

# SISTEM PENDETEKSI GAMBAR PORNO

Dadet Pramadihanto, Sritrusta Sukaridhoto, Muhammad Fatoni.A, Rudy Cahyadi H.P  
Jurursan Teknologi Informasi - Politeknik Elektronika Negeri Surabaya  
Institut Teknologi Sepuluh Nopember  
Kampus ITS Sukolilo, Surabaya 60111, INDONESIA  
email : [dadet@eepis-its.edu](mailto:dadet@eepis-its.edu), [dhoto@eepis-its.edu](mailto:dhoto@eepis-its.edu),  
[fatoni@student.eepis-its.edu](mailto:fatoni@student.eepis-its.edu), [rudy@student.eepis-its.edu](mailto:rudy@student.eepis-its.edu)

## ABSTRAK

Perkembangan teknologi dan internet saat ini, memungkinkan banyak orang untuk menikmati berbagai materi dengan bebas dan cuma-cuma yang akhirnya menjadi informasi yang bermanfaat. namun, ada beberapa materi yang seharusnya tidak dikonsumsi oleh semua kalangan dan bahkan tidak cocok dengan budaya kita seperti pornografi.

Pada paper ini kami menjelaskan bagaimana membangun sistem yang dapat mengidentifikasi serta memfilter apakah sebuah gambar tergolong porno atau tidak. Penyeleksian yang didasarkan pada warna kulit telah terbukti dan merupakan suatu syarat utama yang sangat bermanfaat. Skin detection mempunyai arti yang sangat penting dalam mendeteksi gambar porno. namun itu saja tidak cukup Maka dari itu penelitian kami menggabungkan analisa warna kulit dengan object detection yaitu proses mengidentifikasi bagian-bagian tubuh manusia supaya tingkat keakuratan yang dihasilkan bisa lebih baik. Dan proses yang terakhir adalah melakukan klasifikasi untuk memutuskan apakah gambar tersebut tergolong pornografi atau tidak. Semua itu dapat diimplementasikan secara open source dengan sistem operasi debian, pemrograman c/c++ dan library opencv.

Hasil penelitian ini mampu membantu menghadapi tantangan yang serius yaitu melindungi anak-anak kita sebagai bentuk kepedulian dan tanggung jawab dengan mengontrol, menyeleksi dan mengendalikan informasi yang masuk baik dari internet, televisi dan lain sebagainya dengan memanfaatkan open source.

Kata kunci: pornografi, deteksi warna kulit, object detection, OpenCV.

## 1. PENDAHULUAN

Perkembangan teknologi saat ini, memungkinkan lebih banyak orang untuk dapat menikmati internet. Ada banyak materi tersedia di internet dengan bebas dan cuma-cuma yang akhirnya menjadi informasi untuk kita. Walaupun demikian, ada beberapa materi yang seharusnya tidak dikonsumsi oleh semua kalangan seperti pornografi, padahal untuk mengaksesnya sangat mudah. Bahkan Jumlah gambar ini tidak sedikit dan hampir semuanya gratis.

Untuk memastikan hal itu maka diperlukan sebuah sistem penyeleksian yang dapat mengidentifikasi materi pornografi. Berbagai macam sistem yang dirancang berdasarkan *text content* untuk mencegah akses tersebut. proses penyeleksian didasarkan pada perbandingan alamat *IP / URLS* dan teks dengan daftar teks dan alamat yang tidak disetujui di database sistem tersebut. Namun pendekatan ini kurang akurat karena internet yang bersifat dinamis dan perkembangannya sangat cepat sehingga masih banyak materi pornografi yang lolos. Kemudian muncul ide yaitu penyeleksian berdasarkan gambar. Metode ini dapat meningkatkan keakuratan sehingga dapat handle bermacam-macam variasi materi pornografi yang beredar di internet.

Penyeleksian berdasarkan gambar terdiri dari *skin detection* yaitu proses segmentasi atau pengelompokan area warna kulit, *object detection* yaitu proses

mengidentifikasi bagian-bagian tubuh manusia dan yang terakhir melakukan proses klasifikasi untuk memutuskan apakah gambar tersebut tergolong pornografi atau tidak.

## 2. LANDASAN TEORI

### 2.1 Skin Detection

Untuk mengklasifikasi skin color dan non-skin color pada image digunakan 2 cara yaitu physic model atau training model. Tujuan keduanya sama-sama menyeleksi warna kulit.

#### 2.1.1 Physic Model

Physic model yang akan kami gunakan adalah modifikasi dari model Peer[9]. Model tersebut secara langsung dipakai untuk menyeleksi warna kulit.

$$r > 95 \text{ and } g > 40 \text{ and } b > 20 \text{ and } \max(r, g, b) - \min(r, g, b) > 15 \text{ and } \text{abs}(r - g) > 15 \text{ and } r > g \text{ and } r > b \quad (1)$$

#### 2.1.2 Training Model

Kemudian untuk model training Pertama kali yang dilakukan adalah memilih image warna kulit sebagai pembelajaran system, kemudian mengekstraksi nilai pixel dengan menggunakan *color space* RGB/chromatic. Setelah semua nilai pixel kulit terkumpul kemudian, baru kita cari nilai centre of space (cof) dan jarak rata-rata cof dengan masing-masing data.

Untuk proses klasifikasinya maka diperlukan jarak rata-rata tersebut. Jarak inilah yang dijadikan acuan ukuran apakah pixel input tergolong salah satu cluster atau tidak sama sekali.

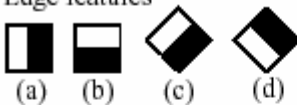
## 2.2 Object Detection

Untuk mengidentifikasi obyek dalam hal ini bagian tubuh manusia, menggunakan metode Cascade of Boosted Classifier using Haar-like Features.

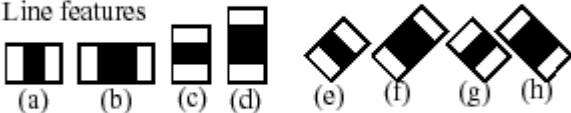
### 2.2.1 Haar-like Features

Setiap Haar-like feature terdiri dari gabungan kotak-kotak hitam dan putih:

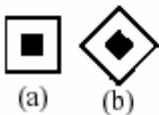
#### 1. Edge features



#### 2. Line features



#### 3. Center-surround features



Gambar 1. Satu set extended Haar-like features

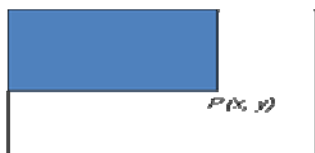
Nilai dari Haar-like feature adalah perbedaan antara jumlah nilai-nilai piksel gray level dalam daerah kotak hitam dan daerah kotak putih:

$$f(x) = \text{Sum}_{\text{black rectangle}} (\text{pixel gray level}) - \text{Sum}_{\text{white rectangle}} (\text{pixel gray level}) \quad (2)$$

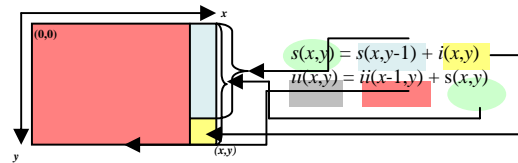
### 2.2.2 Integral Image

Kotak (rectangular) Haar-like feature dapat dihitung secara cepat menggunakan "integral image".

Integral image pada lokasi x,y mengandung jumlah nilai piksel diatas dan dikiri dari lokasi x,y :

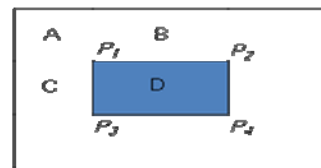


$$P(x, y) \equiv \sum_{x' \leq x, y' \leq y} i(x', y')$$



Gambar 2(a) perhitungan integral image

Pada gambar diatas,  $i(x,y)$  adalah nilai piksel dari image pada posisi x,y.  $s(x,y)$  adalah kumulatif jumlah kolom, Kita dapat menghitung integral image dengan sekali jalan (single pass).



Gambar 2(b) perhitungan integral image

Dengan menggunakan integral image, nilai jumlah piksel rectangular dapat dihitung dalam waktu yang konstan. Sebagai contoh jumlah nilai piksel didalam kotak D, dapat dihitung sbb:

$$ii(P4) + ii(P1) - ii(P2) - ii(P3) \quad (3)$$

### 2.2.3 Stage Classifier

Stage classifier dibangun dengan menggunakan algoritma adaptive-boost (AdaBoost). Algoritma tersebut mengkombinasikan performance banyak weak classifier untuk menghasilkan strong classifier. Weak classifier dalam hal ini adalah nilai dari haar-like feature. Jenis AdaBoost yang digunakan adalah Gentle AdaBoost.

### 2.2.4 Cascade Classifier

Cascade classifier adalah sebuah rantai stage classifier, dimana setiap stage classifier digunakan untuk mendeteksi apakah didalam image sub window terdapat obyek yang diinginkan ( object of interest).

## 3. METODOLOGI PENELITIAN

### 3.1 Skin Detection

Skin detection Classifier yang digunakan adalah perpaduan antara mapping training model dengan classifier buatan PEER tahun 2003.

$$\text{red} > 95 \ \&\& \ \text{green} > 40 \ \&\& \ \text{blue} > 20 \ \&\& \ \text{maxrgb} - \text{minrgb} > 15 \ \&\& \ \text{abs}(\text{red} - \text{green}) > 15 \ \&\& \ \text{red} > \text{green} \ \&\& \ \text{red} > \text{blue} \ \&\& \ \text{green} > \text{blue} \quad (4)$$

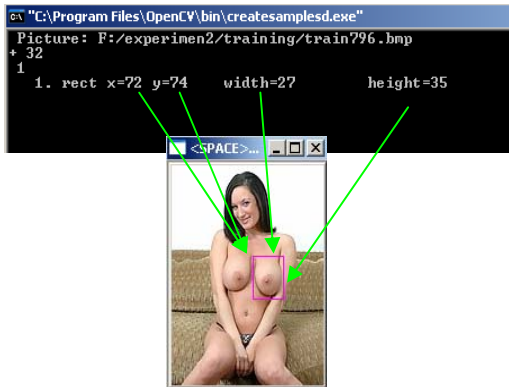
### 3.2 Training Cascade Classifier

Step 1 : Persiapan

Objectmarker.cpp, createsamples.cpp, haartraining.cpp.

Datanegative: adalah sekumpulan gambar yang tidak mengandung obyek yang akan dideteksi.

Data positif: adalah sekumpulan gambar yang mengandung obyek yang akan dideteksi. Untuk training nama-nama file data positif beserta bounding rectangle dari obyek yang telah ditandai (menggunakan objectmarker.cpp ).



Gambar 3. Output program objectmarker

Step 2: Menciptakan sample data set positif

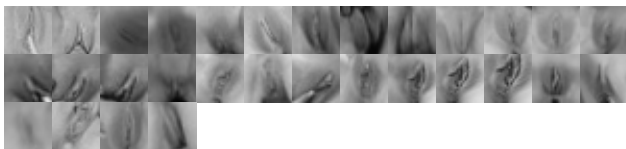
Menciptakan sample data set positif untuk training menggunakan createsample.cpp.

Pada step ini obyek yang telah ditandai dalam bounding rectangle disimpan kedalam file .vec

Berikut ini adalah contoh sample data positif untuk breast ( untuk merubah data vector kedalam bentuk gambar menggunakan vec2img.cpp ) :



Gambar 4. Sample data positif breast



Gambar 5. Sample data positif pussy

Step 3: Training

Pada step training ini akan dihasilkan sebuah cascade classifier yang berisi sejumlah stage yang didalamnya terdapat kumpulan data classifier. Cascade classifier dihasilkan dengan menggunakan haartraining.cpp.

Berikut ini adalah sample data XML classifier untuk breast detection :

```

=<opencv_storage>
=<cascadebreast1981 type_id="opencv-haar-
=<classifier">
=<size>24 24</size>
=<stages>
=<_>
=<trees>
=<_>
=<_>
=<_>
=<feature>
=<rects>
=<_>5 16 13 6 -1.</_>
=<_>5 19 13 3 2.</_>
=</rects>
=<tilted>0</tilted>
=</feature>
=<threshold>-0.0202897991985083</threshold>
=<left_val>0.5025501251220703</left_val>
=<right_val>-0.4522908926010132</right_val>
=</_>
=</_>
...
</trees>
=<stage_threshold>-
1.3585350513458252</stage_threshold>
=<parent>-1</parent>
=<next>-1</next>

```

### 3.3 Object Detection

Didefinisikan obyek-obyek porno (bagian tubuh manusia) adalah : Payudara, vagina, hand touch pussy dll.

Diberikan sebuah gambar yang berasal dari sebuah file atau live video, objek porno detector menguji setiap lokasi gambar dan mengklasifikasikannya sebagai "objek porno" atau "bukan objek porno". Klasifikasi mengasumsikan sebuah ukuran skala tetap untuk objek porno, katakan ukuran 24\*24 pixel, walaupun objek porno dalam sebuah input gambar lebih kecil atau lebih besar daripadanya, Classifier menguji gambar beberapa kali, untuk mencari objek porno dalam sebuah skala range.

Classifier menggunakan data yang disimpan dalam sebuah file XML untuk memutuskan bagaimana mengklasifikasikan setiap lokasi gambar. Untuk setiap objek porno masing-masing mempunyai satu file data XML.

#### 4. HASIL PENELITIAN

Tabel 1. detail spesifikasi hardware dan software

<b>NEC VERSA L2100 - 1800D RC</b>	
Processor	Intel® Core™ Duo Processor T2350 (1.86GHz, 2MB L2 Cache, 533 Mhz FSB)
Chipset	Intel® 945GM chipset
Memory	512MB (1x512MB) DDR2 533MHz
Hard Disk Drive	80GB PATA 5400RPM
Video Card	Intel® Graphics Media Accelerator (GMA) 950™
Notebook Display	14.1" WXGA Super Shine View LCD
Operating System	Debian 4 kernel 2.6.18-3-686
Library OpenC	OpenCV 0.97

##### 4.1 Deteksi Obyek Dan Deteksi Warna Kulit

Berikut ini adalah hasil deteksi obyek yang ( object of interest ) diinginkan ditambah dengan deteksi warna kulit.



Gambar 6. Breast Detection + skin detection yang tepat



Gambar 7. Breast detection yang kurang tepat + skin detection



Gambar 8. Face + breast + hand touch pussy detection yang tepat



Gambar 9. breast + hand touch pussy detection tepat



Gambar 10. Pussy detection yang salah



Gambar 11. Breast + hand touch pussy + pussy detection yang tepat serta pussy detection yang salah



Gambar 12. Breast detection yang tepat walaupun ditutupi pakaian transparan



Gambar 13. Gambar yang tidak terdeteksi adanya object of interest sama sekali



Gambar 14. Face detection + breast detection + hand touch pussy + pussy detection

#### 4.2 Perhitungan Performance

Berdasarkan keputusan sederhana ya atau tidak, apakah suatu gambar terdeteksi adanya object of interest atau tidak, maka terdapat 4 penilaian sebagai berikut :

Tabel 2. Performance keputusan sederhana

	gambar terdeteksi adanya object of interest	gambar tidak terdeteksi adanya object of interest
Gambar terdapat object of interest	True Positive ( TP )	False Negative( FN )
Gambar tidak terdapat object of interest	False Positive ( FP )	True Negative ( TN )

Untuk mengukur tingkat keefektifan suatu sistem klasifikasi, berikut ini terdapat empat definisi/ukuran yang digunakan :

**Presi :** prosentase tingkat ketepatan klasifikasi item didalam set jumlah gambar yang terdeteksi adanya object of interest dengan rumus  $TP/(TP+FP)$ . Ukuran ini mengindikasikan berapa banyak gambar yang benar-benar terdapat object of interest dari seluruh gambar seluruh gambar yang ditandai terdapat object of interest.

**Recall/sensitifitas :** prosentase dari jumlah item gambar yang terdeteksi adanya object of interest didalam seluruh set item gambar terdapat object of interest dengan rumus  $TP/(TP+FN)$ . Kuran ini mengindikasikan seberapa bagus filter dalam mengidentifikasi object of interest.

**Spesifisitas/kekhususan :** prosentase dari jumlah item gambar tidak terdeteksi adanya object of interest didalam seluruh set item gambar yang tidak terdapat object of interest dengan rumus  $TN/(TN+FP)$ . Ukuran ini mengindikasikan seberapa bagus filter didalam mengidentifikasi object non of interest.

**Akurasi :** prosentase dari ketepatan hasil klasifikasi didalam seluruh jumlah item-item dengan ukuran  $(TP+TN)/(TP+TN+FP+FN)$ .

#### 4.3 Performance Breast Detection + Skin Detection

Berikut ini adalah hasil performance dari gambar terdapat breast (184 gambar, jumlah stage 26, ukuran 270x360 piksel), gambar-no human tidak terdapat breast ( 120 gambar, ukuran 320x240 piksel ), dan gambar-human tidak terdapat breast ( 100 gambar, ukuran 320x240 piksel ):

Tabel 2. Performance keputusan sederhana

	gambar terdeteksi adanya breast	gambar tidak terdeteksi adanya breast
Gambar terdapat breast (+)	160	24
Gambar-no human tidak terdapat breast (-)	1	119
Gambar-human tidak terdapat breast (-)	2	118

#### Tingkat keefektifan :

##### Presi :

$$\text{gambar(+)} \ \& \ \text{gambar(-)-no human} : (160/(160+1)) * 100\% = 99\%$$

$$\text{gambar(+)} \ \& \ \text{gambar(-)-human} : (160/(160+2)) * 100\% = 99\%$$

$$\text{gambar(+)} \ \& \ \text{gambar(-)-no human} \ \& \ \text{gambar(-)-human} : (160/(160+1+2)) * 100\% = 98\%$$

$$\text{Recall/sensitifitas: } (160/(160+24)) * 100\% = 87\%$$

##### Spesifisitas:

$$\text{gambar(+)} \ \& \ \text{gambar(-)-no human} : (119/(119+1)) * 100\% = 99\%$$

$$\text{gambar(+)} \ \& \ \text{gambar(-)-human} : (118/(118+2)) * 100\% = 98\%$$

$$\text{gambar(+)} \ \& \ \text{gambar(-)-no human} \ \& \ \text{gambar(-)-human} : (119+118/(119+118+1+2)) * 100\% = 99\%$$

##### Akurasi:

$$\text{gambar(+)} \ \& \ \text{gambar(-)-no human} : ((160+119)/(160+119+24+1)) * 100\% = 92\%$$

$$\text{gambar(+)} \ \& \ \text{gambar(-)-human} : ((160+118)/(160+118+24+2)) * 100\% = 91\%$$

$$\text{gambar(+)} \ \& \ \text{gambar(-)-no human} \ \& \ \text{gambar(-)-human} : ((160+119+118)/(160+119+118+24+1+2)) * 100\% = 94\%$$

## 5. KESIMPULAN

1. Metode modifikasi peer untuk mendeteksi warna kulit adalah metode yang paling baik

2. Metode untuk mendeteksi tekstur menggunakan co-occurrence matrik membutuhkan waktu proses yang cukup lama
3. Hasil pendeteksian / identifikasi gambar porno paling bagus dari sisi akurasi dan waktu proses adalah penggabungan skin detection dengan object detection  
Didalam metode skin detection hasil yang paling baik adalah menggunakan stage 26.
4. Semakin kecil jumlah stage yang digunakan maka semakin besar pula tingkat kesalahan deteksi. Namun jika menggunakan stage terlalu besar maka system akan sangat berhati-hati sekali dalam mendeteksi object yang diinginkan.

## 6. DAFTAR PUSTAKA

1. Robin Hewitt. seeing with OpenCV Part 2 – Finding Face in Image, Jan 2007.
- 2 Qing Chen. Hand Detection with a Cascade of Boosted Classifiers Using Haar-like Features. Discover Lab, SITE, University of Ottawa. May 2006.
3. Rainer Lienhart and Jochen Maydt. An Extended Set of Haar-like Features for Rapid Object Detection. IEEE ICIP 2002, Vol. 1, pp. 900-903, Sep. 2002.
4. Paul Viola and Michael J. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. IEEE CVPR, 2001.
5. Rainer Lienhart, Alexander Kuranov, Vadim Pisarevsky. Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection.
6. JureKovac, PeterPeer, and FrancSolina. HumanSkinColourClusteringforFaceDetection . UniversityofLjubljana Faculty of Computer and Information Science. Slovenia
7. Gady Agam. Introduction to programming with OpenCV. 2006
8. Open Computer Vision Library. <http://sourceforge.net/projects/opencvlibrary/>
9. PINKBLOCK. Pornography Image Filter Effectiveness. Jan. 2007
10. Florian Adolf . OpenCV's Rapid Object Detection How-to build a cascade of boosted classifiers based on Haar-like features. Sep. 2003
11. Gary Bradski. Object Detection and Classification Using Machine Learning. Intel. Stanford