IEICE TRANS. INF. & SYST., VOL.E95–D, NO.9 SEPTEMBER 2012

PAPER

Detecting Objectionable Images Using a New Skin Detection Method

Ali NADIAN GHOMSHEH[†a)], *Member and* **Alireza TALEBPOUR**[†], *Nonmember*

SUMMARY In this paper, a new skin detection method using pixel color and image regional information, intended for objectionable image filtering is proposed. The method consists of three stages: skin detection, feature extraction and image classification. Skin detection is implemented in two steps. First, a Sinc function, fitted to skin color distribution in the Cb-Cr chrominance plane is used for detecting pixels with skin color properties. Next, to benefit regional information, based on the theory of color image reproduction, it's shown that the scattering of skin pixels in the RGB color space can be approximated by an exponential function. This function is incorporated to extract the final accurate skin map of the image. As objectionable image features, new shape and direction features, along with area feature are extracted. Finally, a Multi-Layer Perceptron trained with the best set of input features is used for filtering images. Experimental results on a dataset of 1600 images illustrate that the regional method improves the pixel-based skin detection rate by 10%. The final classification result with 94.12% accuracy showed better results when compared to other methods.

key words: *skin detection, bidirectional reflection distribution function, Objectionable image filtering*

1. Introduction

Many internet users are potential victims of being exposed to unwanted pornographic content through SPAM E-mail or commercial pop-up pages. But the problem becomes more concerning when knowing that 25% of children who gain access to the internet have viewed such content [1]. Filtering such content is conducted in many countries of the world, but the level of filtering of course varies in different communities. Over-blocking of normal web pages as a result of misclassification is a big problem when filtering is applied and reducing the error rate is crucial [2].

Two approaches are incorporated in order to decide on the content of web pages. The first method, applied in European countries, uses black lists, updated manually based on human intelligence [3]. This method is not very efficient since it needs continuing human efforts to keep the list up to date. The second method, Dynamic filtering, can be incorporated to solve these objections, where information flow is checked for potential data that needs to be filtered. [4]. Dynamic filtering can be approached by incorporating textual and visual content of information flow. If textual features are used, a number of keywords are extracted from the web page, and if the frequency and plurality of the offending words exceed a certain threshold, that page is blocked [5]. When visual features are used, presence of human skin is the backbone for feature extraction and classification of images into normal and harmful pages [6].

To block objectionable images, [7] used a simple Bayesian classifier to detect skin regions in images. From these regions they extracted texture features by employing Daubechies wavelets and used them for classification by image retrieval techniques. In [8] a maximum entropy model was built to detect skin pixels. Local and global ellipses were fitted to skin regions and number of parameters calculated from the fitted ellipses served as the feature vector which was utilized for image classification. In [9] a learning-based chromatic-matching scheme was used for skin detection; area, shape, and location features were extracted and Adaboost classifier was used for decision making afterwards. In [10] HSV color space and MPEG-7 edge descriptors [11] were employed to extract features from skin regions and image retrieval technique was used to detect harmful images. In some objectionable image filtering systems, textual and visual features are combined together for image filtering [12]. The WebGaurd system incorporated Support Vector Machines to classify images based on textual and visual features [13]. In [14], object's contour-based features extracted for image recognition, along with continues text features were utilized for image filtering.

Accurate detection of skin regions is very important for objectionable image filtering, because it is employed in all filtering methods. Skin detection can be divided into two categories: pixel-based and regional-based skin detection [15]–[17]. Explicit rules, Bayes classifier, statistical models such as: Gaussian distribution and GMM and the Sinc function have been widely used for pixel-based skin detection [15], [18]–[20]. It has been shown that pixel-based methods are very fast, however, the accuracy of these methods is not very high [15], [20]. Low detection rate makes these methods unsuitable for objectionable image filtering, if they are solely used for skin detection.

Regional-based methods on the other hand incorporate different techniques such as: Gabor wavelets [21], Contourlets [22], histogram matching [23], and physics based approaches to extract regional features for skin detection [24], [25]. These methods require complex processing steps, which is not preferred for the application of objectionable image filtering. Because in practice, most images viewed in internet are normal images. Complex processing for all images increases the overall time spent to detect objectionable

Manuscript received November 21, 2011.
Manuscript revised April 15, 2012.

[†]The authors are with the Shahid Beheshti University, GC, Tehran, Iran.
a) E-mail: a_nadian@sbu.ac.ir

DOI: 10.1587/transinf.E95.D.2288

images.

Conventional pornographic image detection methods use pixel-based or regional-based skin detection method independently. As reviewed above, pixel-based skin detection is a very time efficient approach for finding a prior estimate of skin regions and regional-methods, although are not as fast as pixel-based methods, they are very efficient for finding skin regions with high accuracy.

To overcome these limitations, and also benefit from the advantages of both pixel-based and region-based methods, this paper proposes a new skin detection method that combines pixel-based and regional-based method together to improve the skin detection rate. In the first step, to find an initial estimate of skin pixels in the image, a 2D Sinc function was used to model the skin color distribution in the Cb-Cr chrominance plane. In the second step, physics of color reproduction was utilized to show that, the scattering of the skin pixels in the RGB color space can be approximated with an exponential function. Based on this illustration, the histogram of the initial skin pixels was approximated by an exponential function. The distance between each color value and the exponential function was used as the metric to detect the final skin region of the image.

The advantages of this method, apart from obtaining higher accuracy in each step of skin detection compared to conventional methods, is the ability to reduce much redundant processing by discarding many images with none-skin regions after first step of the skin detection. The final skin detection method is also reliant on the color information present in the image, not on the prior skin color distribution. This will help to reduce the false positive detection rate.

After skin detection, from the final skin mask, three types of features are extracted to describe the detected skin region: area, shape and direction (a total of 271 features). In the classification phase, ten different combinations of features were fed to ten MLPs, and based on the training results the best trained MLP was used for classification. For train and test purposes, 1600, images equally divided into two classes of normal and objectionable, were used.

The rest of the paper is organized as follows: Sect. 2 explains the two steps used for skin detection in detail. Section 3 describes features extracted for image classification. Experimental results are explained in Sect. 4, and the paper is concluded in Sect. 5.

2. Skin Detection

When visual features are used for objectionable image filtering, accurate detection of skin regions is the most critical step. Accurate skin detection, apart from excluding images that have false skin-like regions (regions that don't belong to human skin but have a similar color), makes the features more effective in the classification phase.

Different steps of the proposed algorithm for objectionable image filtering are shown in Fig. 1. In the skin detection stage, after each step, images with small skin area are classified as normal images. By completing the second step skin detection, the skin map of the image is extracted. This mask is then used for feature extraction. The extracted features are then utilized to classify an image as normal or objectionable. The two steps used for skin detection are next explained in detail.

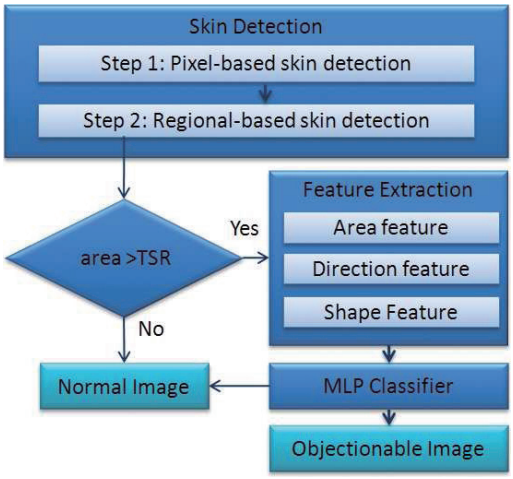

Fig. 1 Block diagram of the proposed method.

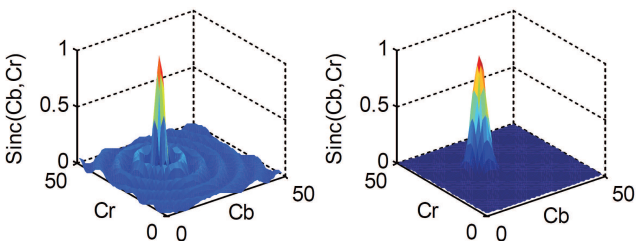

Fig. 2 2D Sinc function: (right) *n=1* and (right) *n=5*.

2.1 Pixel-Based Skin Detection

When pixel-based skin detection methods are incorporated, two points have to be considered: (1) the color space representing the skin color distribution and (2) the function that models this distribution. It has been shown that YCbCr is a better space when chrominance information is used for skin detection [15], [20]. Thus, this color space is used for pixel-based skin detection. We have shown that the Sinc function is more effective compared to statistical models. Thus, it was used in this paper to model the skin color distribution [20]. To use this function, it is required to find its parameters so that it fits to the skin color distribution in the Cb-Cr chrominance plane. The Sinc function raised to the power of a positive integer n, loses its oscillation and depending on n, the major envelope of the Sinc function changes its width. Figure 2 shows a 2D Sinc function for $n=1$, and $n=5$.

The 2D Sinc can be formulated as:

$$f(Cb, Cr) = sinc(\sqrt{m_1(Cb - t_1)^2 + m_2(Cr - t_2)^2})^2 \quad (1)$$

To fit the Sinc function to the skin histogram, a surface

fitting process is required to minimize the cost function:

$$CF=\sum_{Cb=1}^{256}\sum_{Cr=1}^{256}(f(Cb,Cr)-h(Cb,Cr))^2 \quad (for\, h(Cb,Cr\neq 0)) \quad (2)$$

where *h(Cb,Cr)* is the histogram of skin color distribution in the Cb-Cr plane. The optimization process tries to minimize the distance between the 2D Sinc function and the skin color distribution. Five variables defined in Eq. 2, represented with vector $X = [m_1, m_2, t_1, t_2, n]$, have to be found in the minimization process. The steepest descent method is a well-known algorithm for solving such minimization problems [26].

In an iterative process, the steepest descent method updates X in the inverse direction of the gradient of the cost function. At each step X is updated by:

$$X^{K+1} = X^K + \alpha s^K \quad (3)$$

where α shows the pace in each step, and $S=-\nabla f$ is in the inverse direction of the gradient of CF with respect to X in iteration K and is denoted by:

$$\nabla f = \left[\frac{\partial f}{\partial m_1}\frac{\partial f}{\partial m_2}\frac{\partial f}{\partial t_1}\frac{\partial f}{\partial t_2}\frac{\partial f}{\partial n}\right] \quad (4)$$

The iterative algorithm converges when $X^{K+1}-X^K<eps$. Since CF is raised to the power of two, the convergence to the global minimum is guaranteed. A proof of convergence for such functions is presented in [26].

After calculating the optimized coefficients of the Sinc function it can be used for skin detection. For an input chrominance pair (Cb, Cr), $f(Cb, Cr)$ is calculated. This value is defined as the skin probability of (Cb, Cr) and is denoted by P_s. If $P_s > P_\theta$ (a predefined threshold) then the pixel is classified as a skin. To avoid calculation of P_s for all the pixels in the image, a Look Up Table (LUT) which is filled with prior calculated P_s is stored. Thus, for each pixel, P_s is directly read from the LUT. By finding P_s for all image pixels, the initial skin mask of the image is extracted. The images with high skin area are moved to the next step for regional-based skin detection. The images with none or small skin area are regarded as normal images and are not further processed.

2.2 Regional-Based Skin Detection

After pixel-based skin detection, a regional-based method is used to refine the results. In general, three errors might occur in the Pixel-based stage:

1. Non-skin regions are detected as skin.
2. The skin region is detected partly.
3. The existing skin region is not detected at all.

By combining the regional stage with the pixel-based stage, it is possible to increase the detection accuracy by correcting errors 1 and 2.

In order to explain the idea of the proposed regional method, the physics of color image reproduction is explained briefly. The Intensity measured by a camera sensor can be formulated as:

$$I(X_l) = \int E(\lambda, X_l)S(\lambda)d\lambda \quad (5)$$

where I is the intensity measured by the sensor. E is the irradiance of wavelength λ falling onto an infinitesimal small patch on the sensor array located at position X_l. Irradiance is a measure of power per square meter area $W/m2$), and $S(\lambda)$ is a vector of the senor's response time. Irradiance is directly proportional to the radiance given by the object being captured in a scene. Assuming a scaling factor one for simplicity, then one leads to $E(\lambda, X_l) = L(\lambda, X)$, where L is the radiance showing the amount of light radiated per unit area, per unit angle from the surface of an object.

Using this information, then Eq. 5 can be written as:

$$I(X_l) = \int E(\lambda, X_{obj})S(\lambda)d\lambda \quad (6)$$

where, I is the integration of sensor response to the radiated light from the surface of an object at position X_{obj} for all wavelengths. If reflectance models are used, then Bidirectional Reflectance Distribution Function (BRFD) assuming Lambertian surface, is a constant that shows the fraction of the incident light that is reflected for any given wavelength λ [27]. This constant, reflectance, is denoted by $Rf(\lambda, X_{obj})$. Radiance given off by a surface is the product of the radiance given by the light source, reflectance and a scaling factor defined as the normal vector of surface and direction of the light source denoted by $Gf(X_{obj})$, i.e.:

$$L(\lambda, X_{obj}) = Rf(\lambda, X_{obj})L(\lambda)Gf(X_) \quad (7)$$

which follows:

$$I(X_l) = \int Rf(\lambda, X_{obj})L(\lambda)Gf(X_{obj})S(\lambda)d\lambda \quad (8)$$

The sensor response can be approximated by delta function [28]; $S(\lambda) = (\lambda - \lambda_i)$; then:

$$I_i(X_l) = Rf(\lambda, X_{obj})L(\lambda_i)Gf(X_{obj}) \quad (9)$$

where $I_i(X_l)$ is the intensity measure by sensor i at position X_l on the sensor array. From this equation it is evident that the illuminant only scales the response of the sensor. If the Gamma factor is assumed to be one, then the intensity values stored in the image $c = [c_r, c_g, c_b]$ would be obtained as:

$$c_i(x, y) = I_i(x, y) = Gf(x, y)Rf_i(x, y)L_i(x, y) \quad (10)$$

where object position X_{obj} is imaged into image pixel (x, y). If uniform lighting condition is considered Eq. 10 can be simplified into:

$$c_i(x, y) = Gf(x, y)Rf_i(x, y)L_i \quad (11)$$

By Eq. 11, measurements made by the sensor can be viewed as a linear transformation of the incident light.

If an object has the color c_i, represented by color vector

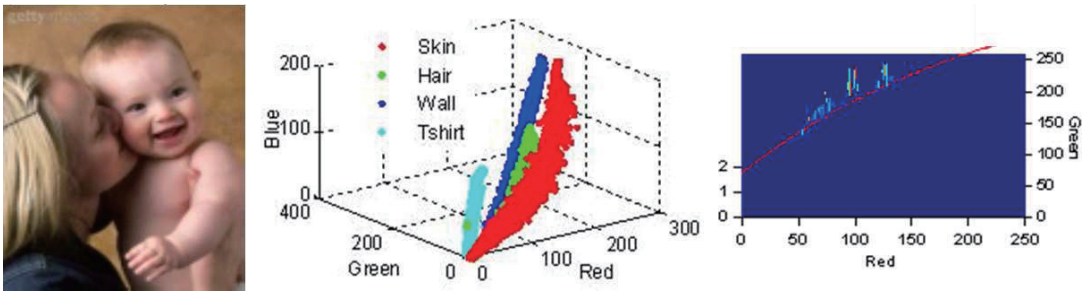

Fig. 3 Color pixel distribution in the RGB color space.

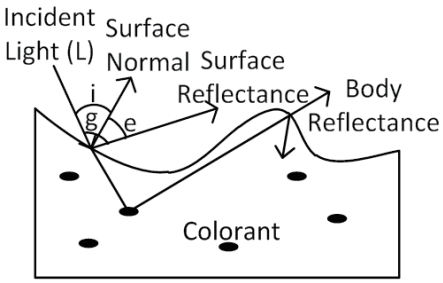

Fig. 4 Dichromatic Reflection Model.

$i \in \{R, G, B\}$, in the 3D RGB color space, then the color that represents point X_{obj} on the object's surface can be defined as the reflected light L_i multiplied by $Gf(x, y)Rf_i(x, y)$:

$$c_{RGB} = Rf_i(x, y)Gf(x, y)L_i \tag{12}$$

Figure 3-a shows an image consisting of skin and a background that has skin-like color. For each region in the image, scattering of pixels in the 3D RGB space is also shown (Fig. 3-b). As it can be seen, for all regions except skin, the corresponding pixels of other regions distribute along a line in the space that passes through the origin. Equation 12 can be used to explain the reason for this manner of distribution. For these regions the actual color of each object is only scaled by $Rf_i(x, y)Gf(x, y)$ therefore a straight line is formed in the space. In order to explain what has happened to skin, BRDF is revisited. The Irradiance at X_{obj} can be more exactly defined as a result of two reflections, the surface reflectance L_s, and the body reflectance L_b; therefore Eq. 7 can be rewritten as:

$$L(\lambda, \theta) = L_s(\lambda, \theta) + L_b(\lambda, \theta) \tag{13}$$

Each component of (13) can be formulated as:

$$L(\lambda, X_{obj}, \theta) = Rf(\lambda, X_{obj})L(\lambda)Gf(X_{obj})\rho(\lambda) \tag{14}$$

where θ includes the viewing angle e, the phase angle g, and the illumination direction angle i(Fig. 4). ρ is the material's spectral surface or body reflectance. By Eqs. 9 and 13 it can be implied that the color of pixels for which $L_s \gg L_b$ resembles the color of the incident light. For pixels that $L_s \approx 0$ the color of those pixels shows only the color of the object multiplied by $Gf(x, y)$; therefore such pixels will distribute along a line that passes the origin of the color space (Since $Gf(x, y)$ and $Rf_i(x, y)$ are both scalars). If L_s and L_b are both present in pixels that belong to one object, pixels that show L_b distribute along a straight line that passes through the origin of space, and object pixels that mostly resemble L_s start to deviate from a straight line and bend towards the white point of the space (these are the specular regions of an object). Human skin has many convex and concave regions and also it's found that skin has a ρ_s of 5% [24]; therefore L_s and L_b are both present and skin color scatters in the space starting from the origin of the space and then bending towards the color of the illuminant.

In this paper, the explained scattering behavior of skin pixels is approximated by an exponential function:

$$f(x) = 1 - \exp^{ax+b} \tag{15}$$

where a and b are the parameters that control the shape of the exponential, and they depend on the color of a region and coefficient ρ (for surface and body reflectance). Since the color Blue is not very significant in the skin color thus in this paper it is not considered and only Red and Green components are used. This will simplify the fitting process to finding an exponential on the 2D plane (rather than in a 3D space).

To find the approximated skin distribution, the skin color histogram on the Red-Green ($R - G$) plane is found. If Green is then defined as a function of color Red in the R-G plane, the exponential that has to be fitted to the skin histogram can be written as:

$$R = 1 - \exp^{-(aG+b)} \tag{16}$$

and further simplified into:

$$\ln(R - 1) = -(aG + b) \tag{17}$$

Thus, the fitting process is simplified to a line fitting problem and can be calculated through maximum likelihood estimation [29]. Figure 5 shows steps computed to find the final skin region in an image.

An exponential is fitted to the pixels of each block that more than 50% of its pixels have been detected as skin. If the calculated a and b have values in the expected range, that block is considered as skin. By performing the block processing step, error 3 is corrected. In this step, for each block, if the parameters of the fitted exponential are not in the predefined range, that block is considered as non-skin. This

Fig. 6 Step by step results of the regional-based skin detection algorithm. a) Original image, b) pixel-based skin detection using the Sinc function, c) skin blocks selected and merged, d) post processing and final skin mask.

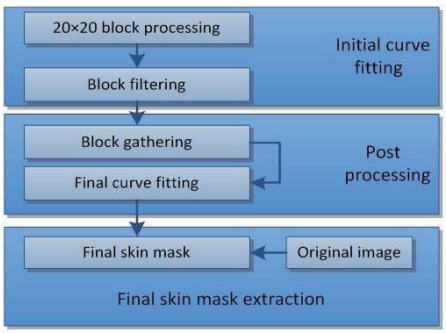

Fig. 5 Diagram of the regional-skin detection process.

results in decreasing the FPR. In this step pixels are classified according to their relative block. Thus a pixel cannot be classified as skin or non-skin independently; therefore this step is considered regional-based.

After block processing, many skin pixel are not detected because they are place in blocks with less skin pixels, or in blocks that are purely conditioned; therefore the values of the approximated exponential for their corresponding blocks does not classify them as skin. To detect such pixels, in a post-processing step, a final exponential is fitted to all those pixels that their corresponding blocks have been detected as skin (Block Gathering). Then if the distance of a pixel is smaller than a predefined threshold T, that pixel is classified as skin. The values for parameters a, b, and T were found by testing them on 400 images in the objectionable image test set and it was found that for skin, a is in an interval [5.7 6], b is in [−1.2 −1.4] and T was set to 35. The final curve fitting step corrects error 2. Here, the calculated exponential curve shows how an individual skin region is distributed in the R-G space. Each pixel is classified as skin or non-skin based on its distance to the calculated exponential curve. In this step, the missed skin pixels in the pixel-based stage can be also recovered. This step improves the accuracy by increasing the TPR and also decreasing the FPR. In this step, parameters of the final curve is dependent on the information of all the skin blocks of the image, thus it uses regional properties of the image to detect the final skin mask.

An output sample after regional-based skin detection is shown in Fig. 6. In Fig. 6-b the result of pixel-based skin detection shows using only color information, many pixels are misclassified as skin. Figure 6-c shows the blocks that where found as skin regions in the image. In Fig. 6-d its shown how the information provided by the correctly detected skin blocks, helps to correctly classify the skin pixels in the image, while the non-skin pixels are correctly detected as non-skin pixels.

3. Feature Extraction

After detecting skin pixels in an image, the binary map of skin, I_{bin}, is extracted. From this map three different types of features are extracted: area, shape, and direction of skin. These features describe the skin map of the image.

3.1 Area Feature

Objectionable images mostly have high skin area and the skin area in an image can be used as a crisp threshold for filtering. To utilize the area feature, morphological operations are employed to find connected areas of the binary image and each connected area is labeled with a different number. From the labeled image the following features are extracted:

n :number of skin regions in the image.

$S(n)$:area of each skin region.

$TSR = \sum_1^n S(n)/A$:Total skin area to image area ratio.

$LSR = S(n)_{max}/A$:Largest skin to image area ratio.

where A is the number of rows in the image times the number of columns of I_{bin}. Area related features are stored in vector $FV_{area} = [n, TSR, LSR]$.

3.2 Direction Feature

To show the direction in which the skin map of the image has distributed along, direction feature is introduced. In this paper Principle Component Analysis (PCA) was used to find the directions for which the skin map has the most and the least direction along (Fig. 7). The covariance matrix of I_{bin}, is a 2×2 matrix and is obtained by:

$$Cov = \frac{1}{K} \sum_{i=1}^{2} \sum_{j=1}^{2} (I_{bin} - \mu_i)^T (I_{bin} - \mu_j) \tag{18}$$

where k is a normalization constant and is equal to $\sum S(n)$. If V and E represent to Eigen vectors and Eigen values of matrix Cov respectively, then V gives a set of two vectors

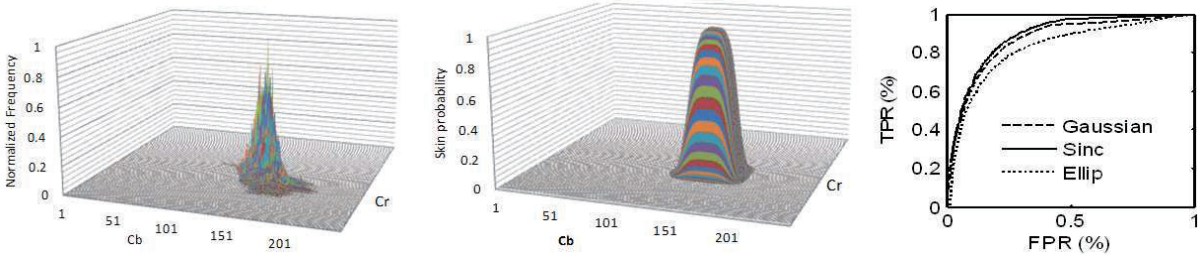

Fig. 7 (a) Skin histogram (b) the fitted sinc function, (c) ROC curve for skin detection on test set.

which show the most and least distribution directions (with respect to $Emax$ and $Emax$). From V and E the following features are extracted:

Most Distributed DIR: $MaxDD = V_{max}$.
Mean Distribution DIR: $MinDD = V_{min}$.
DIR of V with respect to X axis: $(DEV_{max}X, DE_{min}X)$.
DIR of V with respect to Y axis: $(DEV_{max}Y, DE_{min}Y)$.
Total direction features are store in vector FV_{dir}.

3.3 Shape Feature

To encode the data that relates to the shape of the skin area I_{bin} is divided into 256 equal size blocks. The size of the blocks are determined with respect to the size of I_{bin}. Feature vector FV_{shape} with size 1×256 holds the information of all the blocks. If the n'th block holds more than 50% of pixels that belong to the skin class, then $FV_{shape}(n) = 1$, otherwise $FV_{shape}(n) = 0$. This feature is utilized to describe position, and compactness of the skin region.

The final feature vector has 275 elements placed in $FV_{total} = [FV_{area}, FV_{direction}, FV_{shape}]$.

4. Experimental Results

To evaluate different steps of the proposed method, 1600 images were used and equally divided into two classes of normal and objectionable images. Each class was also divided into two equal numbers for train and test purposes. Since a standard database is not available for image filtering, some characteristics of our data set which can significantly affect the classification results are shown in Table 1. Presence of one or more person in an image makes a difference in regional-based skin detection. For example, when histogram matching is used, presence of more than one persons affects the resulted skin histogram, and this, in turn, could lead to wrong classification of skin. Many animals, due to their hair or skin color are often detected as skin, and they might be classified as objectionable. Many features encode shape, edge, and compactness of the skin region. If skin and regions with skin-like color are both present in an image, then the resulted feature vector might not correctly represent the true skin region. Sand has a very skin-like color and has a great impact on the filtering results. Further, all the skin images were chosen randomly and no special skin-color or illumination was considered in the data set, as it is common in the literature of skin detection (e.g., [9], [10], [12], [30]).

Table 1 Information about the database used in this paper.

	Frequency of occurrence			
	Porn		Non-porn	
Image type	**Test**	**Train**	**Test**	**Train**
Objectionable images	**400**	**400**	**0**	**0**
Normal images	**0**	**0**	**400**	**400**
One person in an image	**19**	**31**	**19**	**25**
Multiple people	**381**	**369**	**13**	**11**
True and False skin present	**289**	**223**	**13**	**22**
Human present in image	**400**	**400**	**31**	**37**
Animals present in image	**0**	**0**	**51**	**42**
Nature scenes	**0**	**0**	**209**	**151**
Beach scenes	**100**	**112**	**16**	**11**
Other	**0**	**0**	**48**	**101**

Fig. 8 Two vectors showing the direction of the skin pixel distribution.

4.1 Result of Pixel-Based Skin Detection

Using all the images in the training dataset, the mixed distribution of all skin color pixels was obtained in the Cb-Cr chrominance plane. Figure 8-a shows the histogram of skin color pixels in Cb-Cr plane. The equivalent fitted Sinc function is also shown in Fig. 8-b. Based on the results of the training set, the threshold for classifying a pixel as skin was chosen at $P_\theta = .0112$. To achieve this threshold, the probability value, for which TPR=95% was chosen. Based on this threshold, skin regions in the test set were extracted. Figure 8-c shows the Receiver Operating Characteristics (ROC) curve for the training images. When $P_\theta = .0112$ was used, TPR=93.2% and FPR=26.2% were obtained on the test set (accuracy of 83.5%). To assess the accuracy of the proposed method, our result was compared with those of the Gaussian distribution and the elliptic model (Fig. 8-c). The ROC curves, clearly show the superiority of the proposed method.

4.2 Result of Regional-Based Skin Detection

By employing the regional-based skin detection step, it is expected that a large amount of false detected skin regions

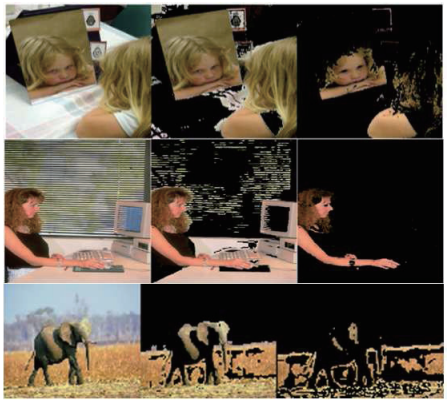

Fig. 9 (a) original image, (b) result of pixel-based and (c) regional methods for skin detection.

Table 2 Result of regional-based skin detection results.

	TPR%	FPR%	Acc.%
Proposed method	95.02	7.80	93.61
[7]	91.00	28.30	81.35
[30]	92.70	10.06	91.32

would be eliminated.

Figure 9 shows the result of some correct and incorrect examples of skin detection. In the first two images, the pixel-based method was able to detect all of the skin pixels, however many non-skin pixels have also been detected as skin. As it can be seen, by employing the regional-based step, the non-skin pixels have been excluded from the skin mask and the correct skin region is obtained. In the third image, the pixel-based method and the regional method have failed to detect the skin pixels correctly, because the color pixels in this image show the same regional properties of human skin. The result of skin detection based on TPR and FPR for the training and the test dataset is shown in Table 2.

To compare the result of the proposed method with regional-based skin detection methods, histogram matching [7], and adaptive skin detection, based on hierarchical color clustering [3] methods were tested on the data set. Histogram matching was performed in the RGB color space, and the skin histogram was obtained from the 500 adult test images. The results shown in Table 2 indicate 2.5% improvement of the presented method compared to [30], and more than 11% compared to [7].

Figure 10 shows the result of skin detection using methods of [7] and [30] for images of Fig. 9. As it can be seen, the proposed method shows much better results in comparison to [7]. The presented method was also more effective than that of [30] for eliminating skin-like pixels from true skin pixels.

After completing the skin detection stage, it is possible to let many normal images with no skin or very small TSR pass the filter. To do so, a TSR is set after each step of skin detection and images with small TSR are rejected in later processing steps. The ROC curves for employing TSR and

Fig. 10 Skin detection using: a) method of [7], and b) method of [30].

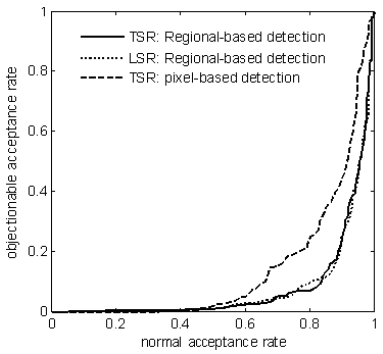

Fig. 11 Rate of filtering for TSR and LSR.

Table 3 TSR and LSR thresholds for initial image filtering.

	pixel-based (%)		regional-based (%)		pixel-based (%)	
TPR	**TSR**	**TNR**	**TSR**	**TNR**	**LSR**	**TNR**
99%	26	48.50	8	50.25	4	50.25
98%	30	51.25	12	57.00	5.5	54.20
95%	37	60.50	20	70.75	15	73.25

LSR after each step of skin detection are shown in Fig. 11.

Table 3 shows the True Negative detection Rate (TNR) for fixed TPRs. For each TRP, the corresponding threshold (TSR or LSR) was also calculated from Fig. 10 and shown in the table. It can be seen that for only 1% percent error on the objectionable image set, it is possible to let more than 45% of normal images pass the filter. The, LSR feature is not much more effective than TSR when used as a single threshold. Also, TNR is nearly the same for both pixel-based and regional-based skin detection methods (for TPR=90%); therefore, a threshold of TSR=26% was chosen after the pixel-based skin detection step. By using this threshold it is possible to avoid regional skin detection for about 25% of images in the data set; which in turn will reduce the overall computation time.

4.3 Classification of Images Using MLP

After applying TSR, F_{Vtotal} was extracted for the remaining images. Three types of features were explained for filtering: area, direction, and shape. These features describe the skin mask of each image. MLP is a monolithic strong classifier and has been shown to be a very efficient classification tool [31]. In this paper, an MLP with one hidden layer

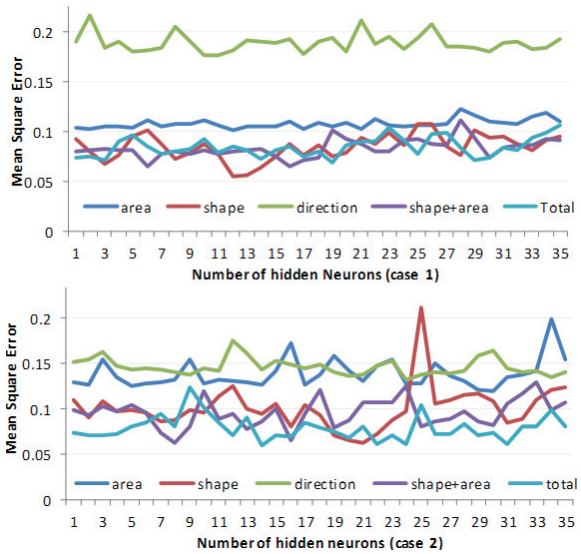

Fig. 12 MSE in the training process for 10 MLPs in case 1 and case 2.

and one neuron as output was used for classification. In the learning process, 80% of the training images were used for training and the rest were used for validating the weights. The number of hidden neurons was changed from 1 to 35, in order to find the better number of hidden neurons. Area, shape, and direction features, along with their combinations were tested to find out which of them are more effective. The effect of discarding images with small skin area, as explained in the previous subsection was also investigated.

Figure 12 shows the Mean Square Error (MSE) after training the MLP with different hidden layers changing from 1 to 35. Ten MLPs were trained with different inputs as follow:

1. Shape features independently (FV_{shape}).
2. Direction features independently (FV_{dir}).
3. Area features independently (FV_{area}).
4. Area and Shape features ($FV_{shape+area}$).
5. Area, Shape, and Direction features (FV_{total}).

where for five MLPs the FV_{total} was extracted without applying TSR after skin detection (case1), and for the other five, FV_{total} was extracted when TSR was applied after pixel-based skin detection (case 2).

As it can be seen in Fig. 12, in case 1, direction feature caused a high MSE compared to other combination of features. This is due to high skin FPR, as the original TSR was not applied after skin detection. In case 2, where TSR was applied, the direction feature for the remaining images show to be more effective. However, the MSE produced by FV_{dir} is still not acceptable. Shape feature in case 1 and 2 has a ver low MSE for different number of neurons. This shows that the proposed feature, which is also very easy and fast to compute, was very effective for classifying objectionable images. Also, when hidden neurons were set between 10 and 15 lower MSEs were obtained; therefore, MLP with 12 hidden neurons was chosen for classification on the test set.

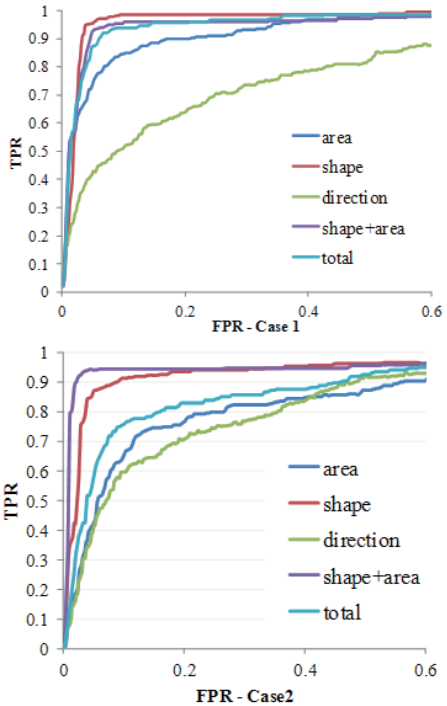

Fig. 13 ROC curves for training with the10 MLPs.

Table 4 Result of final classification on the test set.

	Case 1		Case 2	
	shape	**area+shape**	**shape**	**area+shape**
TPR(%)	92.75	92.25	90.25	94.75
FPR(%)	9.50	7.00	6.00	6.50
Accuracy(%)	91.60	92.60	92.12	94.12

ROC curves obtained from image classification on the test set are shown in Fig. 13. In case 1, where shape feature was used independently, better results were obtained, as expected. Whereas, in case 2, shape and area features showed better result. Also, Fig. 13 shows that when FV_{total} was used, classification rate was decreased due to the presence of ineffective FV_{dir}.

It should be mentioned that in case 2, for the non-porn data set, only normal images that had high TSR were input to the MLP (207 out of 400 normal images in the test set). Thus, it might seem Case 1 has better results, but if the number of rejected images after using TSR were added to the results of case 2, the ROC curves would have shown better results. Based on overall inspection of the trained MLPs, four MLPs were used for classification of images in the test set: the MLPs trained in case1 and 2 using FV_{shape} and $FV_{shape+area}$. Table 4 shows the final classification result on the test set.

Table 4 clearly shows that in case 2 when $FV_{shape+area}$ was used, the best result was obtained. In this case, the obtained accuracy of classification was 94.12%.

In the literature of pornographic content detection, the works of [9] and [10] have specifically focused on visual features and detection of pornographic images. Both of

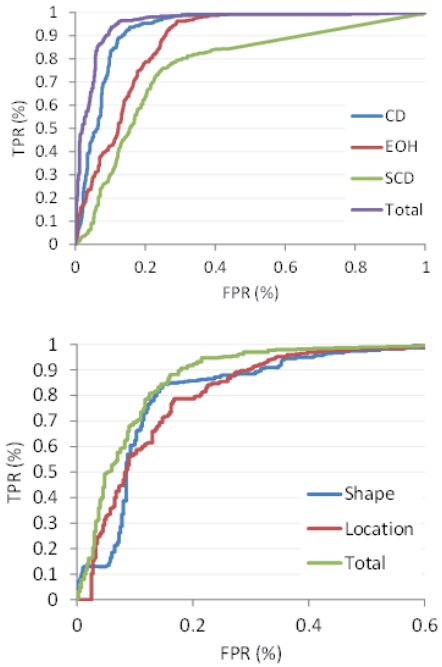

Fig. 14 Classification results using features presented in: (a) [10] and (b) [9].

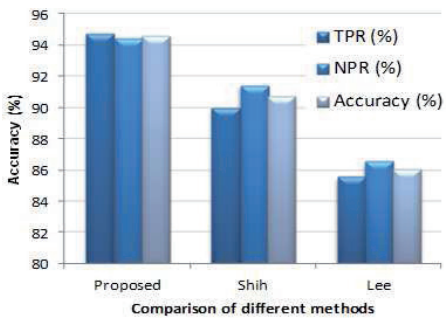

Fig. 15 Accuracy of different test cases.

these methods use human skin as the region of interest and extract features to find out if the image contains inappropriate content. Since these methods are more analogous to the method presented in this paper, they were chosen for comparison of the results. Features used in these two methods are: Scalable Color Descriptors (SCD), Edge Oriented Histograms (EOH), Compactness Descriptors (CD), Location (LOC) and Shape feature. The number of histogram bins, whenever required, was set as stated in [9] and [10]. Figure 12-a and 12-b show the results of classification for each set of features of [9] and [10] separately.

Figure 14-a shows that the SCD feature was not very effective for adult image classification alone, because all the regions that were left for feature extraction had the same color properties. The CD that encodes the shape of the detected skin region showed more effective compared to other features of this figure. The result of classification using features of [9] was not significant, because only binary information were used for adult image classification (Fig. 14-b).

Figure 15 shows the final comparison of the results, tested in this paper. On the bases of the classification accuracy, the result of the proposed method was improved by 4.4% and 9% compared to Shih et, al, [10] and [9] respectively.

5. Conclusions

In this paper, a new combinational method incorporating skin color pixel information and regional information was proposed for objectionable image filtering. For pixel-based skin detection, a 2D Sinc function was fitted to skin histogram, and by applying this model, skin pixels were detected with TPR=93.2% and FPR=26.2%. In the regional skin detection step, based on the distance between the exponential and the histogram of skin regions, the final skin mask was extracted with an accuracy of 93.61%. This step improved the accuracy of pixel-based skin detection by more than 10%. In comparison with the other regional methods, the proposed method was improve by more than 10% compared to the histogram matching method, and 2.31% compared to the adaptive skin detection method. In the classification phase, new shape and direction features along with area feature were extracted from each skin mask. By combining different features, ten MLPs were trained and based on MSE comparisons, it was found that the shape+area feature was more significant than the other features. The final results for objectionable image filtering showed that the accuracy of the proposed method (94.12% with only 6.5% FPR) was higher than that of the methods mentioned in this paper. The effectiveness of the proposed skin detection method and high detection accuracy makes the method significant for the application of objectionable image filtering.

References

[1] F. D'Orlando, "The demand for pornography," J. Happiness Studies, vol.12, no.1, pp.51–75, 2011.
[2] N. Kranich, "Why filters won't protect children or adults," Library Administartion and Management, vol.18, no.1, pp.8–14, 2004.
[3] W.P. Stol, H. Kaspersen, J. Kerstens, E.R. Leukfeldt, and A.R. Lodder, "Governmental filtering of websites: The Dutch case," Computer Law and Security Review, vol.25, no.3, pp.251–262, 2009.
[4] B. Haselton, Report on accuracy rate of FortiGuard Filter, pp.1–6, Bellevue, 2007.
[5] A. Khodaei, C. Shahabi, and C. Li, "Hybrid indexing and seamless ranking of spatial and textual features of web documents," Database and Expert Systems Applications, vol.6261, pp.450–466, 2011.
[6] W.A. Arentz and B. Olstad, "Classifying offensive sites based on image content," Computer Vision and Image Understanding, vol.94, no.1-3, pp.295–310, 2004.
[7] J.Z. Wang, J. Li, G. Wiederhold, and O. Firschein, "System for screening objectionable images," Computer Communications Journal, vol.21, no.15, pp.1355–1360, 1998.
[8] W. Zeng, W. Gao, T. Zhang, W. Gao, and Y. Liu, "Image guarder: An intelligent detector for adult images," Asian Conference on Computer Vision, pp.1080–1084, 2004.
[9] J.S. Lee, Y.M. Kuo, P.C. Chung, and E.L. Chen, "Naked image detection based on adapitve and extensible skin color model," Pattern

Recognit., vol.40, no.8, pp.2261–2270, 2006.
[10] J.L. Shih, C.H. Lee, and C.S. Yang, "An adult image identification system employing image retrieval technique," Pattern Recognit. Lett., vol.28, no.16, pp.2367–2374, 2007.
[11] C.S. Won, D.K. Park, and S.J. Park, "Efficient use of MPEG-7 edge histogram descriptor," ETRI Journal, vol.24, no.1, pp.23–30, 2002.
[12] A. Ahmadi, M. Fotouhi, and M. Khaleghi, "Intelligent classification of web pages using contextual and visual features," Applied Soft Computing, vol.11, no.2, pp.1638–1647, 2010.
[13] M. Hammami, Y. Chahir, and L. Chen, "WebGuard: A web filtering engine combining textual, structural, and visual content-based analysis," IEEE Trans. Knowl. Data Eng., vol.18, no.2, pp.272–284, 2006.
[14] W. Hu, O. Wu, Z. Chen, Z. Fu, and S. Maybank, Recognition of pornographic web pages by classifying texts and images, IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.6, pp.1019–1034, 2007.
[15] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," Pattern Recognit., vol.40, no.3, pp.1106–1122, 2007.
[16] A. Soetedjo and K. Yamada, "Skin color segmentation using coarse-to-fine region on normalized RGB chromaticity diagram for face detection," IEICE Trans. Inf. & Syst., vol.E91-D, no.10, pp.2493–2502, Oct. 2008.
[17] M. Decker and M. Sawaki, "Lighting independent skin tone detection using neural networks," IEICE Trans. Inf. & Syst., vol.E90-D, no.8, pp.1195–1198, Aug. 2007.
[18] A. Guerrero-Curieses, J. Rojo-Alvarez, P. Conde-Pardo, I. Landesa-Vazquez, J. Ramos-Lopez, and J.L. Alba-castro, "On the performance of kernel methods for skin color segmentation," Eurasip Journal on Advances in Signal Processing, vol.2009, no.30, pp.1–13, 2009.
[19] M.J. Jones and J.M. Rehg, "Statistical color models with application to skin detection," Int. J. Comput. Vis., vol.46, no.1, pp.81–96, 2002.
[20] A. Nadian and A. Talebpour, "Pixel-based skin detection using sinc function," IEEE Symposium on Computers and Informatics, pp.751–755, Kuala Lumpur, 2011.
[21] A. Kaushal and J. Raina, "Face detection using neural network and gabor wavelet transform," Int. J. Comput. Sci. Technol., vol.1, no.1, pp.58–63, 2010.
[22] M. Fotouhi, M.H. Rohban, and S. Kasaei, "Skin detection using contourlet-based texture analysis," Computer Society of Iran Computer Conference, pp.59–64, Tehran, 2009.
[23] D. Wang, J. Ren, J. Jiang, and S.S. Ipson, "Skin detection from different color spaces for model-based face detection," Communications in Computer and Information Science, vol.15, no.14, pp.487–494, 2008.
[24] M. Storring, H.J. Andersen, and E. Granum, "Physics-based modelling of human skin colour under mixed illuminants," Robotics and Autonomous Systems, vol.35, no.3-4, pp.131–142, 2001.
[25] A. Nadian, A. Talebpour, and M. Basseri, "Regional skin detection based on eliminating skin-like lambertian surfaces," IEEE Symposium on Computers and Informatics, pp.745–759, Kuala lumpur, Malaysia, 2011.
[26] R. Fletcher, Practical Methods of Optimization: Vol.2: Constrained Optimization, John Wiley & Sons, New York, 1981.
[27] S.A. Shafer, "Using color to separate reflection components," Color Research and Application, vol.10, no.4, pp.210–218, 1985.
[28] G.J. Klinker, S.A. Shafer, and T. Kanade, "A physical approach to color image understanding," Int. J. Comput. Vis., vol.4, no.1, pp.7–38, 1990.
[29] S. Chatterjee and A.S. Hadi, Regression analysis by example, 4th ed., Wiley-Intersection, New Jersey, 2006.
[30] H.-M. Sun, "Skindetection for singleimages using dynamicskincolormodeling," Pattern Recognit., vol.43, no.4, pp.1413–1420, 2010.
[31] B.D. Ripley, Pattern recognition and neural networks, Cambridge Univ. Press., New York, 2008.

Ali Nadian Ghomsheh received the B.S. degree in Electrical Engineering from Shahid Chamran University, Iran, in 2005, and the M.S. degree in Electrical Engineering from Shahid Beheshti University, Iran, in 2007. He is currently a Ph.D. candidate in the Department of Electrical and Computer Engineering, Shahid Beheshti University, Iran. His research interests include image processing and pattern recognition.

Alireza Talebpour worked for several years in the private sector after his first degree in Electrical Engineering in Iran. He obtained his Ph.D. in image processing from Surrey University in 2004, and since then he has been a lecturer at Shahid Beheshti University in Iran. His research interests are in multimedia and signal and image processing.