

Prototype Pemblokir Gambar Pornografi Menggunakan Algoritma Viola and Jones

Muhammad Fikri Hidayattullah^{1*}, Miftakhul Huda²

¹Program Studi Teknik Informatika, Politeknik Harapan Bersama, Tegal

²Program Studi Teknik Komputer, Politeknik Harapan Bersama, Tegal

^{1,2}Jl. Mataram No. 9, Pesurungan Lor, Margadana, Kota Tegal, Jawa Tengah 52147

Email : ¹fikri@poltektegal.ac.id, ²miftah.puskom.it@poltektegal.ac.id

Copyright ©2019, Politeknik Harapan Bersama, Tegal

Abstract : The spread of pornographic content extends very quickly through internet media. Various types of porn sites increase dramatically every day. Various prevention efforts to overcome the spread of pornographic content have been carried out. Starting from URL blocking, text filtering to be more specific directly to pornographic images through skin detection. However, all of these methods have various weaknesses. An effective step to reduce the spread of pornographic images is to block the breast area that is characteristic of the pornographic images. In this study a breast area blocking prototype was developed on pornographic images using the Viola and Jones algorithm.. From the test results obtained a precision value of 82.61%, 77.55% recall and accuracy of 78.26%.

Abstrak : Persebaran konten pornografi meluas dengan sangat cepat melalui media internet. Berbagai situs porno jumlahnya meningkat drastis setiap harinya. Berbagai upaya pencegahan untuk menanggulangi tersebarnya konten pornografi telah dilakukan. Mulai dari pemblokiran URL, pemfilteran teks hingga lebih spesifik mengarah langsung ke gambar porno melalui deteksi kulit. Namun, semua metode ini memiliki berbagai kelemahan. Langkah yang efektif untuk mengurangi persebaran gambar porno adalah dengan memblok area payudara yang menjadi ciri dari gambar porno tersebut. Dalam penelitian ini dikembangkan sebuah *prototype* pemblokir area payudara pada gambar porno menggunakan algoritma Viola and Jones. Dari hasil pengujian didapatkan nilai *precision* 82,61%, *recall* 77,55% dan akurasi 78,26%.

Kata Kunci : *blokir, pornografi, Viola and Jones, internet*

I. PENDAHULUAN

Pornografi ibarat candu, siapa yang mencoba berinteraksi dengannya akan terkena efek adiksi. Berbagai dampak negatif dimunculkan akibat dari pengaruh buruk pornografi, seperti malas melakukan aktivitas yang positif, mengalami perilaku seks yang menyimpang, terkena efek kecanduan parah, melakukan pelecehan seksual hingga terjatuh ke dalam tindak pidana perkosaan.

*) **Corresponding author:** (Muhammad Fikri Hidayattullah)
Email: fikri@poltektegal.ac.id

Pornografi tidak hanya merusak akal dan pikiran seseorang, namun juga memberi dampak buruk terhadap perkembangan tubuh dan kestabilan emosi para pecandunya [1]. Seorang pecandu pornografi seakan di dalam pikirannya tercipta sebuah *file* dengan nama "*Porn Library*". Semakin banyak dirinya mengakses dan menyaksikan tayangan pornografi, maka semakin bertambah pula koleksi *database* pornografi di dalam "*Porn Library*" yang telah tertanam di pikirannya. Sehingga setiap waktu yang ada di benak pikirannya hanya bayangan vulgar hal-hal berbau pornografi.

Persebaran pornografi kian hari semakin masif. Sarana terbesar tersebarnya konten pornografi adalah media internet. Pada tahun 2005 saja jumlah pengguna internet yang pernah mengakses laman pornografi sebanyak 42,7% dan setiap detik sebanyak 28.258 pengguna internet mengakses laman pornografi [2]. Jumlah tersebut merupakan angka yang sangat besar mengingat pada tahun-tahun tersebut akses penggunaan internet tidak secepat dan semudah saat ini. Setiap dua tahun kapasitas internet bertambah sebanyak dua kali lipat dan akan menanjak hingga sepuluh kali lipat antara tahun 2015 sampai 2020. Pada tahun 2020 diperkirakan jumlah laman pornografi akan mencapai 5,72 trilyun laman atau sebesar 2.016.000 Peta Bytes yang diakses secara *online* [3]. Hal ini tentu saja menjadi ancaman yang serius bagi perkembangan mental dan psikologis seseorang.

Berbagai metode pencegahan persebaran konten pornografi melalui media internet telah dikembangkan oleh para peneliti. Mulai dari pemblokiran alamat *website* (URL) yang berisi konten pornografi [4][5][6] dan juga pemblokiran laman *website* yang terindikasi terdapat kata-kata porno di dalamnya (pemfilteran teks) [7][8]. Namun metode ini memiliki kelemahan yang bersifat fundamental, yaitu terkait *update*-an setiap waktu alamat URL yang dinilai berisi konten pornografi. Padahal jumlah domain URL mencapai jutaan dengan tingkat pertumbuhan yang sangat cepat. Ditambah lagi metode pemblokiran URL ini sangat mudah untuk dikelabui dengan membuat URL *website* porno menggunakan domain yang sama sekali tidak berbau porno, misal: www.pendidikan-anak.com. Sedangkan untuk metode pemfilteran teks pada laman sebuah *website* juga tidak efektif untuk dilakukan. Hal ini disebabkan seringkali sistem

yang dibuat berbasis metode ini melakukan kesalahan pemblokiran pada sebuah website non-porno dikarenakan laman *website* tersebut mengandung kata-kata yang dianggap bermuatan unsur porno. Kata-kata tersebut dianggap sesuai dengan “kamus porno” yang ada di dalam sistem. Misalkan, ada sebuah laman website yang memuat informasi seputar kesehatan reproduksi wanita. Di dalam laman tersebut terdapat beberapa kata yang dianggap bermuatan unsur porno oleh sistem, seperti “payudara” dan “vagina”. Maka, secara otomatis web non-porno tersebut dikategorikan web porno dan langsung diblokir.

Metode pemblokiran URL dan pemfilteran teks dirasa kurang efektif untuk diterapkan. Untuk itu dikembangkan metode selanjutnya, yaitu metode pemblokiran berbasis pengenalan gambar pornografi. Metode yang sering digunakan untuk mendeteksi gambar porno dengan cara melakukan deteksi area kulit pada suatu gambar (*skin detection*). Luasan persentase area kulit akan dibandingkan dengan luasan persentase area non-kulit. Jika luasan area kulit melebihi nilai *threshold* (ambang batas) yang telah ditentukan oleh sistem maka akan dianggap sebagai gambar telanjang. Lin, Tsen & Fuh [9] menggunakan metode *Support Vector Machine* (SVM) untuk melakukan klasifikasi area kulit dan non-kulit. Agbinya *et al.* [10] memanfaatkan ruang warna HSV untuk melakukan segmentasi area kulit pada gambar porno. Metode pengenalan gambar porno menggunakan deteksi kulit ini pun memiliki kelemahan fundamental. Kelemahan tersebut adalah jika suatu gambar porno dengan objek visual pornografi berukuran kecil dibanding dengan area *background*-nya, maka akan dianggap bukan gambar porno. Hal ini disebabkan karena luasan persentase kulit pada gambar tersebut lebih kecil dibanding nilai *threshold*. Begitu juga sebuah gambar non-porno akan dianggap sebagai gambar porno jika memiliki luasan area kulit yang dominan. Bahkan, objek manusia berpakaian tertutup pun akan dianggap sebagai gambar porno jika pakaian yang digunakan menyerupai warna kulit manusia.

Salah satu solusi utama untuk mendeteksi secara efektif suatu gambar termasuk kategori porno atau non-porno adalah dengan langsung mendeteksi organ intimnya. Sebuah gambar dikategorikan pornografi jika menampilkan organ intim. Salah satu organ intim yang dapat dideteksi adalah payudara. Para peneliti yang telah menggunakan langkah ini diantaranya Wang, Hu & Yao [11] dengan mengembangkan algoritma pendeteksi payudara pada gambar porno, dan juga Xen, Wei & Qian [12] dengan melakukan penggabungan beberapa organ erotik untuk dideteksi.

Metode deteksi organ intim untuk mengenali gambar porno terbilang cukup efektif. Namun, untuk meningkatkan efektifitas nilai kemanfaatannya perlu dilakukan lagi langkah lanjutan sampai tahap pemblokiran. Tahap pemblokiran organ intim berupa payudara mampu menghilangkan atau menutupi area utama yang menjadi ciri gambar pornografi. Oleh karena itu penelitian ini berfokus untuk mengembangkan sebuah *prototype* pemblokir area payudara wanita pada gambar pornografi dengan menggunakan algoritma Viola and Jones.

II. PENELITIAN TERKAIT

Wang, Hu & Yao [11] mengembangkan sebuah algoritma untuk mendeteksi payudara wanita pada gambar porno. Algoritma ini memanfaatkan deteksi kulit menggunakan ruang warna YcYbCr. Tahap deteksi kulit diperlukan untuk segmentasi area. Setelah itu dilanjutkan dengan

pengoperasian *Nipple Feature* (NF) sebagai dasar dari deteksi payudara. Pada tahap penentuan kandidat area NF digunakan algoritma klastering *Fuzzy C-Means*. Tahap selanjutnya adalah ekstraksi fitur dari kandidat yang diperoleh pada tahap sebelumnya. Kemudian dilanjutkan tahap terakhir berupa pelatihan fitur dan klasifikasi menggunakan algoritma *Forward Propagation*. Penelitian ini menghasilkan presisi sebesar 93,18% dan *recall* sebesar 90,01%.

Xen, Wei & Qian [12] melakukan penelitian seputar deteksi gambar pornografi. Penelitian ini menggabungkan beberapa organ tubuh untuk dideteksi hingga didapatkan gambar porno yang valid. Tahap pertama yang dilakukan adalah deteksi wajah. Setelah didapatkan area wajah dilanjutkan deteksi area tubuh. Kemudian dilanjut tahap ketiga berupa deteksi kulit menggunakan ruang warna YcCbCr. Setelah tahap deteksi kulit, langkah berikutnya adalah deteksi organ intim payudara dan rambut vagina. Tahapan metode yang ditempuh dalam penelitian ini cukup selektif untuk meningkatkan akurasi deteksi. Dari hasil pengujian diperoleh nilai *true positive* sebesar 80,31% dan *false positive* sebesar 19,69%.

Wang *et al.* [13] dalam penelitiannya melakukan pengenalan gambar porno melalui deteksi puting payudara. Langkah pertama yang dilakukan adalah melakukan pencarian area puting payudara menggunakan algoritma *AdaBoost*. Setelah itu dilanjutkan dengan melakukan deteksi puting payudara menggunakan informasi bentuk dan hubungan warna kulit antara area puting dan non-puting. Akurasi pendeteksian yang dihasilkan sebesar 75,6%.

Kejun & Jian [14] mengembangkan fitur *Haar-Like* di tahap ekstraksi fitur dalam algoritma Viola and Jones. Fitur *Haar-Like* yang dikembangkan tersebut merupakan turunan dari fitur *Haar-Like (Extended Haar-Like Feature)* yang telah dikembangkan sebelumnya. Algoritma Viola and Jones yang telah di-*improve* ini digunakan untuk mendeteksi puting payudara wanita pada gambar pornografi. Setelah dikomparasi dengan algoritma Viola and Jones standar terbukti penggunaan *Extended Haar-Like Feature* mampu meningkatkan akurasi deteksi dari 75,64% menjadi 90,37%.

Thaweekote, Songram & Jareanpon [15] juga melakukan pendeteksian gambar pornografi melalui deteksi puting payudara. Namun, dalam penelitian ini untuk menghindari kesalahan deteksi area puting terlebih dahulu dilakukan deteksi wajah. Setelah berhasil mendeteksi area wajah akan dilanjut melakukan deteksi area tubuh. Kemudian dilanjutkan ke proses deteksi kulit. Langkah terakhir setelah area kulit tersegmentasi dilanjutkan tahap deteksi area puting payudara. Dari hasil pengujian diperoleh nilai *true positive* sebesar 95,25% dan *false positive* sebesar 4,75%.

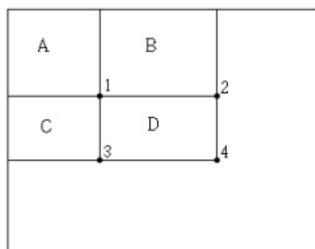
III. ALGORITMA VIOLA JONES, ADABOOST & CASCADE CLASSIFIERS

A. Algoritma Viola and Jones

Algoritma *Viola and Jones* merupakan algoritma yang dikembangkan oleh dua orang peneliti yang bernama Paul Viola dan Michael Jones. Algoritma ini diperkenalkan melalui paper mereka yang berjudul *Rapid Object Detection using Boosted Cascade of Simple Features* pada tahun 2001 [16]. Terdapat tiga kontribusi utama dalam algoritma ini. Pertama, penggunaan *integral image* untuk mempercepat perhitungan area piksel gambar. Proses perhitungan piksel-piksel ini menggunakan sebuah fitur yang disebut fitur *Haar-Like*. Kedua, penggunaan algoritma *AdaBoost* untuk

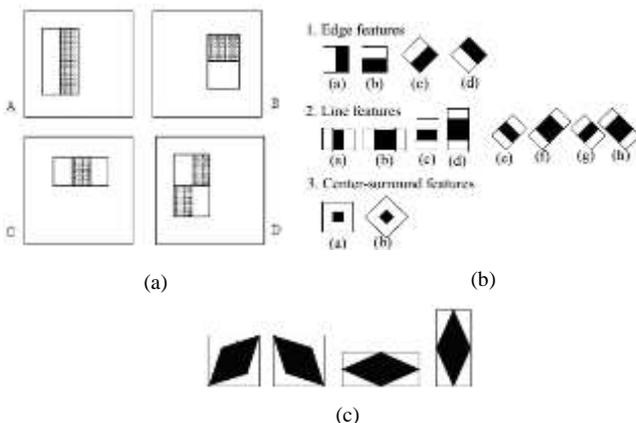
melakukan *feature selection*. AdaBoost mampu meminimalisir jumlah fitur *Haar-Like* yang terlalu besar, bahkan bisa melebihi jumlah piksel. Kontribusi ketiga adalah *cascade classifiers*. *Cascade classifier* merupakan metode untuk menggabungkan beberapa *classifier* yang lebih kompleks secara berturut-turut dalam struktur bertingkat (*cascade*). Melalui metode ini kecepatan deteksi dapat meningkat dengan sangat cepat.

Integral image (citra integral) mampu menghitung dengan cepat total area piksel pada sebuah fitur *Haar-Like* tanpa melakukan penjumlahan keseluruhan piksel. Namun, perhitungan total keseluruhan piksel tersebut cukup dengan empat referensi *array*.



Gbr 1. Perhitungan Penjumlahan dengan *Integral Image*

Fitur yang digunakan dalam perhitungan *integral image* dinamakan fitur *Haar-Like* disebabkan karena menyerupai fungsi *Haar Basis* yang diperkenalkan oleh Papageorgiou et al. [17]. Pada tahun 2002 fitur *Haar-Like* dibuat turunannya oleh Lienhart sebanyak tiga jenis, yaitu *edges features*, *line features* dan *center-surround features* [18]. Selanjutnya pada tahun 2012, Kejun & Jian [14] mengembangkan turunan dari fitur *Haar-Like* yang lebih spesifik untuk deteksi puting payudara wanita pada gambar pornografi.



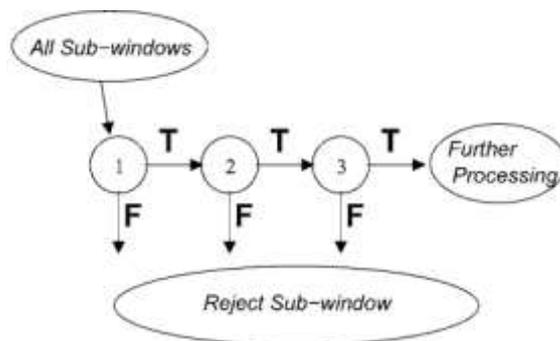
Gambar 2. a. Fitur *Haar-Like* Generasi Pertama yang Digunakan Viola and Jones, b. Fitur *Haar-Like* Turunan oleh Lienhart, c. Fitur *Haar-Like* Turunan oleh Kejun & Jian.

B. AdaBoost

Algoritma *AdaBoost* digunakan untuk menyeleksi fitur dan melatih *classifiers*. Algoritma ini merupakan salah satu algoritma *Boosting* yang awalnya digunakan untuk mempercepat performa klasifikasi pada algoritma pembelajaran sederhana. Inti utama dari penggunaan *AdaBoost* di dalam algoritma *Viola and Jones* untuk mempercepat dan mengefektifkan pendeteksian objek.

C. Cascade Classifiers

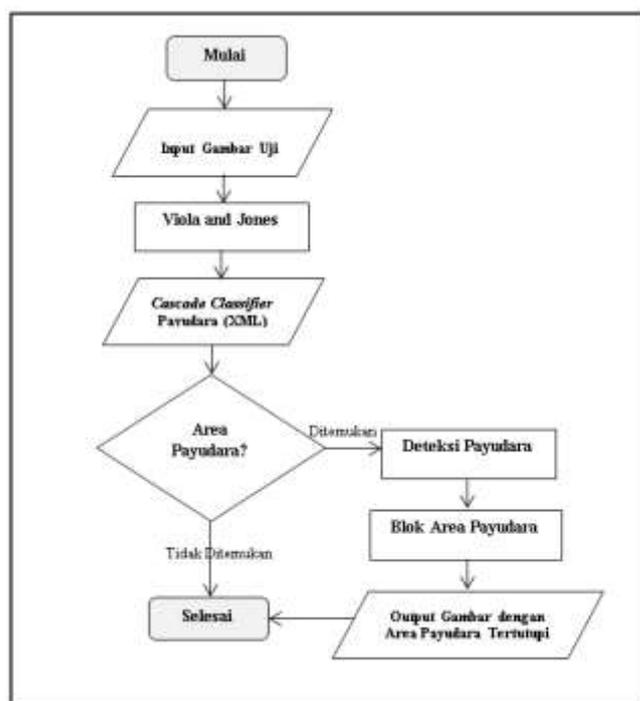
Tahapan ini sangat membantu detektor untuk meningkatkan kecepatan dan ketepatan deteksinya. *Cascade classifiers* dapat diibaratkan seperti seleksi bertingkat untuk mendapatkan area objek yang akan dideteksi. Seleksi di tingkat (*stage*) pertama sangat longgar sekali. Kemudian jika berlanjut ke tingkat kedua sudah tidak selonggar tingkat pertama. Begitu seterusnya hingga diperoleh *stage* yang pas untuk kasus yang akan dijadikan objek deteksi. Semakin tinggi *stage* tidak menjamin akan semakin tepat pendeteksiaannya, karena semakin ketat proses penyaringan *sub-window* yang dilakukan. Bahkan seringkali yang terjadi adalah objek gagal terdeteksi. Oleh karena itu, ketepatan penentuan batas *stage* sewaktu proses training sangat berpengaruh sekali bagi *classifiers* untuk membantu proses deteksi.



Gbr 3. Skema *Cascade Classifiers* dalam Mendeteksi

IV. METODE PENELITIAN

Tahap pertama sebelum melakukan pemblokiran adalah pendeteksian area payudara. Algoritma yang digunakan adalah Viola and Jones. Algoritma ini dikenal kuat dan cepat dalam mendeteksi objek [16]. Setelah diperoleh area payudara pada gambar pornografi, maka secara otomatis akan dilakukan pemblokiran (penutupan) area tersebut.



Gbr 4. Skema Model yang Diusulkan

Gambar uji (*data testing*) merupakan *dataset* yang berisi sekumpulan gambar uji positif dan negatif yang akan diujikan pada metode yang diusulkan. Gambar uji positif merupakan gambar yang berisi objek area payudara wanita yang akan diblok. Sedangkan gambar uji negatif merupakan gambar yang tidak mengandung objek area payudara wanita. Jumlah gambar uji positif sebanyak 50 gambar, sedangkan gambar uji negatif sebanyak 20 gambar.

Dalam penelitian ini *cascade classifiers* berbentuk *file XML (Extensible Markup Language)*. *File* ini merupakan hasil implementasi dari serangkaian proses pelatihan data positif dan data negatif, perhitungan fitur Haar-Like dengan *integral image*, seleksi fitur dan pelatihan *classifiers*. Dari seluruh tahapan ini dihasilkan sebuah file detektor objek payudara wanita berbentuk *file XML*. *Template XML* ini tersusun dari 34 *stages*.

V. HASIL DAN PEMBAHASAN

Setiap gambar porno yang diuji di dalam penelitian ini berisi hanya satu pose wanita telanjang. Total gambar uji positif yang diujikan berjumlah 50 dan gambar uji negatif berjumlah 20. Gambar porno yang telah selesai diuji secara otomatis akan tertutupi area payudaranya dengan blok persegi warna merah.



Gbr 5. Hasil Pengujian Gambar Porno

Beberapa area payudara yang tidak berhasil terblokir oleh sistem disebabkan karena beberapa faktor, diantaranya adalah area payudara pada gambar porno tidak tampil secara utuh atau tertutupi sebagian areanya, pose payudara yang menghadap ke samping tidak ke depan dan tingkat pencahayaan yang terlalu tinggi. Sedangkan kesalahan deteksi (*false positive*) terjadi pada beberapa area lekukan pada tubuh yang mirip dengan bentuk tepi payudara. Dari hasil pengujian terhadap seluruh gambar uji diperoleh nilai seperti pada Tabel I.

TABEL I
 HASIL PENGUJIAN

Gambar Uji	True Positive	False Positive	True Negative	False Negative
Positif	38	4	-	11
Negatif	-	4	16	-
Total	38	8	16	11

$$Precision = \frac{TP}{FP + TP} * 100\% = \frac{38}{8+38} * 100\% = 82,61\%$$

$$Recall = \frac{TP}{FN + TP} * 100\% = \frac{38}{11+38} * 100\% = 77,55\%$$

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

$$Akurasi = \frac{38+16}{38+16+4+11} * 100\% = 78,26\%$$

Nilai *precision* yang diperoleh sebesar 82,61%, *recall* sebesar 77,55% dan akurasi sebesar 78,26%.

VI. KESIMPULAN

Pemblokiran area payudara pada gambar porno menggunakan algoritma *Viola and Jones* terbukti cukup efektif. Beberapa area payudara yang gagal terblokir disebabkan karena area payudara tidak tampil secara utuh, pose payudara yang menghadap ke samping dan tingkat pencahayaan yang terlalu tinggi. Oleh karena itu untuk meningkatkan akurasi deteksi dan pengeblokan perlu dikembangkan lagi *cascade detector* hasil pemrosesan dari data latih yang lebih bervariasi lagi. Penggabungan dengan metode lain seperti deteksi kulit dan juga pengembangan dari algoritma *Viola and Jones* akan berdampak lebih baik untuk meningkatkan akurasi.

DAFTAR PUSTAKA

- [1] P. F. Fagan and D. Ph, "The Effects of Pornography on Individuals, Marriage, Family, and Community," *Marriage & Religion Institute*, pp. 1-26, 2009.
- [2] B. J. Ropelato, "Internet Pornography Statistics," pp. 1-10, 2014.
- [3] EDsmart, "Internet Pornography Stats," *Pornography by the Numbers*, 2015..
- [4] S. N. Hamade, "Internet Filtering and Censorship," pp. 1082-1087, 2008.
- [5] J. Zhang, P. Drive, and J. Qin, "The Role of URLs in Objectionable Web Content Categorization 1," 2006.
- [6] S. Sen, "Adult Website Classifier Url Text Classification Url Features," 2010.
- [7] Z. Chen, O. Wu, M. Zhu, and W. Hu, "A Novel Web Page Filtering System by Combining Texts and Images," no. 95, pp. 0-3, 2006.
- [8] Z. Gao, G. Lu, H. Dong, S. Wang, H. Wang, and X. Wei, "Pornographic Web Filtering," pp. 270-273, 2008.
- [9] Y. Lin, H. Tseng, and C. Fuh, "Pornography Detection Using Support Vector Machine," *16th IPPR Conference on Computer Vision, Graphics and Image Processing (CVGIP 2003)*, vol. 19, no. Cvgip, pp. 123-130, 2003.
- [10] J. I. Agbinya, B. Lok, Y. S. Wong, and S. Da Silva, "Automatic Online Porn Detection and Tracking," *Faculty of Engineering, University of Technology, Sydney, 1 Broadway, Sydney 2007*, 2007.
- [11] X. Wang, C. Hu, and S. Yao, "A Breast Detecting Algorithm for Adult Image Recognition," pp. 341-344, 2009.
- [12] X. Shen, W. Wei, and Q. Qian, "The Filtering of Internet Images Based on Detecting Erotogenic-part," no. Incn, pp. 3-7, 2007.
- [13] Y. Wang, J. Li, H. Wang, and Z. Hou, "Automatic Nipple Detection Using Shape and Statistical Skin Color Information," *Proceeding MMM'10 Proceedings of the 16th international conference on Advances in Multimedia Modeling Pages 644-649 Springer-Verlag Berlin, Heidelberg ©2010*, pp. 644-649, 2010.
- [14] X. Kejun and W. Jian, "Automatic Nipple Detection Using Cascaded AdaBoost Classifier," vol. 1, no. 3, pp. 1-6, 2012.
- [15] V. Thaweekote, P. Songram, and C. Jareanpon, "Automatic Nipple Detection based on Face Detection and Ideal Proportion Female using Random Forest," no. C, pp. 11-15, 2013.
- [16] P. Viola, "Rapid Object Detection using a Boosted Cascade of Simple Features," *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, no. 2001, pp. I-511 - I-518 vol.1, 2001.
- [17] C. P. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection," in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, 1998, pp. 555-562.
- [18] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," *Proceedings. International Conference on Image Processing*, vol. 1, pp. 1-900-1-903, 2002.