Journal of Computations & Modelling, vol.3, no.1, 2012, 153-167
ISSN: 1792-7625 (print), 1792-8850 (online)
Scienpress Ltd, 2013

Multicolor Skin Modeling with Application to Skin Detection

Samy Sadek[1], Mahmoud A. Mofaddel[2] and Bernd Michaelis[3]

Abstract

In literature, different researchers consistently argue different color models to be the best choice for skin detection. However, to the best our knowledge, no significant work has been reported in the literature that attempted to utilize more than one color model for skin detection and evaluate the performance for identifying adult image contents. In this paper, we propose a rapid statistical framework for skin detection with an application of adult image identification, based on Multicolor Skin Modeling (MSM). From a high level, the proposed approach proceeds in two consecutive steps (levels). At the first step, the underlying goal is to identify and isolate skin regions of interest (ROI) in each image. At the second step, the suspected skin regions are fed into a specialized statistical analyzer which first extracts some statistical features form these regions. Then, an SVM classifier is used on the extracted features to verify the presence/absence of an adult content in images. Quantitative evaluation shows that our method compares favorably with the state-of-the-art methods in terms of detection rate and false alarm, while reducing the computational costs by at least a factor of six compared with Forsyth's method.

[1] Otto-von-Guericke-University Magdeburg (IESK), Magdeburg, Germany
[2] Department of Math. & Comp. Sci., Faculty of Science, Sohag University, Egypt
[3] Otto-von-Guericke-University Magdeburg (IESK), Magdeburg, Germany
e-mail: samy.technik@gmail.com and samy.bakheet@gmail.com

Article Info: *Received* : December 21, 2012. *Revised* : January 28, 2013
Published online : March 31, 2013

Keywords: Skin detection; adult image content; object recognition; content-based image retrieval

1 Introduction

Over the course of the past two decades protecting children from harmful materials on the Internet such as adult contents has received a great deal of attention and is still a matter of great concern for many researchers in the field of computer vision and pattern recognition. The pioneering work on identifying nude pictures is that of Forsyth *et al.* [1]. In this approach, the images containing large areas of skin are first identified. Then by using a specially defined human structure, the images containing human figures are recognized as nude images. In their approach, only the images that are at least covered in one third by the skin area are fed to a geometric analyzer. While the images used in [1] are 128×192, it takes about six minutes for the method to process a single image. Though, the final recall of the method is less than 50%. Thereafter, additional contributions were made in this field [2, 3, 4, 5, 6, 7]. Generally speaking, skin detection plays an important role in various applications related to computer vision and pattern recognition such as searching and filtering image content on the web, face detection and tracking. Much work has been done in this regard. References [8, 9, 10, 11, 12] discuss detection of human skin and the effect of different cameras, light-setting, human race and color spaces on the recognition process. Furthermore, references [1, 13] use texture information as a component in the skin detection. As a third component in object recognition, many researchers have looked at the shape of the object as a last stage in the information gathering process [14, 15, 16]. Arentz and Olstad [17] have proposed a method for helping to identify adult web sites by using image-contents as a mean to detect erotic material. The average detection error rates for offensive sites were 14.1%, as compared to 9.8% for non-offensive as indicated in their paper. In [18] Veltkamp and Hagedoorn have written a survey of different state-of-the-art shape matching methods. With a set of features describing the image, such as color and shape for segmented objects, it is possible to build fully system for adult image detection.

The remainder of this paper is structured as follows. In Section 2, the proposed method is described in detail. Experimental setup and results are given in Section 3. Finally, conclusions are drawn in the last section with some discussion on future work.

2 Suggested Methodology

Indeed, it is intuitively plausible that it would be very difficult to find a precise definition of "skin" based on attributes of color alone, which can correctly identify skin pixels and non-skin pixels in all cases since there are too many outside factors such as lighting conditions that might change the apparent color of skin, and of course, different people have different colore skin. In addition, objects in the background may have the same color as a person's skin and there is no clear way of telling the difference using these methods.

In order to deal with or tackle such problems, it seems reasonable to take an approach that uses a set of color model to get a new skin detection algorithm that gives higher accuracy. It was stated that using a set of color spaces is a good idea to precisely extract a more defined skin region. In the implementation of the proposed MSM, there are some main steps viz. 1) apply skin filter to identify the skin regions and apply threshold if necessarily, 2) apply median filter to get rid of impulsive salt-pepper noise and 3) the skin-colored regions of areas smaller than predefined threshold, are deleted. These regions are too small to be counted as human skin. An overview of the MSM is shown in Figure 1.

2.1 Multicolor skin modeling

As stated earlier, a robust skin detector is the primary need of many fields of computer vision, including face detection, gesture recognition, and adult content filtering. In the following subsections, a statistical multi-color model for skin detection that makes use of RGB, $Normalized\ RGB$, $YCrCb$ and HSI color spaces is described.

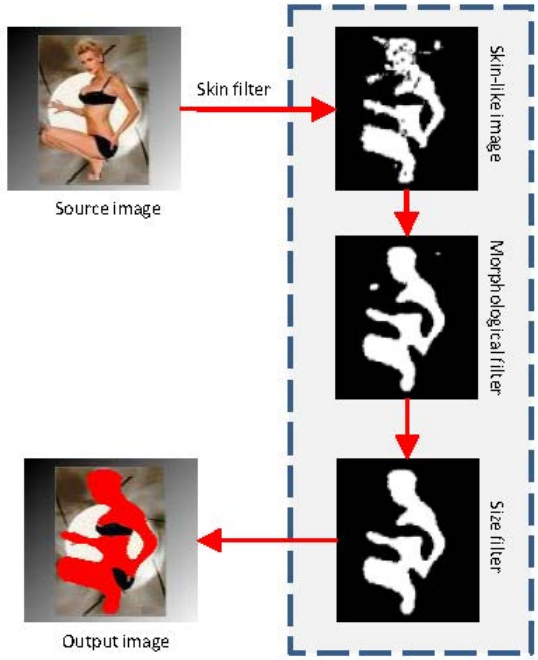

Figure 1: An overview of multi-color skin modeling

2.1.1 Skin filter in RGB space

Indeed, one of the simplest methods for detecting skin pixels is to use an explicitly defined skin region. The simplicity of these methods has attracted (and still does) many researchers [19, 20]. Normally, in these methods, the pixels for skin region can be detected by defining explicitly (through a number of rules) the boundaries skin cluster in RGB color space which provides a fast means of skin detection. For example:

$$\begin{aligned}&\text{if } r > r_m, g > g_m, b > b_m, r - b > \alpha, r - g > \beta \\ &\text{then } skin(r, g, b) = 1 \\ &\text{else } skin(r, g, b) = 0\end{aligned} \tag{1}$$

where r_m, g_m, b_m, α and β are constants that are estimated from the training data. The obvious advantage of this algorithm is the simplicity of skin detection rule which leads to the establishment of a very rapid classifier. The main difficulty to achieve high recognition rates with this approach is the need to find precise decision rules empirically.

2.1.2 Skin filter using RGB channels ratio

It was observed that pixels belonging to skin region regularly contain a significant level of red. Using this observation certain values of the two ratios: $g = G/R$ and $b = B/R$ can be used as skin presence indicator [21]. Formally speaking, let the thresholds be chosen as (g_1, g_2), (b_1, b_2) then a pixel is classified to have skin tone if $g \in (g_1, g_2)$ and $b \in (b_1, b_2)$.

2.1.3 Skin filter in normalized RGB space

The skin detection algorithm used here is based on normalized RGB, or chromaticity space. The chromaticities are defined as

$$\begin{aligned} r &= \frac{R}{R+G+B} \\ g &= \frac{G}{R+G+B} \end{aligned} \tag{2}$$

The reason for using this color space is due to evidences that the human skin color is more compactly represented in it than it is in other color spaces, such as RGB, HSI, SCT and YQQ[22]. Here, we use the chromaticities r and g to describe the color. Skin color distribution can be modeled by an elliptical Gaussian joint probability density function (pdf), defined as:

$$p_{skin}(x) = \frac{1}{2\pi|\Sigma_s|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_s)^T \Sigma_s^{-1}(x-\mu_s)} \tag{3}$$

Here, x is a color vector and μ_s and Σ_s are the distribution parameters (mean vector and covariance matrix respectively). The model parameters are estimated from the training data using the following formulas,

$$\mu_s = \frac{1}{n}\sum_{k=1}^{n} x_k \tag{4}$$

$$\Sigma_s = \frac{1}{n}\sum_{k=1}^{n}(x_k - \mu_s)^T(x_k - \mu_s) \tag{5}$$

where n is the total number of samples and x_k is the vector representing the k sample. The probability can be used directly as the measure of how "skin-like" the color is [23]. Gaussian skin modeling has been also employed in [24, 22, 25].

2.1.4 Skin filter in $YCrCb$ space

In this color space, the two chroma components Cr, and Cb can be efficiency used to define explicitly skin region. The thresholds be chosen as (Cr_{max}, Cr_{min}) and (Cb_{max}, Cb_{min}) , a pixel is classified as skin pixel if the values (Cr, Cb) fall within the thresholds.

2.1.5 Skin filter in HSI space

Skin filter in HSI space The value of Hue and saturation (H, S) are adequate to segment the skin region from non-skin region. To detect skin region, the values of H, and S should be explicitly determined through defining their intervals as shown below:

$$
\begin{aligned}
&\text{if } H \in (H_{min}, H_{max}) \wedge S \in (S_{min}, S_{max}) \\
&\text{then } skin(r, g, b) = 1 \\
&\text{else } skin(r, g, b) = 0
\end{aligned} \tag{6}
$$

where $H_{min}, H_{max}, S_{min}$, and S_{max} are experimentally determined using the training data.

2.2 Adult image content detection

One interesting application of skin detection is as part of a larger system for detecting adult content in photos. An adult content recognizer that works reliably to recognize adult image contents could be a valuable tool for image search services in digital libraries [26], as well as for image categorization. The main goal of the proposed adult content recognition system is to determine whether or not an input image contains an adult content by feeding the output of the skin detector to an SVM classifier, which attempts to classify image contents based on simple few statistical feature extracted from the image contents.

The main steps involved in the proposed system for adult image recognition can be summarized as,

1. *Step-1 (Skin Detection):* The first main step in building a system for adult images identification is the skin-colored regions segmentation. This stage can be accomplished by the proposed MSM described previously at the beginning of this section.

2. *Step-2 (Analysis):* In this stage, the skin colored regions are fed to a specialized statistical analyzer which first extracts few features from the skin regions. Then, an SVM classifier is used in order to decide whether or not any image regions contains an adult content. Figure 2 shows a simple flowchart of these two steps of the adult content detection.

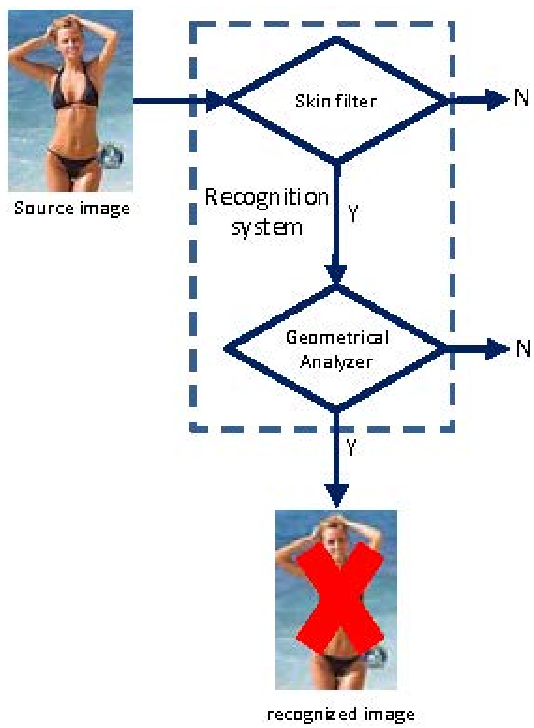

Figure 2: A simple flowchart diagram showing the main steps of the approach

By taking advantage of the fact that adult images are often sized to frame a standing or reclining figure, some features can be computed from the output image of the skin detector. These features play an important role in increasing overall recognition rate of the proposed system, which include: 1) edge of connected components of skin, 2) height and width of the largest region of skin, 3) percentage of pixels detected as skin, and 4) number of connected components of skin These features can all be computed in a single pass, before feeding the image into the geometric analyzer that allows the adult content detection system to be extremely fast compared with the Forysth's system counterpart [20].

Features extraction

The main objective of the feature extraction is to capture the most relevant and discriminant characteristics from the image contents to be recognized. This relevant information allows for discrimination between the two image content classes (i.e., adult, non-adult). It is worthwhile to mention here that for recognition systems based the features extracted from skin regions, it was experimentally found that portrait images present a big challenge for such systems, as these images show abundance of skin color as adult images. Therefore, for the purpose of differentiating between adult images and portrait ones, significant skin regions are firstly approximated by ellipses. Then, statistical features that describe image objects are extracted based on the best-fit ellipsoids. These features includes: percentage of skin in the entire image, total number of skin regions, relative height and width of the largest skin region, skin percentage in the largest skin region, etc. Finally, feature vectors are normalized to zero mean and unit standard deviation before feeding into the SVM classifier for final decision.

SVM classification

In this section, we formulate the adult content recognition task as a binary class learning problem, where there are two categories of image contents (i.e., adult and non-adult), and the goal is to classify a new image to one and only one of these two classes. There are various supervised learning algorithms by which an adult content recognizer can be trained. Support Vector Machines (SVMs) are used in our framework due to their outstanding generalization capability and reputation of a highly accurate paradigm. SVMs [27] are based on the Structure Risk Minimization principle from computational theory, and are a solution to data overfitting in neural networks.

Originally, SVMs were designed to handle dichotomic classes in a higher dimensional space where a maximal separating hyperplane is created. On each side of this hyperplane, two parallel hyperplanes are conducted. Then SVM attempts to find the separating hyperplane that maximizes the distance between the two parallel hyperplanes (see Figure 3). Intuitively, a good separation is achieved by the hyperplane having the largest distance. Hence the larger the margin the lower the generalization error of the classifier. More formally, let

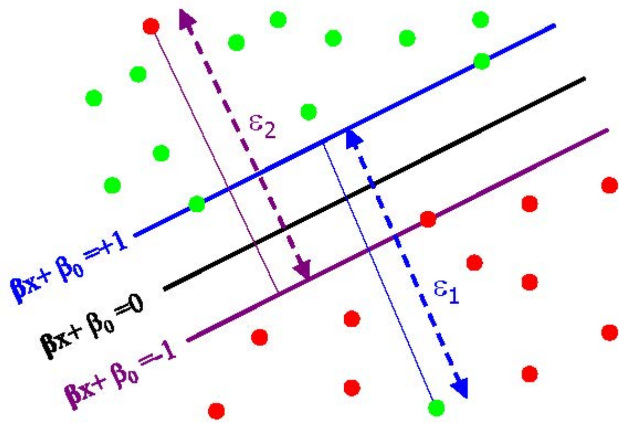

Figure 3: Generalized optimal separating hyperplane

$\mathcal{D} = \{(\mathbf{x}_i, y_i) \,|\, \mathbf{x}_i \in R^d, y_i \in \{-1, +1\}\}$ be a training dataset, Coretes and Vapnik [27] show that this problem are best addressed by allowing some examples to violate the margin constraints . These potential violations are formulated using some positive slack variables ξ_i and a penalty parameter $C \geq 0$ that penalize the margin violations. Thus the optimal separating hyperplane is determined by solving the following QP problem:

$$\min_{\boldsymbol{\beta}, \beta_0} \quad \frac{1}{2}\|\boldsymbol{\beta}\|^2 + C \sum_i \xi_i \tag{7}$$

$$\text{subject to} \quad (y_i(\langle \mathbf{x}_i, \boldsymbol{\beta} \rangle + \beta_0) \geq 1 - \xi_i \quad \forall i) \wedge (\xi_i \geq 0 \quad \forall i).$$

Geometrically, $\boldsymbol{\beta} \in R^d$ is a vector going through the center and perpendicular to the separating hyperplane. The offset parameter β_0 is added to allow the margin to increase, and to not force the hyperplane to pass through the origin that restricts the solution. For computational purposes it is more convenient to solve SVM in its dual formulation. This can be accomplished by forming the Lagrangian and then optimizing over the Lagrange multiplier $\boldsymbol{\alpha}$. The resulting decision function has weight vector $\boldsymbol{\beta} = \sum_i \alpha_i \mathbf{x}_i y_i$, $0 \leq \alpha_i \leq C$. The instances $\mathbf{x}_i$ with $\alpha_i > 0$ are called *support vectors*, as they uniquely define the maximum margin hyperplane. In the current approach, two classes of image contents are created. An SVM classifier is trained using the features extracted from images in the training dataset. All feature vectors are then fed into the SVM classifier for the final decision.

3 Experimental Results

Due to the lack of any standard image databases for testing and comparison of adult content detection systems, the proposed system has been tested using our own images database. This database consists of 562 test adult images and 1580 assorted control control images, containing some images of people but none of adult images. All images were taken of type RGB format 8 bits / pixels in each color channel. The test images were collected from the internet. They show a very wide range of postures. Some depict several people including naked or scantily-dressed. Some depict only small parts of the bodies of one or more people. Most people in the images are Caucasian; a small number are Blacks or Asians. To evaluate the results of the proposed adult content detector two different metrics are used. TP (true positive) is the number of adult images identified correctly divided by the number of all test images. FP (false positive) is the number of non-adult images identified as adult images divided by the number of all control images. Receiver operating characteristics (ROC) curve that shows the relationship between correct detections and false detections of the proposed adult content detector as a function of the detection threshold is shown in Figure 4.

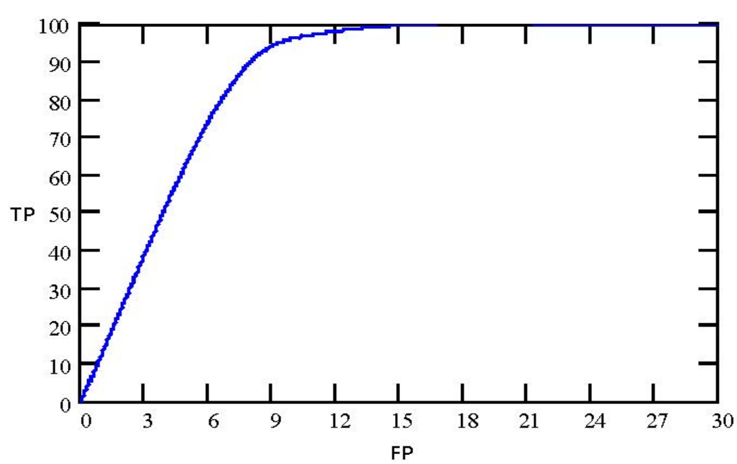

Figure 4: ROC curve of the proposed adult content detector

Figure 5 shows the capability of the proposed system to detect a sample of clothed images in the test set. These images correctly classified as containing adult content. Mistakes by the proposed pornography detector may occur for

Table 1: Comparison with other state-of-the-art systems

Method	TP	FP
Proposed Method	**89.3%**	**7.4%**
Arentz-Olstad [17]	85.9%	9.8%
Jones-Rehg [28]	86.7%	9%
Forsyth-Fleck [1]	43%	4.2%

several reasons. In some images, adult contents are too small to be detected. In others, most or all of the skin area is desaturated, so it fails the skin detector.

Some control images pass the skin detector because they contain people, particularly several close-up portrait shots. Other control images contain material whose color closely resembles that of human skin; particularly wood, sand, and skin or fur of certain animals.

The proposed recognition system can be variously configured. In one configuration of the system, it can successfully identify 89.3% of the test images (true detection), but only 7.4% of the control images (false alarm). Table 1 provides a quantitative comparison between our system and those of other investigators in the literature. As can be clearly seen from the table, our method yields encouraging results and is relatively superior compared to the other methods. Such results have the potential to compare favorably with those reported in the existing literature in terms of detection rate and false alarm rate. Furthermore, our system is relatively fast; first-stage,i.e. skin detection, takes a trivial amount of time while the geometric analyzer processes pictures at a rate of less than a minute per picture (approx. 40 sec). On the other hand the method developed by Forsyth and Fleck [20] takes about 6 minutes on a workstation for the figure grouper to process a suspect image passed by the skin filter.

Figure 5: Performance obtained by the proposed adult content detector: (i) source images; (ii) result of skin filter; (iii) result of adult-image recognizor.

4 Conclusions and Future Work

In this paper, we have introduced a statistical method to detect and identify adult contents in color images using a combination of a simple visual color cue and geometric human figure characteristics. The proposed system can be flexibly configured to meet the users needs. The results obtained showed that our method performs comparably well as or better than the existing methods in terms of detection rate and false alarm rate, while it is simple and relatively rapid. The future work will be along two major axes. The first will be the further improvement of the method by using more elaborated feature sets; for example, by adding texture features, while the second will be to examine the robustness of the method by performing more experiments on real-word video data.

Acknowledgements. The doctoral scholarship provided by University of Sohag (Egypt) to the first author is gratefully acknowledged.

References

[1] David A. Forsyth and Margaret M. Fleck, Identifying nude pictures, *IEEE Workshop on the Applications of Computer Vision*, (1996), 103-108.

[2] S. Sadek and A. Al-Hamadi and B. Michaelis and U. Sayed, A Robust Neural System for Objectionable Image Recognition, *International Conference on Machine Vision (ICMV'9)*, Dubai, UAE, (2009), 32-36.

[3] H. Zhu and S. Zhou and J. Wang and Z. Yin, An algorithm of pornographic image detection, *International Conference on Image and Graphics (ICIG'07)*, (2007), 801-804.

[4] L.-L. Cao and X.-L. Li and H.-H. Yu and Z.-K. Liu, Naked people retrieval based on adaboost learning, *First International Conference on Machine Learning and Cybernetics*, Beijing, China, (2002), 1133-1138.

[5] R. Schettini and C. Brambilla and C. Cusano and G. Ciocca, On the detection of pornographic digital images, *Visual Communications and Image Processing*, **5150**, (2003), 2105-2113.

[6] Y.-C. Lin, H.-W. Tseng and C.-S. Fuh, Pornography detection using support vector machine, *16th IPPR Conference on Computer Vision, Graphics and Image Processing (CVGIP 2003)*, (2003), 123-130.

[7] Deselaers and L. Pimenidis and H. Ney, Bag-of-Visual-Words Models for Adult Image Classification and Filtering, *ICPR*, (2009).

[8] E. Angelopoulou, Understanding the color of human skin, *SPIE Conf. On Human Vision and Electronic Imaging VI (SPIE)*, **4299**, (2001), 243-251.

[9] L. Bretzner, I. Laptev and T. Lindeberg, Hand gesture recognition using multi-scale color features, hierarchical models and particle filtering,

Proc. 5th IEEE International Conference on Automatic Face and Gesture Recognition, Washington D.C., (2002).

[10] R.-L. Hsu, M. Abdel-Mottaleb and A. Jain, Face detection in color images, *IEEE Trans. Patt. Anal. Mach. Intell*, (2002), 696-706.

[11] M.C. Shin, K.I. Chang and L.V. Tsap, Does color space transformation make any difference on skin detection?, *IEEE Workshop on Applications of Computer Vision*, Orlando, FL, (2002), 275-279.

[12] M.A. Mofaddel and S. Sadek, Adult Image Content Filtering: A Statistical Method Based on Multi-Color Skin Modeling, *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT'10)*, Luxor, Egypt, (December, 2010), 366-370.

[13] J.M. Rehg, K.P. Murphy and P. W. Fieguth, Vision-based speaker detection using Bayesian networks, *IEEE Conf. on Computer Vision and Pattern Recognition*, Ft. Collins, Colorado, (1999), 110-116.

[14] Qing-Fang Zheng, Wei Zeng, Wei-Qiang Wang and Wen Gao, Shape-based Adult Image Detection, *International Journal of Image and Graphics*, **6**(1), (2006), 115-124.

[15] J.S. Park and D.H. Chang, 2-d invariant descriptors for shape-based image retrieval, *Pattern Recognition Letters*, (2001).

[16] M. Safar, C. Shahabi and X. Sun, Image retrieval by shape: a comparative study, *IEEE Internat. Conf. on Multimedia and Expo*, (2000), 141-144.

[17] W.A. Arentz and B. Olstad, Classifying offensive sites based on image content, *Computer Vision and Image Understanding*, (2004), 295-310.

[18] R. Veltkamp and M. Hagedoorn, State-of-the-art in shape matching, *Technical Report, UU-CS-1999-27, Utrecht University, the Netherlands*, (1999).

[19] P. Peer, J. Kovac and F. Solina, Human skin color clustering for face detection, *EUROCON 2003- International Conference on Computer*, (2003).

[20] David A. Forsyth and Margaret M. Fleck, Automatic detection of human nudes, *International Journal of Computer Vision*, (1999), 63-77.

[21] J. Brand and J.A. Mason, Comparative assessment of three approaches to pixel level human skin-detection, *Proc. of the International Conference on Pattern Recognition*, **1**, (2000), 1056-1059.

[22] E. Littmann and H. Ritter, Adaptive color segmentation, A comparison of neural and statistical methods, *IEEE Transactions on Neural Networks*, **8**(1), (1997), 175-185.

[23] B. Menser and M. Wien, Segmentation and tracking of facial regions in color image sequences, *Proc. SPIE Visual Communications and Image Processing*, (2000), 731-740.

[24] U. Sayed, S. Sadek and B. Michaelis, Two Phases Neural Network-Based System for Pornographic Image Classification, *5th Int. Conf. of Sciences of Electronic, Technologies of Information and Telecommunications (SETIT'9)*, (2009), 1-6.

[25] H. Kruppa and M.A. Bauer and B. Schiele, Skin Patch Detection in Real-World Images, *DAGM Symposium*, Zurich, Switzerland, (2002), 109-116.

[26] J.Z. Wang, J. Li, G. Wiederhold and O. Firschein, Classifying objectionable websites based on image content, *Notes in Computer Science, Special issue on interactive distributed multimedia systems and telecommunication services*, **21**(15), (1998), 113-124.

[27] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.

[28] M. Jones and J. Rehg, Statistical Color Models with Application to Skin Detection, *Int. Journal of Computer Vision*, **46**(1), (2002), 81-96.