data is the measurement input for the local filter. The data from sensors

are dedicated to corresponding local filter, after calculation was completed than supply their resulting solutions to the main filter for the master update, yielding a global solution.

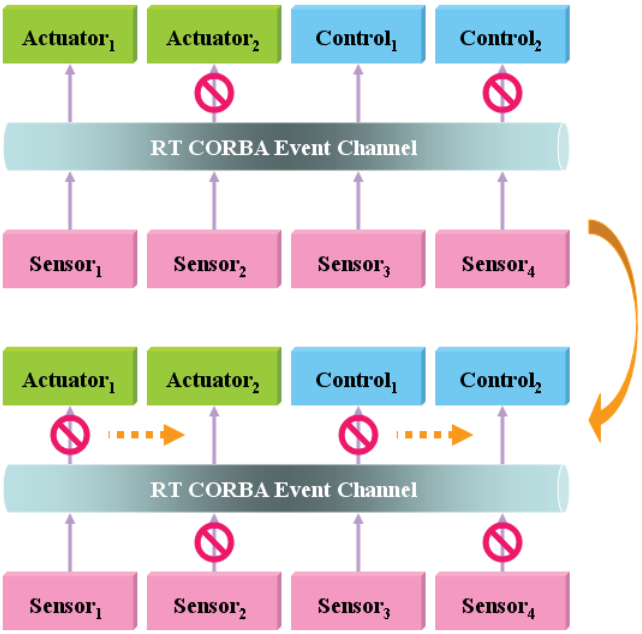


Fig. 2. Open Control Platform Rapid Changing

An OCP can accommodate rapidly changing application requirements, easily incorporate new technology (i.e. new hardware platforms or sensor technology), interoperate in heterogeneous, unpredictable, and changing environments as illustrated in fig. 2.

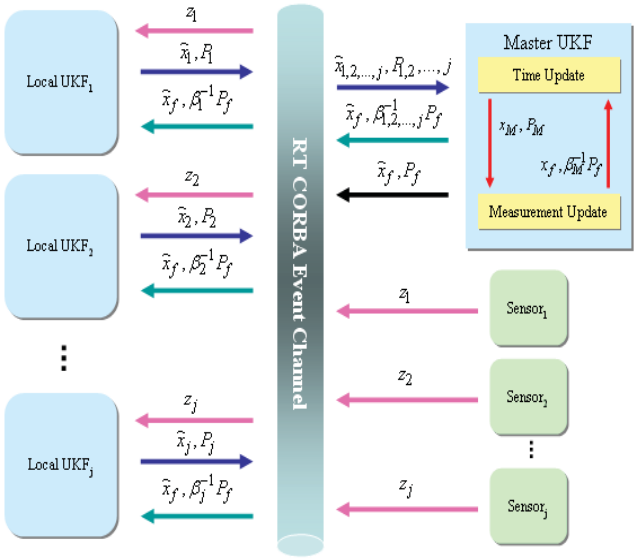


Fig. 3. Sensor Fusion Implementation

Using OCP structure will enable to accommodate changing application requirements and maintain viability in changing environments, which is suitable to integrate a high number of sensors and actuators as illustrated in fig. 3.

To develop a code that estimate the position, attitude, speed and other parameters of vehicle is one of the most safety-critical parts of software. Automatic code generator or program fusion techniques have reported in [9], to develop KF code from high-level declaration specification of state estimation problem can help to solve this predicament by completely automating the coding phase.

The algorithms and software code generation presented are designed to deliver an online robust and accurate state estimates (control and trajectory planning), predictive model (control, trajectory planning, fault detection, and recovery), uncertainty bounds (control, trajectory planning, and fault detection).

IV. SIMULATION RESULTS

A simulation environment that contains many of the complexities is required to validate the algorithm development. To solve this problem, we have a several steps: 1) developing a system dynamic model more detail, instead of Autonomous Vehicle, 2) to create some simulation scenarios that reflects the important situations, especially to test an algorithm and model in critical condition of implementation (e.g. when mean and covariance of Gaussian white noise sequences are bias).

The entire local UKF have a same state variables and same dynamics. To obtain a global optimal estimation of decentralized UKF, information from each local filter will be used, and than fuse the global estimation of the master filter, as depicted in fig 3.

For simulations, we have a simple scenario, that a sensor fusion technique should be able to track an Autonomous Vehicle that experiences a set of complicated and highly nonlinear forces. The current position and velocity of the vehicle, as well as on certain characteristics are not precisely known.

All the simulations for local filter and master filter have been carried out using Matlab. A random measurement noise from GPS is included in the simulations. Here, we assume that range sensor computations are sampled in 1ms, rotary sensor is provided at 1 Hz, and GPS updates are available at 30 Hz. In the simulations, all sensor are alignment in 1 Hz, and will be running for 400 iteration periods. The initial values $x_1(0)$ and $x_2(0)$ of the state vector x_1 and x_2 is chosen as zero and the initial alignment error are also assumed to be zero. $P_1(0), P_2(0), Q_1, Q_2, R_1$ and R_2 are chosen for a medium accuracy.

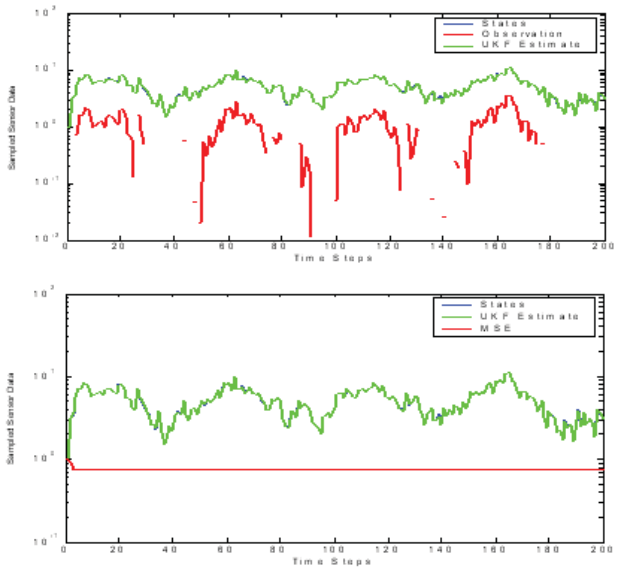


Fig. 4. Local Filter Estimation and Its MSE

A local filter estimation results are depicted in first picture of fig. 4 in red line, while green line is measurement signal from local sensor. Second picture show it's MSE. In a similar illustration, a master filter estimation results are depicted in fig 5 in first picture, while it's MSE in second picture.

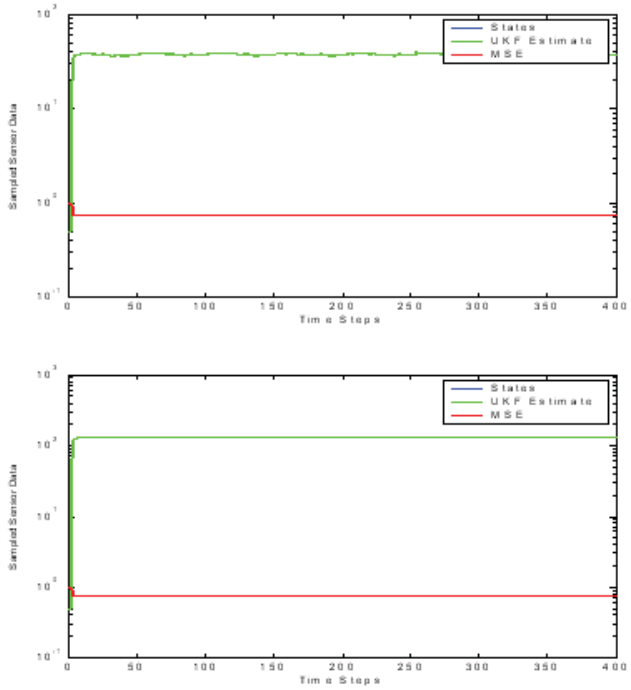


Fig. 5. Master Filter Estimation for X_1 and X_2

A multisensory data fusion algorithm based on decentralized UKF is computationally intensive, but it promise to utilize in real world. While simulation is carried out, a test bad of an multisensory data fusion is prepared, it consist of several sensors and actuators system. Which are attached to the two PCs was connected by wireless networked. A sensor system using 4 unit's rotary sensor AUTONICS ENB-500-3-1 act as odometer and its data is the measurement input for the local filter. Data from 8 units ultrasonic range sensor SHARP GP2Y0A02YK is dedicated to corresponding local filter. A GPS and all sensors dedicated to supply the local solutions to the main filter for yielding a global solution. Otherwise, an actuator using 4 unit DC motor drive to actuate the control command of steering, braking, accelerating and camera direction.

V. CONCLUSSIONS

An UKF based multisensor data fusion algorithm in decentralized application is computationally intensive, but it is promising to perform fault detection and isolation technique easier. An OCP is practical, that contains the same process model and makes their available for control designer, in order to draw from concepts to implementations.

The simulation results in this paper are not capture the whole system representation; we can say that the results is to simple. In the future work, we will utilize the true sensor data from the test bed of an online decentralized multisensor data fusion that have been under preparation. This is way to validate the algorithm have been proposed in this paper.

REFERENCES

[1] Tariq Samad, Gary Balas, "*Software-Enabled Control: Information Technology for Dynamic Systems*", IEEE Press, John Willey & Sons Inc. 2003.

[2] A. Gelb, "*Applied Optimal Estimation"*, MIT press, Cambridge Mass, 1974.

[3] P. S. Mayback, "*Stochastic Models, Estimation and Control"*,Academic Press, 1979.

[4] Jassemi-Zargani R., Necsulescu D., "Extended Kalman filter-based sensor fusion for operational space control of a robot arm", IEEE Transactions on Instrumentation and Measurement, Volume 51, Issue 6, Dec. 2002

[5] Sasiadek J.Z., Hartana P., "Sensor data fusion using Kalman filter", Proceedings of the Third International Conference on Information Fusion, FUSION 2000, Volume 2, 10-13 July 2000.

[6] S. J. Julier and J. K. Uhlmann, "A new extension of the kalman filter to nonlinear systems," Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls,Orlando, Florida, 1997.

[7] Wiston Li, Henry Leung, "Simultaneous registration and fusion of multiple dissimilar sensors for cooperative driving", IEEE Transactions on Intelligent Transportation Systems, Volume 5, Issue 2, June 2004

[8] Jamshaid Ali, Fang Jiancheng, "Multisensor Data Fusion using Federated Form of Unscented Kalman Filtering", Proceedings of IEEE International Conference on Industrial Technology, ICIT 2005, 14-17 Dec. 2005.

[9] Richardson J., Schumann J., Fischer B., Denney E., "Rapid Exploration of the Design Space During Automatic Generation of Kalman Filter Code", Proceedings of IEEE Conference Aerospace, 05-12 March 2005.

Robust Residual Generation for Actuator Fault Isolation

A Case Study: Magnetic-Tape-Drive MIMO System

Samiadji Herdjunanto, Adhi Susanto and Oyas Wahyunggoro
Department of Electrical and Information Technology Faculty of Engineering Gadjah Mada University
Jalan Grafika 2, Yogyakarta 55281, Indonesia
Tel. +62-274-90-2202, fax. +62-274-552305
e-mail: samiadji@jteti.gadjahmada.edu

***Abstract*— This paper concerns with robust residuals generation such that the designed signatures can be used to isolate actuator faults simultaneously and is not sensitive to exponential autocorrelation stochastic disturbance signal which contaminates one of plant's inputs. The proposed solution of this problem consists of two stages. The first stage is to derive an internal model of stochastic disturbance signal. The internal model is then augmented to the original plant's model so that a new plant's model is obtained. The second stage is to generate a transformation matrix such that each original feature vector of actuator fault is converted into the corresponding designed directional residual with its vector direction as its signature. Based on the designed of signature vectors the modes of residual actuator fault can be decoupled one from others and robust to the stochastic disturbance signal. The transformation matrix is constructed row by row in order each row can block the disturbance signal so that strict robustness can be achieved. To demonstrate the effectiveness of the method a simulation study on a Magnetic –Tape-Drive MIMO system which has multi input and multi output is excuted. The result shows that the method has been successfully implemented.**

Keywords—robust; fault isolation; internal model; transformation matrix; signature vector

I. INTRODUCTION

Systems approaches have been gaining in popularity in a number of different fields, including social, economic, information, and engineering. A green system can be defined as an assemble or combination of elements or parts which to coorperate each other to reach clearly defined green system objective.Many aspects of green design were discussed in [1]. Computation time reduction as a strategy to reduce energy consumption in green design is also investigated [2]. Futhermore, energy prediction in green buildings is given in recent paper [3]. One of green objectives among others is that to reduce the amount of material so that contamination of environment due to broken devices can be minimized. Nowdays it can be observed that hardware redundancy is still used to maintain the reliability of some industrial systems. In other words hardware redundancy is used for solving fault detection and isolation problem. Unfortunately this method is expensive and needs a lot of material for implementation. Therefore analytical redundancy gradually is replacing the former method. This approach has been gaining in popularity as an alternative method of solving this problem. This phenomenon can be shown by observing a lot of publications based on this new concept [4]-[6]. Some application papers also have been written [7]-[9]. Variety of methods can be seen from several papers [10]-[17].An interesting method among others is that it uses vector's direction as a signature to isolate an actuator fault instead of using fault magnitude. In other word this scheme is not tuned to certain mode of fault signal. A particular filter based on this concept is called a detection filter. From observation from many publications about detection filter it is seldom found a discussion about robustness of detection filter [18]-[23]. This unfortunate situation is caused ,one out of several reasons, by certain property namely mutually detectable which is hard to satisfy in general. Therefore issue of robust filters based on the concept directional vector for actuator fault is still needed.

In this paper a contribution to solve the robustness problem is proposed so that more than one actuator fault can be isolated independently from each other and robustness of all residuals against a stochastic disturbance signal with exponential autocorrelation function which enters into one out of several channels of plant's inputs can be achieved. The novelty of the proposed method is that to interweave new plant's model with a transformation matrix which converts an original feature vector into a designed directional residual which is called signature vector. The new plant's model is the result of augmentation of internal model of disturbance to the original plant's model. Then based on this new plant's model a transformation matrix is created row by row in order each row will block disturbance signal so that the mode of signature vectors of actuator faults are not contaminated by disturbance signal.

II. THEORY

A. New Plant's Model Contaminated by a Disturbance Signal

There are three types of uncertainties namely unstructured ,structured and disturbance. In this paper the uncertainty come from a disturbance signal which enter into the first plant's

input. Moreover in this paper the disturbance has exponential autocorrelation stochastic signal. The state model of a plant with its first input channel contaminated by a disturbance v(t) can be shown, by the following

$$\dot{\mathbf{x}}_n(t) = A_n\mathbf{x}_n(t) + B_n\mathbf{u}(t) + b_1(u_1(t) + v(t)) \quad (1)$$

Internal model of a stochastic disturbance v(t) can be thought as the output of a dynamic system driven by a white noise w(t). The general form of state equation of internal model of stochastic disturbance signal can be seen , by the following

$$\dot{d}(t) = A_d d(t) + w(t)$$

$$v(t) = C_d d(t) \quad (2)$$

The internal model of stochastic disturbance signal is then augmented to the old plant's model so that the new state space model of the plant can written , as in

$$\begin{bmatrix} \dot{\mathbf{x}}_n(t) \\ \dot{d}(t) \end{bmatrix} = \begin{bmatrix} A_n & b_1C_d \\ 0 & A_d \end{bmatrix}\begin{bmatrix} \mathbf{x}_n(t) \\ d(t) \end{bmatrix} + \begin{bmatrix} b_1 & B_n \\ 0 & 0 \end{bmatrix}\begin{bmatrix} u_1(t) \\ \mathbf{u_n}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(t) \quad (3)$$

B. Model of a Stochastic Disturbance Signal[24]

The autocorrelation function of the stochastic disturbance signal is assumed known.The state space model of the stochastic disturbance can be derived from the knowledge of this autocorrelation function $R_d(\tau)$.

$$R_d = e^{-k|\tau|} \quad (4)$$

Fourier transform of this autocorrelation function $R_d(\tau)$ is equal to the power spectral density $P_v(\omega)$ of stochastic signal v(t). This power spectral density $P_v(\omega)$ is equal to the multiplication of H(s) and H(-s) with s equal to jω. The transfer function H(s) has its output stochastic disturbance V(s) and its input white noise W(s) in s-plane. Then H(s) is chosen based on pole location of $P_v(s)$ which is located on left half plane of s-plane.

$$F\{R_d(\tau)\} = P_v(\omega) = H(s)H(-s) = \frac{\sqrt{2k}}{(s+k)}\frac{\sqrt{2k}}{(s-k)}$$

$$H(s) = \frac{\sqrt{2k}}{(s+k)} \quad (5)$$

State variable realization of H(s) based on phase variable form is shown , as in

$$\dot{d}(t) = -kd(t) + w(t)$$

$$v(t) = \sqrt{2k}d(t) \quad (6)$$

C. Analysis for robust actuator fault isolation[25]

From equations (2),(3) and(6) it can be observed that the final state space model of the plant with actuator fault and stochastic disturbance signal can be represented, as in

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{u}(t) - \mathbf{f}_a\boldsymbol{\mu}(t) + Mw(t)$$

$$A = \begin{bmatrix} A_n & b_1\sqrt{2k} \\ 0 & -k \end{bmatrix}; B = \begin{bmatrix} b_1 & B_n \\ 0 & 0 \end{bmatrix}; M = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (7)$$

If an actuator fault is related to the first input channel of the plant then f_a is equal b_1. Furthermore matrix M is related to white noise which excites the internal model of stochastic disturbance . When the size of matrix A_n is (n-1) then the size of matrix A is n.

A filter is made as a tool to get error signal which is defined by e(t)=s(t)-Tx(t) where T is called a transformation matrix.The dynamic model of the filter is given, as in

$$\dot{\mathbf{s}}(t) = \mathbf{E}\mathbf{s}(t) + \mathbf{F}\mathbf{y}(t) + \mathbf{G}\mathbf{u}(t) \quad (8)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t); \mathbf{C} = \begin{bmatrix} \mathbf{C}_1 = \mathbf{I}_m & \mathbf{0} \end{bmatrix} \quad (9)$$

The error signal is not discarded but it is kept because it contains rich of information about actuator fault signal. In order to investigate the error signal thoroughly the dynamic model of error signal is developed and shown , as in

$$\begin{aligned} \dot{\mathbf{e}}(t) &= \dot{\mathbf{s}}(t) - \mathbf{T}\dot{\mathbf{x}}(t) = \mathbf{E}\mathbf{s}(t) + \mathbf{F}\mathbf{C}\mathbf{x}(t) + \mathbf{G}\mathbf{u}(t) - \mathbf{T}\mathbf{A}\mathbf{x}(t) \\ &- \mathbf{T}\mathbf{B}\mathbf{u}(t) - \mathbf{T}\mathbf{M}w(t) + \mathbf{T}\mathbf{f}_a\mu(t) \end{aligned} \quad (10)$$

According to [25] the final form of state equation of error signal is given, as in

$$\begin{aligned} \dot{\mathbf{e}}(t) &= \mathbf{E}\mathbf{e}(t) + (\mathbf{E}\mathbf{T} + \mathbf{F}\mathbf{C} - \mathbf{T}\mathbf{A})\mathbf{x}(t) + (\mathbf{G} - \mathbf{T}\mathbf{B})\mathbf{u}(t) \\ &- \mathbf{T}\mathbf{M}w(t) + \mathbf{T}\mathbf{f}_a\mu(t) \end{aligned} \quad (11)$$

Note that f_a is equal b_1 if actuator fault come from the first channel of plant's inputs. Moreover from equation (11) it can be seen that error response depends on plant's input u(t),state of the plant x(t),white noise w(t) and fault signal μ(t). Therefore in order error response depends only on fault signal some constraints must be fulfilled [25], as in

$$\mathbf{T}\mathbf{M} = \mathbf{0} \quad (12)$$

$$\mathbf{E}\mathbf{T} + \mathbf{F}\mathbf{C} - \mathbf{T}\mathbf{A} = \mathbf{0} \quad (13)$$

$$\mathbf{G} - \mathbf{T}\mathbf{B} = \mathbf{0} \quad (14)$$

Note that matrix A contains internal model of stochastic disturbance signal. If equations (12), (13) and (14) are satisfied then the final equation can be written , as in

$$\dot{\mathbf{e}}(t) = \mathbf{E}\mathbf{e}(t) + \mathbf{T}\mathbf{f}_a\mu(t) \tag{15}$$

It is assumed that only m states can be measured. Therefore matrices C and A can be expressed , as in

$$\mathbf{C}=[\mathbf{C}_1\ \mathbf{0}]\ ;\mathbf{A}=[\mathbf{A}1\quad \mathbf{A}2] \tag{16}$$

Note that the size of matrix **A**1 is n by m and the size of matrix **A**2 is n by (n-m). If equation (13) is scrutinized it can be broken down into two separate equations [26] , as in

$$[\mathbf{TA}-\mathbf{ET}]\begin{bmatrix}\mathbf{I}_m\\ \mathbf{0}\end{bmatrix} = \mathbf{F}\mathbf{C}_1 \tag{17}$$

$$[\mathbf{TA}-\mathbf{ET}]\begin{bmatrix}\mathbf{0}\\ \mathbf{I}_{n-m}\end{bmatrix} = \mathbf{0} \tag{18}$$

Eigenvalues of matrix E is free to be chosen as far as they are located on left half s-plane so that the filter is stable. The ith *eigenvalue* of matrix E and the ith row of matrix T have a relationship , as in

$$\mathbf{t}_i^T\left[\mathbf{A}\,2 - \lambda_i\begin{bmatrix}\mathbf{0}\\ \mathbf{I}_{n-m}\end{bmatrix}\right] = \mathbf{0} \tag{19}$$

Equation (19) can be written concisely , as in

$$\mathbf{t}_i^T[\mathbf{L}_i] = 0 \tag{20}$$

Therefore $\mathbf{t}_i^T$ is located at *left nullspace* of matrix $\mathbf{L}_i$. In order to find $\mathbf{t}_i^T$ it is required to compute basis of *left nullspace* of matrix $\mathbf{L}_i$. Equation (21) is used to find each of m members of the basis of *left nullspace* of matrix $\mathbf{L}_i$.

$$\mathbf{h}_{ij}^T\left[\mathbf{A}2 - \lambda_i\begin{bmatrix}0\\ \mathbf{I}_{n-m}\end{bmatrix}\right] = \mathbf{0} \quad \text{for j=1,…….m} \tag{21}$$

Matrix $\mathbf{L}_i$ is not square therefore it is required two *orthonormal* transformation matrices so that *left nullspace, column space, row space and nullspace* can be explicitely shown. This can be achieved through SVD (*Singular Value Decomposition*) concept [27].

In order to obtain robustness property $\mathbf{t}_i^T$ should satisfy both equation (20) and equation (12) . Therefore it is concluded that $\mathbf{t}_i^T$ is *intersection* between *left nullspace* of matrix $\mathbf{L}_i$ and *left nullspace* of matrix **M** [25], as in

$$\mathbf{t}_i^T = \mathbf{L}_i^{\perp} \cap \mathbf{M}^{\perp} = [\,\mathbf{L}_i\ \ \mathbf{M}\,]^{\perp} \tag{22}$$

If the number of basis members of $[\,\mathbf{L}_i\ \ \mathbf{M}\,]^{\perp}$ are r for ith e*igenvalue* of matrix E then $\mathbf{t}_i^T$ can be expressed as a linear combination of its basis members, as in

$$\mathbf{t}_i^T = c_{i1}\mathbf{h}_{i1}^T + c_{i2}\mathbf{h}_{i2}^T + \ldots\ldots + c_{ir}\mathbf{h}_{ir}^T \tag{23}$$

Matrix transformation T will convert original feature vectors of actuator fault $\mathbf{f}_a$ to signature vectors Q which is chosen as orthonormal vectors so that they are independent each other . If the number of actuator faults are p then matrix Q is given [25], as in

$$\mathbf{Q} = \mathbf{I}_p = \mathbf{T}\mathbf{f}_a \tag{24}$$

The coefficients $c_{i1},c_{i2,}\ldots,c_{ir}$ can be computed from ith row of (24) , as in

$$[c_{i1}\ \ .\ \ .\ \ c_{ir}]\begin{bmatrix}\mathbf{h}^T{}_{i1}\\ .\\ .\\ \mathbf{h}^T{}_{ir}\end{bmatrix}[\mathbf{f}_{a1}\ \ .\ \ \mathbf{f}_{ap}] = [q_{i1}\ \ .\ \ .\ \ q_{ip}] \tag{25}$$

Other coefficients from another row of (24) can be solved similarly to that one in (25).

III. Simulation Result

Since magnetic-tape drive MIMO system has multi input and multi output therefore it is chosen as a case study to demonstrate the effectiveness of the proposed method. State space equations of the magnetic-tape drive MIMO system is given below [28], as in

$$\dot{\mathbf{x}}_n(t) = A_n\mathbf{x}_n(t) + B_n\mathbf{u}(t) \qquad \mathbf{y}(t) = \mathrm{C}_n\mathbf{x}(t)$$

$$A_n = \begin{bmatrix} 0 & 0 & -r & 0 \\ 0 & 0 & 0 & r \\ \frac{Kr}{J} & -\frac{Kr}{J} & -\frac{br^2}{J} & -\frac{br^2}{J} \\ \frac{Kr}{J} & -\frac{Kr}{J} & -\frac{br^2}{J} & -\frac{br^2}{J} \end{bmatrix} \qquad B_n = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{K_m}{J} & 0 \\ 0 & \frac{K_m}{J} \end{bmatrix};$$

$$C_n = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

where
K_m is motor torque constant.
J is motor and capstan inertia.
K is tape spring constant.
R is capstan radius.
b is tape damping constant.

$\mathbf{x_n} = [x_1 \quad x_2 \quad x_3 \quad x_4]^T$
x_1 ,x_2 = tape positions at capstan
x_3 ,x_4 = angular rates of motor
u(t) are inputs
While parameter of the plant is given below:
K_m =0,544 N-m/A
J =0,006375kg-m^2
K =2113N/m
R =0,1 m
b = 3,75 N-sec/m

Autocorrelation function $R_d(\tau)$ is given, as in

$$R_d = e^{-32|\tau|}.$$

In this simulation the stochastic disturbance signal contaminate the first plant's input. Therefore from (7) the new plant's model with internal model of disturbance inside matrix A is given, as in

$$A = \begin{bmatrix} 0 & 0 & -0{,}1 & 0 & 0 \\ 0 & 0 & 0 & 0{,}1 & 0 \\ 35216 & -35216 & -477.38 & -501.2 & 682.4 \\ 35216 & -35216 & -477.38 & -501.2 & 0 \\ 0 & 0 & 0 & 0 & -32 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 85.3 & 0 \\ 0 & 85.3 \\ 0 & 0 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

From matrix C it can be seen that only 3 states are measured therefore m=3 . From the given autocorrelation function it can be seen that k=32.

In this simulation there are two actuator faults therefore p=2 and from (24) matrix $Q=I_2$. Eigenvalues of matrix E are free chosen on left half s-plane therefore in this simulation they are set to s= -10 and s= -12 so that matrix E is stable..Therefore

$$E = \begin{bmatrix} -10 & 0 \\ 0 & -12 \end{bmatrix}; \; Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \text{ m=3; k=32}$$

Original feature fault vectors are :

$$f_{a1} = b_1 = [0 \quad 0 \quad 85.3 \quad 0 \quad 0]^T$$
$$f_{a2} = b_2 = [0 \quad 0 \quad 0 \quad 85.3 \quad 0]^T.$$

From equation (7) it has been shown that internal model of stochastic disturbance is located at the end row of matrix A therefore matrix M is:

$$M = [0 \quad 0 \quad 0 \quad 0 \quad 1]^T.$$

For robust isolation actuator fault the transformation matrix can be computed from (23) and (25)

$$T = \begin{bmatrix} 7.47x10^{13} & 85.82 & 0.01178 & 0 & 0 \\ -62.44 & 57.36 & 0 & 0.01178 & 0 \end{bmatrix}$$

Matrix G can be calculated from (14)

$$G = \begin{bmatrix} 1 & -0.00016 \\ 0 & 1 \end{bmatrix}$$

Using equation (17) matrix F is computed:

$$F = \begin{bmatrix} 7.47x10^{14} & 445.4 & -74.73x10^{9} \\ -336.5 & 275.4 & 0.6 \end{bmatrix}$$

In first simulation study the first actuator fault signal is a bias signal which is contaminated by an exponentially autocorrelated stochastic disturbance signal. In the first simulation study the filter is not designed for robust actuator isolation.The purpose of the simulation is to investigate the response of first residual.The result is shown in Fig .1. From Fig .1, it is difficult to recognize the bias actuator fault signal because the first residual is not robust to stochastic disturbance signal .

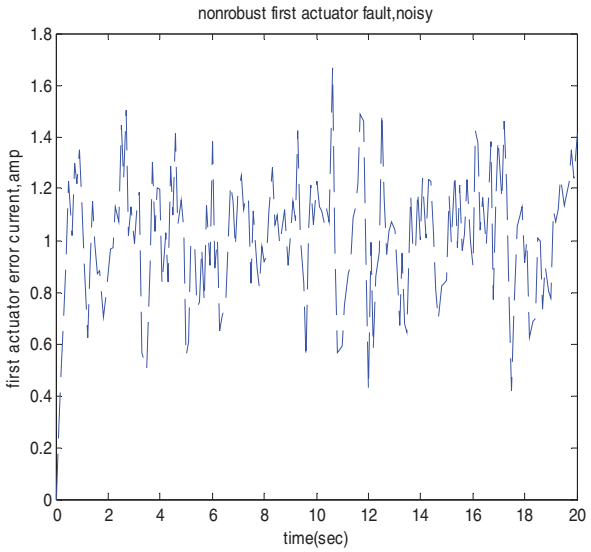


Fig 1. 1st residual mode of actuator error signal with direction $[1 \quad 0]^T$ which is contaminated by stochastic disturbance.

In second simulation study the condition of first input is similar to the one in the first simulation study but the filter is designed for robust actuator isolation. Moreover the second actuator fault is a drift signal which its magnitude is proportional to time. The result is shown in Fig. 2 and Fig. 3. The mode of first residual is shown in Fig. 2. From Fig 2, it can be seen that the first residual mode is clean from stochastic disturbance signal so that a bias fault signal can be seen clearly. Moreover the first residual mode is successfully decoupled from the second actuator fault signal which is a drift signal.

Fig. 3 shows the result of second residual mode which is clearly seen as a drift signal. This signal is not contaminated by stochastic disturbance signal. Furthermore the second residual mode is successfully decoupled from first residual mode.

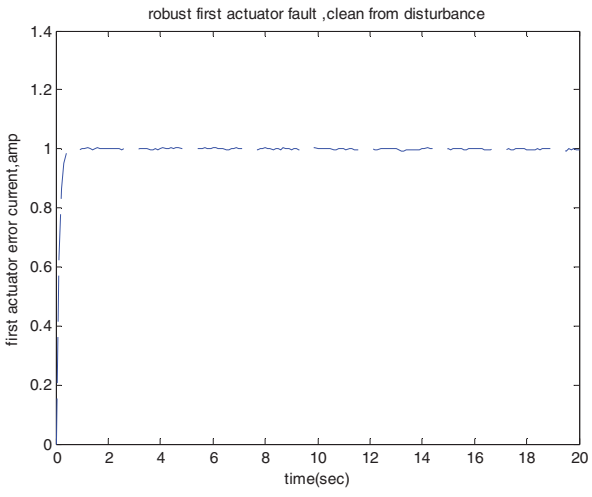


Fig 2. 1st residual mode of actuator error signal with direction $[1 \quad 0]^T$ which is clean and decoupled from second residual mode

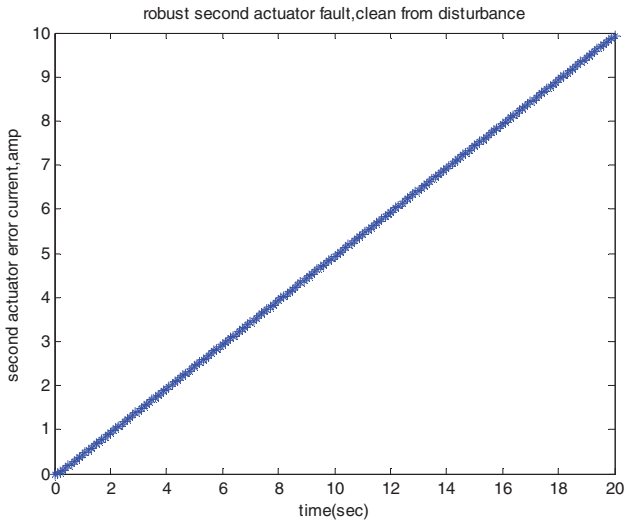


Fig 3. 2nd residual mode of actuator error signal with direction $[0 \quad 1]^T$ which is decoupled from first residual mode

IV. CONCLUSION

Robust residual generation for actuator fault isolation is successfully developed and simulated on magnetic-tape drive MIMO system. Each residual mode is not contaminated by stochastic disturbance. Moreover both residuals are decoupled each other. Since this robust residual generation is based on the concept of analytical redundancy therefore the amount of material for implementing this design is reduced to support green system objective.

REFERENCES

[1] M.A Hersh,1998,"*A Survey of Systems Approaches to Green Design with Illustrations from the Computer Industry*", IEEE Transactions on systems, Man, and Cybernetics-Part C: Apllications and Reviews, vol. 28 no. 24.

[2] S.Herdjunanto,2010,*ComputationTime Reduction as a Strategy to Reduce Energy in Green Design, A case study:Model Parameter Estimation on Unwinding Part of a Web Winding System,Proceedings of ICGC-RCICT* ,pp390-395.

[3] D. Holcomb,2009"*Algorithms for Green Buildings:Learning-Based Techniques for Energy Prediction and Fault Diagnosis,*", Electrical Engineering and Computer Sciences University of California at Berkeley.

[4] Deckert, J.C., Desai, M.N., Deyst, J.J., Willsky, A.S.,1977, *F-8 DFBW Sensor Failure Identification Using Analytic Redundancy,*IEEE Trans on Automatic Control,vol.AC-22,no.5,pp795-803.

[5] Ding, X.,Guo, L., Jeinsch, T.,1999, *A Characterization of Parity Space and Its Application to Robust Fault Detection,* IEEE Trans on Automatic Control,vol.AC-44,no.2,pp 337-343.

[6] Zhong, M., Ding, S.X., Han, Q.L., Ding, Q.,2010, *Parity Space-Based Fault Estimation for Linear Discrete Time-Varying Systems,*IEEE Trans on Automatic Control,vol.AC-55,no.7,pp1726-1731.

[7] Gu, D.W., Poon, F.W.,2003, *A Robust Fault-Detection Approach with Application in a Rolling –Mill Process.* IEEE Trans on Control Systems Technology,vol.11,no.3,pp 408-414.

[8] Kallesoe, C.S., Cocquempot, V., Zamanabadi, R.I.,2006, *Model Based Fault Detection in a Centrifugal Pump Application,* IEEE Trans on Control Systems Technology,vol.14,no.2,pp 204-215.

[9] J.M.Koscielny,M.Bartys,M.Syfert,2012,*Method of Multiple Fault Isolation in Large Scale Systems,*IEEE Trans on Control Systems Technology,vol.20,no.5,pp1302-1310

[10] S.A.Arogeti,D.Wang,C.B.Low,M.Luo,2013,*Energy-Based Mode Tracking of Hybrid Systems for FDI,* IEEE Transactions on Systems,Man ,and Cybernetics:Systems, vol 43, no.1, pp14-28.

[11] Demetriou, M.A., Polycarpou, M.M.,1998, *Incipient Fault Diagnosis on Dynamical Systems Using Online Approximators,* IEEE Trans on Automatic Control,vol.AC-43,no.11,pp1612-1617.

[12] Huang, S. and Tan, K.K.,2009, *Fault Detection and Diagnosis Based on Modeling and Estimation Methods,*IEEE Trans on Neural Network ,vol.20,no.5,pp 872-881.

[13] Patan, K., Witcz, M., Korbicz, J.,2008, *Towards Robustness in Neural Network Based Fault Diagnosis,*Int.J.Appl.Math.Comput.Sci,vol.18,no.4,pp 443-454.

[14] Perhinschi, M.G., Napolitano, M.R., Campa, G., Seanor, B., Burken, J., Larson, R.,2006, *An Adaptive Threshold Approach for the Design of an Actuator Failure Detection and Identification Scheme,* IEEE Trans on Control Systems Technology,vol.14,no.3,pp 519-525.

[15] Vemuri, A.T., Polycarpou, M.M.,1997, *Neural-Network-Based Robust Fault Diagnosis in Robotic Systems ,* IEEE Trans on Neural Networks, vol.8,no.6,pp 1410-1420.

[16] Zhang, X.,2011, Sensor Bias Fault Detection and Isolation in a class of Nonlinear Uncertain Systems Using Adaptive Estimation,IEEE Trans on Automatic Control ,vol.AC-56,no.5,pp 1220-1226.

[17] Massoumnia, M.A.,1986, *A Geometric Approach to the Synthesis of Failure Detection Filters,* IEEE Trans on Automatic Control,vol.AC-31,no.9,pp 839-846.

[18] Massoumnia, M.A., Verghese, G.C., Willsky, A.S.,1989, *Failure Detection and Identification,* IEEE Trans on Automatic Control,vol.AC-34,no.3,pp 316-321.

[19] White, J.E., Speyer, J.L.,1987, *Detection Filter Design : Spectral Theory and Algorithms ,* IEEE Trans on Automatic Control,vol.AC-32,no.7,pp 593-603.

[20] Chen, J., Patton, R.J., Zhang, H.Y.,1996, *Design of unknown input observers and robust fault detection filters,*Int.J.Control.vol .63,no.1, pp 85-105.

[21] Kim, Y., Park, J.,2003, *A Condition of the Eigenvalues of Detection Filters for Disturbance Attenuation: An Invariant Zero Approach*,Proc.42nd IEEE Conf. Decision Control,pp 475-479.

[22] Kim, Y., Park, J.,2005, *On The Approximation of Fault Directions for Mutual Detectability:An Invariant Zero Approach*, IEEE Trans on Automatic Control,vol.AC-50,no.6,pp 851-855.

[23] Meskin, N. and Khorasani, K.,2009, *Actuator Fault Detection and Isolation for a Network of Unmanned Vehicles,* IEEE Trans on Automatic Control ,vol.AC-54,no.4,pp 835-840.

[24] K.S.Shanmugan,A.M.Breipohl,1988,*Random Signals,Detection,Estimation and Data Analysis,John Wiley&Sons,Inc*

[25] S.Herdjunanto,*Reduced Order Filter for Robust Actuator Fault Isolation,*unpublished

[26] Tsui, C.C.,1987, A Complete Analytical Solution to the Equation TA-FT=LC and Its Applications,IEEE Trans on Automatic Control,vol.AC-32.no.8,pp 742-744.

[27] Klema, V.C. and Laub, A.J.,1980, *The Singular Value Decomposition:Its Computation and Some Applications,* IEEE Trans on Automatic Control, vol.AC-25,no.2,pp 164-176.

[28] G.F.Franklin,J.D.Powell,M.Workman,1998,*Digital Control of Dynamic Systems*,3rd edition,Addison Wesley Longman,Inc

Study of Fuzzy Logic Control and Power System Stabilizers Effect on the Stability Enhancement of a SMIB Power System

F. Mayouf[1,3*](Adjeroud), F. Djahli[2], A. Mayouf[3], T. Devers[4]

[1]Department of Electrical Engineering, University of Sétif 1, Algeria.
[2]Department of Electronics, University of Sétif 1, Algeria.
[3]DIMMER Laboratory, University of Djelfa, Algeria.
[4]Department of Industrial Engineering and Maintenance, University of Orléans, France
[*]fa.adjeroud@gmail.com

Abstract— **In a previous work, power system stabilizers (PSS) have been employed in the Excitation and/or in the turbine Governor systems (EPSS, GPSS and EGPSS) for improving more the stability of a Single Machine Infinite-Bus power system (SMIB). Basing on obtained results, the employment of PSS both in excitation and governor systems (EGPSS) has improved more the system stability. In order to more enhance stability and overcome the drawbacks of conventional PSS, we studies in this paper the effect of the implementation of the fuzzy logic controller (FLC) into the excitation and/or turbine governor systems (FLCE, FLCG, FLCEG). Obtained results, by nonlinear simulation using Matlab/Simulink of a SMIB, show the effectiveness of using Fuzzy logic controller both in excitation and governor systems (FLCEG) for large and small disturbances. Our results concern: rotor angle (δ), terminal voltage (V_t), electrical torque (T_e) and speed deviation (Δω) for the four cases: open loop (without PSS), EPSS, GPSS and EGPSS.**

Keywords-modeling and simulation; Fuzzy logic controller; power system stabilizers; SMIB; governor-turbine stabilizer.

I. INTRODUCTION

Large electrical power systems consists of many complex nonlinear systems connected together to form a large, complex and dynamic system capable of generating, transmitting and distributing electrical energy over a large geographical area. They are often subjected to low frequency electromechanical oscillations due to insufficient damping caused by adverse operating. These oscillations with small magnitude and low frequency often persist for long periods of time and in some cases they even present limitations on power transfer capability [1]. The secure operation of power systems, therefore, has become a major concern, and the applications of PSS for dynamic stability enhancement have drawn more attention than ever before. The investigation of power system stabilizer for improving the power system stability has been lead since the late 1960 [2-10].

The power utilities worldwide are currently implementing power system stabilizers as effective controllers for improving power system stability**.** The action of a PSS is to extend the angular stability limits of a power system by providing supplemental damping to the oscillation of synchronous machine rotors through the generator excitation. This supplementary control is very beneficial during line outages and large power transfers [3-13].

Generator excitation controls are a basic stability control, but it is not the unique tool. The stability of the synchronous generator can be enhanced either by controlling the excitation and hence the field current, or by controlling the governor which provides the adjustment of the mechanical torque [5-11].

Including a PSS signal in the turbine governing systems with the aim of improving damping is not new. In 1972 and for the first time some solutions regarding hydro-turbines have been introduced. The principle of providing an additional damping torque from the turbine governor is similar to that used in EPSS. The time constants in the turbine governor introduce a phase shift between the oscillations in the speed deviation (Δω) and the turbine mechanical power [12].

The main advantage of applying a GPSS lies in the fact that the turbine governor dynamics are weakly coupled with those of the rest of the system. Consequently, the parameters of the GPSS do not depend on the network parameters. Wang et al. [5,6,9] show interesting simulation results for systems equipped with a GPSS.

Since power systems are highly nonlinear systems, with configurations and parameters which alter through time, the Conventional PSS (CPSS) design based on the linearized model of the power system cannot guarantee its performance in a practical operating environment. To overcome this drawback, several other categories of stabilizers which consider the nonlinear nature of the plant and adapt to the changes in the environment have been proposed in the literature. In order to improve the performance of CPSS, a variety of controllers have been employed to manipulate it, including optimal controller, adaptive self-tuning, variable structure, artificial neural network based PSS and fuzzy logic controllers [1,14-21].

In recent years, fuzzy logic has emerged as a powerful tool and is starting to be used in various power system applications. The implementation of FLC appears to be the most suitable one whenever a well-defined control objective cannot be specified, the system to be controlled is a complex one, or its exact mathematical model is not available. Due to their robustness, low cost, simplicity and ability to cope with

nonlinear, complex system, fuzzy logic power system stabilizers have been proposed for effective damping of power system oscillations [14-21].

In previous works, PSS have been employed in the EPSS, GPSS and EGPSS for improving the stability of a SMIB. Basing on obtained results, the employment of EGPSS has improved more the system stability [22-25].

In order to further enhance stability and overcome the drawbacks of conventional PSS, we have used a fuzzy logic controller (FLC) as a power system stabilizer.

In this paper, we study the implementation effect of FLC on the stability enhancement in comparison with EGPSS. The developed FLC has two inputs: speed deviation ($\Delta\omega$) and its derivative. So as to obtain the ideal area if the implementation of the developed FLC, we have opted to employ it into excitation and/or turbine-governor systems (FLCE, FLCG, and FLCEG).

To show effectiveness of the FLCEG implemented both in excitation and turbine-governor systems, a nonlinear simulation using Matlab/Simulink of a SMIB has been tested under small and large perturbations. The proposed general control model of SMIB is implemented in Fig.1.

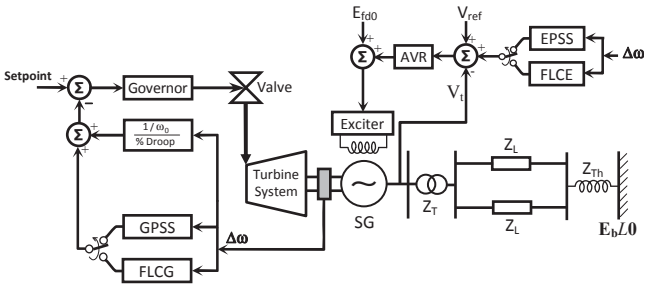


Figure 1. General control model of SMIB power system.

The main objectives of the research work presented in this paper are to present the effectiveness of the FLCEG implementation in the stability enhancement of SMIB power system, to investigate the dynamic performance of the system for all cases (EGPSS, FLCE, FLCG, FLCEG) and finally to compare the performances of the proposed controller (FLCEG) with the other stabilizers: EGPSS, FLCE, and FLCG.

II. MODELING

The SMIB power system considered in this study is shown in Fig. 1. The synchronous generator is driven by a hydraulic turbine equipped with a governor and connected to a large power system through a step-up transformer and double circuit transmission line. The parameters V_t and E_b are respectively the voltages of generator terminal and infinite-bus. Z_{Th}, Z_L and Z_T represent the impedances of Thevenin, transformer, and transmission line.

The control signal generated by the PSS is injected as a supplementary stabilizing signal to the AVR summing point. The system model has been designed to support the EGPSS, FLCE, FLCG and the FLCEG and can be easily extended to support other types of PSS.

For our simulations, the nonlinear model 1.1 of the generator has been used and static excitation system has been considered.

A. SMIB power system model

The model used for representing the synchronous generator is characterized by field circuit and one equivalent damper winding on q-axis. The machine equations are [7,8]:

$$\frac{d\delta}{dt} = \omega_b(S_m - S_{m0}) \tag{1}$$

$$\frac{dS_m}{dt} = \frac{1}{2H}\left[T_m - T_e - D(S_m - S_{m0})\right] \tag{2}$$

$$\frac{dE'_d}{dt} = \frac{1}{T'_{qo}}\left[-i_q(x_q - x'_q) - E'_d\right] \tag{3}$$

$$\frac{dE'_q}{dt} = \frac{1}{T'_{do}}\left[E_{fd} - E'_q + i_d(x_d - x'_d)\right] \tag{4}$$

The electrical torque T_e is expressed as follows:

$$T_e = E'_d i_d + E'_q i_q + (x'_d - x'_q) i_d i_q \tag{5}$$

In order to simplify the formulation, we consider Z_{eq} as the equivalent impedance between the generator and the infinite-bus expressed by:

$$Z_e = R_e + jx_e = Z_T + Z_{Th} + \frac{Z_L}{2} \tag{6}$$

By neglecting the ohmic losses in the stator, the formulation can be reduced as:

$$\begin{cases} v_d = E'_d - x'_q i_q = R_e i_d + x_e i_q - E_b \sin(\delta) \\ v_q = E'_q + x'_d i_d = R_e i_q - x_e i_d + E_b \cos(\delta) \end{cases} \tag{7}$$

Solving the system of equations (7) gives the variables i_d and i_q:

$$\begin{bmatrix} i_d \\ i_q \end{bmatrix} = \begin{bmatrix} x'_d + x_e & -R_e \\ -R_e & -(x'_q + x_e) \end{bmatrix}^{-1} \begin{bmatrix} E_b \cos(\delta) - E'_q \\ -E_b \sin(\delta) - E'_d \end{bmatrix} \tag{8}$$

The IEEE Type-ST1 excitation system is considered in this study:

$$\frac{dE_{FD}}{dt} = \frac{1}{T_A}\left[-E_{FD} + K_A(V_{ref} - V_t + V_{PSS})\right] \tag{9}$$

where K_A and T_A are the gain and time constant of the excitation system.

B. Exciter and Governor PSSs design

The block diagram of a speed input conventional lead-lag exciter-based stabilizer (EPSS) is shown in Fig. 2. The stabilizer contains a washout term, stabilizer gain, phase lead-lag compensation and output limiters. For the governor stabilizer (GPSS), it has the same structure except that the gain is larger that of the EPSS. In this study, the GPSS's gain (K_{GPSS}) is taken 5× K_{EPSS} which is the exciter stabilize gain.

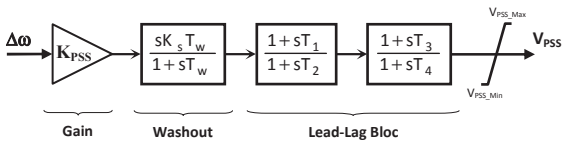


Figure 2. Lead-Lag Power System Stabilizer.

C. FLC design

A power system stabilizer based on the FLC algorithm has been developed. The proposed Fuzzy controller block diagram is given in Fig. 3. Since the goal of this application is to stabilize and improve the damping of the synchronous machine, speed deviation $\Delta\omega$ and acceleration $\Delta\dot{\omega}$, have been selected as the controller inputs. The controller output is then injected into the AVR summing point.

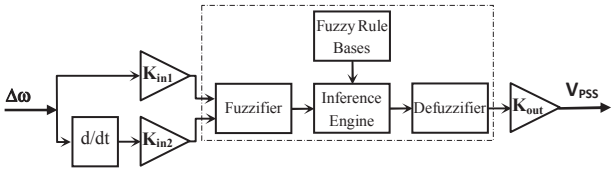


Figure 3. Lead-Lag Power System Stabilizer.

The main phases of the FLC are the fuzzification of controller inputs, rule definitions, rule inference, and defuzzifications.

The fuzzification requires the transformation of control variables to corresponding fuzzy variables. The input variables depend on the nature of the controlled systems. In this paper, we have considered $\Delta\omega$ and its derivative $\Delta\dot{\omega}$, as input signals to the proposed FLC. The fuzzified signals of FLC are interpreted into a number of linguistic variables. We have defined seven different variables namely: NB, NM, NS, Z, PS, PM and PB which stand for Negative Big, Negative Medium, Negative Small, Zero, Positive Small, Positive Medium, and Positive Big, respectively. Each linguistic variable has a label and a membership function to distinguish it from the others; the universe of discourse for each input and output is defined according to the controller designers; however, the universe of discourse or the range of fuzzified output variable of FLC must be limited to avoid saturation and cause system instability. In this paper we have used the most commonly fuzzy inference technique which is so-called Mamdani method [26,27]. The performance of FLC also depends upon the type of membership functions. The most membership functions used are trapezoidal, triangular and Gaussian [28].

In this work, we have chosen the triangular form defined as:

$$\mu_A(x) = \begin{cases} 0 & x \le a, \\ \dfrac{x-a}{c-a} & a \le x \le c \\ \dfrac{b-x}{b-c} & c \le x \le b \\ 0 & x \ge b, \end{cases} \tag{10}$$

Where a and b are the feet of the triangle and c is the peak.

Designing fuzzy logic control (FLC) requires the definition of the control rules. The control rules create relations between input and output fuzzy sets. Usually, they are derived based on different manners. In this paper, we have used the algorithm developed in [27] to generate the rule base with the help of a previous experience of the controlled system dynamics. Table 1 shows the fuzzy logic rules for the fuzzy inference system.

u(t)		NB	NM	NS	ZO	PS	PM	PB
d(Δω)/dt	**NB**	NB	NB	NB	NB	NB	NM	NS
	NM	NB	NB	NB	NM	NM	NS	ZO
	NS	NB	NM	NM	NS	NS	ZO	ZO
	ZO	NS	NS	ZO	ZO	ZO	PS	PS
	PS	ZO	ZO	PS	PS	PM	PM	PB
	PM	ZO	PS	PM	PM	PB	PB	PB
	PB	PS	PM	PB	PB	PB	PB	PB

Tab 1. Fuzzy rule bases.

This configuration implies that the FLC has two input parameters, K_{in1}, and K_{in2}, and one output parameter, K_{out}, as seen in Fig. 3. The selection of these parameters is obtained using a two steps method. These two steps consist of adjusting K_{in1}, and K_{in2} in fact to normalize input and then tuning K_{out} to obtain optimal result. The range of the scaling factors is chosen using the method described in [29].

III. NONLINEAR SIMULATION RESULTS

The dynamic performance of SMIB system has been analyzed with the proposed FLC (EGPSS, FLCE, FLCG, FLCEG), under various disturbances. The performance of the proposed power system stabilizer (FLCEG) has been examined under small perturbation and three-phase fault. Power system toolbox (PST), MATLAB has been used for the analysis [30,31]. The data of the simulated system are given below:

Synchronous generator:
X_d=1.7572 pu, X'_d = 0.4245 pu X_q=1.5845 pu, x'_q = 1.04 pu,
T'_{do} = 6.66 s, T'_{qo} = 0.44 s, H=3.542s, P_e=0.6 pu, Q_e=0.02224 pu.

Exciter: K_A=400, T_A=0.025s.

Transmission line (in pu): R_L=0.08593, X_L=0.8125 , B_c=0.1184.

Transformer: R_T=0 pu, X_T=0.1364 pu.

Thevenin: X_{Th}=0.13636 pu.

To investigate the effectiveness of FLCEG, we have simulated with EGPSS, FLCE, and FLCG a four-cycle three-phase fault (Fig. 4) and a step increase in excitation

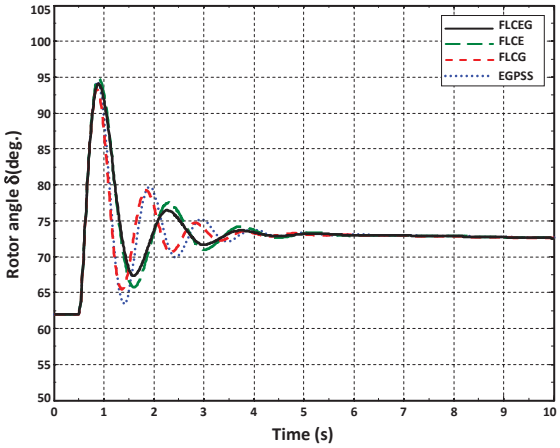
reference voltage (Fig. 5) in SMIB power system. Our results concern: rotor angle (δ), terminal voltage (V_t), electrical torque (T_e) and speed deviation (Δω) for the four cases: EGPSS, FLCE, FLCG, FLCEG.

Fig. 4 shows the system response to a four-cycle three-phase fault applied at the sending end of one of the circuits of the transmission line at time t=0.5s. The fault is cleared by tripping the faulty line. Fig. 5 presents the simulation of a 0.1pu step increase in V_{ref}.

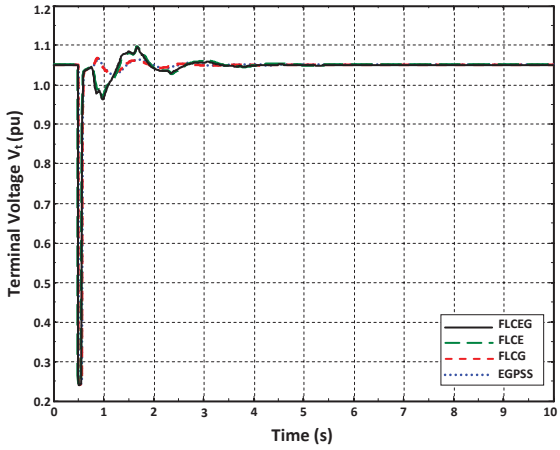
It could be observed for both cases of simulation, a good response with reduced oscillations of the fuzzy controller FLCEG. The system's stability is enhanced in this case. More important oscillations for the other stabilizers were observed (EGPSS, FLCE, FLCG) before reaching the steady state operation point.

On the other hand, it can be ascertained that the electromechanical oscillations are damped quickly in case of the proposed stabilizer (FLCEG) proving its superiority over the rest of the stabilizers: FLCE, FLCG and finally EGPSS. This confirms the effectiveness of the FLC in enhancement of the stability in SMIB power system.

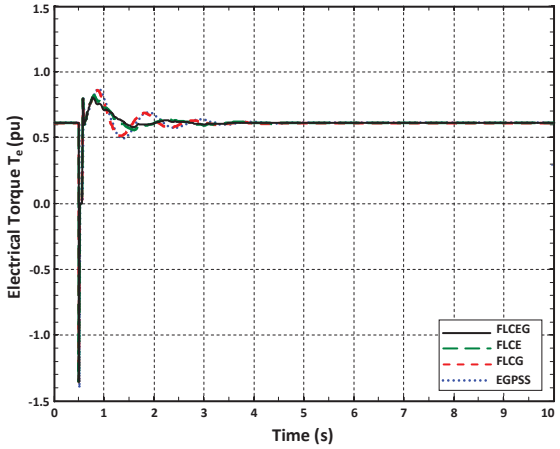
A. Large disturbance test



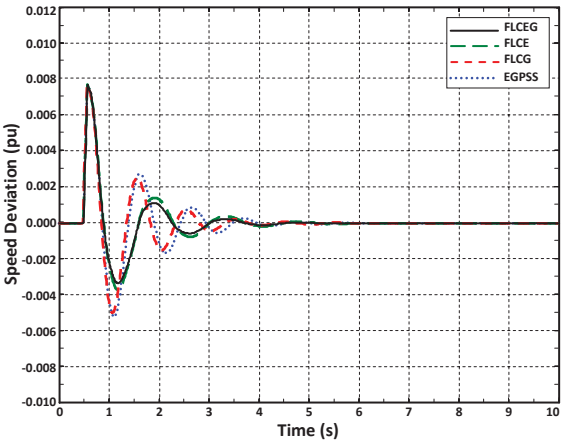
(a) Variation of rotor angle δ(deg.).



(b) Variation of terminal voltage V_t (pu).



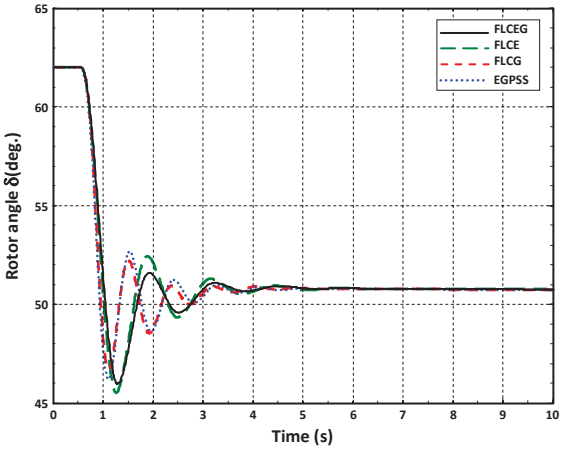
(c) Variation of electrical torque T_e (pu).



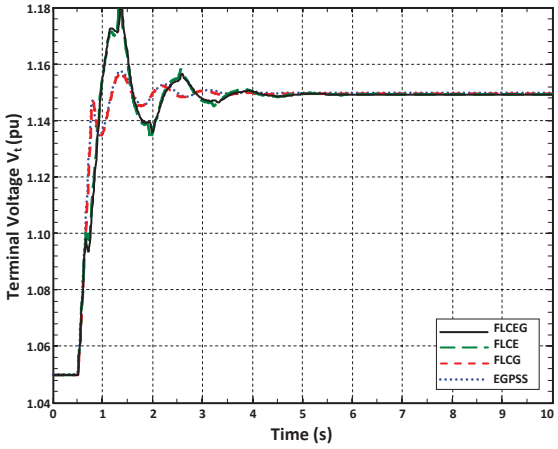
(d) Variation of speed deviation (pu).

Figure 4. System Response to a four-cycle three phase fault.

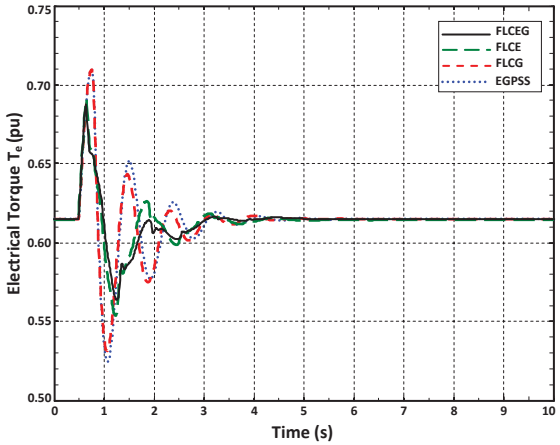
B. Small disturbance test



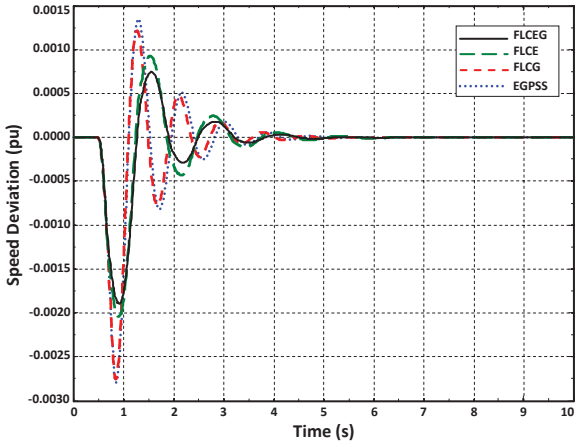
(a) Variation of rotor angle δ(deg.).



(b) Variation of terminal voltage V_t (pu).



(c) Variation of electrical torque T_e (pu).



(d) Variation of speed deviation (pu).

Figure 5. System Response to a 0.1pu step incease in V_{ref}.

IV. CONCLUSION

In this paper we have studied the effect of the implementation of the fuzzy logic controller (FLC) into the excitation and/or turbine governor systems (FLCE, FLCG, FLCEG) in order to more enhance stability and overcome the drawbacks of conventional PSS. Obtained results, show that the oscillations are damped quickly in case of the proposed stabilizer (FLCEG) proving its superiority over the rest of the stabilizers: FLCE, FLCG and finally EGPSS.

REFERENCES

[1] W. Liu, G. K. Venayagamoorthy, and D. C. Wunsch, "A heuristic dynamic programming based power system stabilizer for a turbogenerator in a single-machine power system", IEEE Trans. on industry applications, vol. 41, no. 5, pp. 1377-1385, 2005.

[2] J. I. Dineley, and S. E. Mikhail, "Effect of feedback stabilising signals in excitation and governors/turbine systems on transient and dynamic stability", IEEE. PES Winter Meeting, 1976.

[3] E. Larsen, and D. Swann "Applying power system stabilizers, parts I, II and III", IEEE Trans. Power App. Syst., vol. 100, pp. 3017 -3046, 1981.

[4] P. Kundur, M. Klein, G. J. Rogers, and M. S. Zywno, "Application of power system stabilizers for enhancement of overall system stability", IEEE Trans. Power Syst., vol. 4, no. 2, pp. 614-626, 1989.

[5] H. F. Wang, Y. S. Hao, B. W. Hogg, and Y. H. Yang, "Stabilisation of multi-machine power systems by turbine governor control", IEE International Conference on Advances in Power System Control, Operation and Management, APSCOM-91, 5-8 Nov. 1991, vol.3, pp. 493-497.

[6] H. F. Wang, Y. S. Hao, B. W. Hogg, and Y. H. Yang, "An adaptive power system stabilizer based on turbine governor control", IEE International Conference on Advances in Power System Control, Operation and Management, APSCOM-93, 7-10 Dec. 1993, vol. 1, pp. 285-289.

[7] P. Kundur, Power System Stability and Control, EPRI Power System Engineering Series, Mc Graw-Hill, 1994.

[8] K. R. Padiyar, Power System Dynamics Stability and Control, Wiley, 1996.

[9] H. F. Wang, F. J. Swift, Y. S. Hao, and B. W. Hogg, "Adaptive stabilisation of power system by governor turbine control", Int. J. Electr. Power Energy Syst., vol. 18, no. 2, pp. 131-138, 1996.

[10] IEEE Std. 421.5, IEEE Recommended Practice for Excitation System Models for Power System Stability Studies, 2005.

[11] F. Zhang, and Z. Xu, "Effect of exciter and PSS on SSR damping", Proc. IEEE PES General Meeting, June 2007, pp. 1-7.

[12] J. Machowski, J. W. Bialek, and J. R. Bumby, Power System Dynamics Stability and Control, John Wiley & Sons, Ltd, 2008.

[13] G. Gurrala, and I. Sen, "Power system stabilizers design for interconnected power systems", IEEE Trans Power Syst., vol. 25, no. 2, 1042-1051, 2010.

[14] N. Hosseinzadeh and A. Kalam, "A direct adaptive fuzzy power system stabilizer", IEEE Trans. Energy Conversion, vol. 14, pp.1564 -1571, 1999.

[15] P. Hoang and K. Tomsovic, "Design and analysis of an adaptive fuzzy power system stabilizer", IEEE Trans. Energy Conversion, vol. 11, pp.455-481, 1996.

[16] N. Hosseinzadeh and A. Kalam, "A rule-based fuzzy power system stabilizer tuned by neural network", IEEE Trans. Energy Conversion, vol. 14, pp.773-779, 1999.

[17] A. Hariri and O. P. Malik, "A fuzzy logic based power system stabilizer with learning ability", IEEE Trans. Energy Conversion, vol. 11, pp.721-727, 1996.

[18] G. P. Chen and O. P. Malik, "An adaptive power system stabilizer based on the self-optimizing pole shifting control strategy", IEEE Trans. Energy Conversion, vol. 8, pp.639-645, 1993.

[19] P. He, F. Wen, G. Ledwich, Y. Xue, and K. W. Wang, " Effects of various power system stabilizers on improving power system dynamic performance", Int. J. Electr. Power Energy Syst., vol. 46, no. 2, pp. 175-183, 2013.

[20] Timothy. J. Ross, Fuzzy Logic: with Engineering Applications, Wiley, 2nd edition, 2007.

[21] M. Ramirez-Gonzalez and O. Malik, "Power system stabilizer design using an online adaptive neurofuzzy controller with adaptive input link weights", IEEE Trans. Energy Conversion, vol. 23, no. 3, pp. 914-922, 2008.

[22] F. Mayouf(Adjeroud), F. Djahli and A. Mayouf, "Study of excitation and governor power system stabilizers effect on the stability enhancement of a single machine infinte-bus power system", The 12th Int. Conf. on Environment and Elect. Eng. (EEEIC-2013), 5-8 May 2013, Wroclaw, Poland, pp. 534-538.

[23] F. Mayouf(Adjeroud), F. Djahli, A. Mayouf and T. Devers, "Study of excitation and governor power system stabilizers effect on the sability enhancement of a multimachine power system", The 9th Int. Symposium on Diagnostics for Electrical Machines, Power Electronics & Drives (SDEMPED 2013), 27-30 August 2013, Valencia, Spain, in press.

[24] F. Mayouf(Adjeroud), F. Djahli, A. Mayouf and T. Devers, "A new controller based on fuzzy logic and PSS for enhancing the stability of a single machine ifinte-bus power system", The 9th Int. Symposium on Diagnostics for Electrical Machines, Power Electronics & Drives (SDEMPED 2013), 27-30 August 2013, Valencia, Spain, in press.

[25] F. Mayouf(Adjeroud), F. Djahli, A. Mayouf and T. Devers, "Study of excitation and governor control effects of superconducting generator with high response excitation on the stability of a SMIB power system", The 5th Int. Conf. on Information Technology and Electrical Engineering (ICITEE 2013), 7-8 Oct. 2013, Yogyakarta, Indonesia, in press.

[26] T. Hiyama, "Robustness of fuzzy logic power system stabilizers applied to multimachine power system", IEEE Trans. EC, vol. 9, no. 3, pp.451-459, 1994.

[27] M. E. El-Hawary, Electric Power Applications of Fuzzy Systems, IEEE press, 1998.

[28] H. M. Behbehani, J. Bialek, and Z. Lubosny, "Enhancement of power system stability using fuzzy logic based supervisory power system stabilizer", 2nd IEEE Int. Conference on Power and Energy (PECon 08), December 1-3, 2008, Johor Baharu, Malaysia, pp. 479-484.

[29] M. L. Kothari and T. J. Kumar, "A new approach for designing fuzzy logic power system stabilizer", International Power Engineering Conference, IPEC 2007, 3-6 Dec. 2007, vol. 3, no. 6, pp. 419 – 424.

[30] Simulink User's Guide, The Mathworks, Natick, MA, 1999.

[31] Power System Blockset User's Guide, The Mathworks, Natick, MA, 1998.