Journal of Imaging Science and Technology® 55(2): 020508-1–020508-10, 2011.

An Adult Image Identification System Based on Robust Skin Segmentation

Seok-Woo Jang
Department of Digital Media, Anyang University, 708–113 Anyang 5–Dong, Manan-gu Anyang-shi, Kyonggi-do 430–714, South Korea

Young-Jae Park and Gye-Young Kim
School of Computing, Soongsil University, 1-1, Sangdo 5–Dong, Dongjak-Ku, Seoul 156-743, South Korea
E-mail: gykim11@ssu.ac.kr

Hyung-Il Choi
School of Global Media, Soongsil University, Seoul 156–743, South Korea

Min-Chel Hong
School of Electronic Engineering, Soongsil University, Seoul 156–743, South Korea

Abstract. *In this article, the authors introduce a new algorithm to identify adult images that can effectively filter out images of naked human bodies in the internet. The algorithm detects eyes, which are known as the most salient component of a human face, and makes a statistical skin color distribution model directly from each input image by choosing reliable skin samples in facial areas near the detected eyes. Skin areas over the entire image are segmented robustly with the online constructed skin color model. The authors then extract a set of representative features characterizing naked bodies from the segmented skin areas and verify if the skin regions contain naked bodies through multilayer perceptron neural networked-based learning and inference of the representative features. Experimental results are given to demonstrate that the proposed adult image detection method can identify various types of nude images effectively compared to other conventional methods.* © *2011 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2011.55.2.020508]

INTRODUCTION

With the rapid development of the internet and information technology, it has become very easy to access and browse various types of multimedia contents on the web, including photographs, movie clips, and music files. Meanwhile, objectionable contents, such as pornographic images that would be illegal to sell even in adult bookstores, can be easily transferred to homes and schools through the web or via e-mail image attachments. This can cause juveniles to see such obscene images intentionally or unintentionally with little effort. Therefore, the methods of effectively blocking or filtering out this type of harmful materials have been the subject of great interest in related research areas.

In general, conventional adult image filtering approaches may be classified into three categories: Internet protocol (IP)-based blacklist blocking, textual content-based filtering, and visual content-based filtering.[1] The IP-based blacklist blocking approach first builds a set of uniform resource locators (URLs) of objectionable websites and then prohibits access to some requested web page if its URL is included in the blacklist. However, since the contents on the internet are highly dynamic and it is thus hard to keep the blacklist of all objectionable websites up to date, the IP-based blocking approach seems to be inefficient and impractical. The textual content-based filtering approach attempts to block obscene websites based on the analysis of their textual contents. Each word or phrase in a requested web page is compared with those in the keyword dictionary containing prohibited keywords or phrases, and the access to a requested web page is then prohibited if enough offensive keywords or phrases occur in the web page. The textual content-based filtering approach, however, suffers from the well-known *over-blocking* phenomenon that blocks access to educational websites related to health or sexology. In addition, many adult websites with text incorporated in elaborate images cannot be blocked by textual content analysis. Therefore, many researchers investigate the visual content-based filtering approach to analyze the image contents in a web page or e-mail image attachments.

In the visual content-based approach, the identification of adult images is typically treated as an image classification problem. Usually, human skin pixels are first extracted using some predefined or learned form of skin color model (SCM), and the detected skin pixels are then grouped into candidate skin areas. Subsequently, various representative features are obtained from the detected skin regions, including color, texture, and shape features, and they are used to discriminate benign images from adult images.

We can find many approaches to adult image identification based on analyzing the image contents in related lit-

Received Apr. 5, 2010; accepted for publication Nov. 2, 2010; published online Mar. 10, 2011.
1062-3701/2011/55(2)/020508/10/$20.00.

erature. Forsyth and Fleck suggested a system that finds naked people in an image by using a skin filter and a human figure grouper.[2] Color and texture features are exploited to detect human skin regions, and the detected regions are then fed to the specialized grouper for grouping a human figure using geometric constraints on human structures. If the grouper finds a predefined structure, the system decides that an image of a naked human body is present. Wang et al. proposed a wavelet image pornographic elimination system for screening objectionable images.[3] The system successively eliminates pornographic images by using a combination of an icon filter, a graph photodetector, a color histogram filter, a texture filter, and a wavelet-based shape matching algorithm. Jones and Rehg introduced a skin color detection technique with a statistical color distribution model.[4] Several features are extracted, and a neural network classifier is trained using both adult and nonadult images. Bosson et al. presented a new method to block pornographic images where the likelihood ratio of a quantized color space is computed and the features of the image blobs are then computed and presented as a vector.[5] Finally, a neural network may be utilized to classify whether the image is pornographic or not. Jedynak et al. proposed a statistical model for skin detection.[6] The maximum entropy model is adopted to infer the skin models from the data set. Then, the Bethe tree approximation and belief propagation algorithm are utilized to approximate the skin probability at pixel locations. Zheng et al. suggested an adult image detection method, where an SCM is first applied to extract skin bocks in an image, and features extracted from skin blocks are then fed into the multilayer perceptron (MLP) classifier to identify adult images.[7] Hammami et al. developed the WebGuard system in which visual analysis based on skin color is employed for adult content detection and filtering.[8] A skin color-related visual feature, which represents the percentage of skin pixels within a web page, is used to identify pornographic websites. Lee et al. adopted the YCbCr color space and performed skin segmentation with several SCMs generated to tolerate the chromatic deviation due to special lighting.[9] The texture roughness feature was further utilized to reject false positives coming from skinlike background regions. Several representative features were then used to verify the detected skin areas. Ioffe and Forsyth suggested human body model-based pornographic image detection.[10] The system segments skin color pixels in an image using color and texture information and then finds all connected skin regions that are candidates for trunks and limbs. These skin columns are combined subject to constraints derived from a geometrical model of the human body. If the combination can form the shape of a human body, the image is treated as pornographic. Besides these methods, many other approaches have been reported.[11–13]

As can be seen in these methods, accurate detection of human skin regions is very important for adult image identification, but they still have essential problems in extracting skin color. In other words, the colors of human skins are basically not the same because of individual skin difference

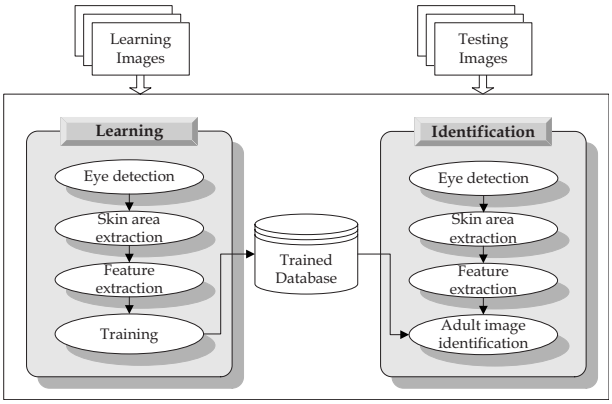

Figure 1. Overall flow of the proposed algorithm.

or different races. Moreover, the skin regions of captured images may not have identical color due to makeup, different cameras used, various illumination conditions, etc. Therefore, most of the existing skin region detection algorithms using predefined or learned SCMs are unable to overcome all of the above mentioned circumstances. The optimal solution for the problem is to reliably select skin samples from an input image and to adaptively make a skin chrominance distribution model suitable for the image itself whenever it is tested instead of utilizing a generalized skin model.

In this article, therefore, we propose a new adult image detection method that robustly segments skin areas with an input image-adapted skin color distribution model and verifies if the segmented skin regions contain naked bodies by fusing several representative features through a neural network scheme. The main difference of our method from previously reported skin detection methods is that we first detect eyes, which are known as the most salient component in a human face,[14,15] and make an image-adapted and statistical SCM adequate for the test image by choosing reliable skin samples in facial areas near the detected eyes. Skin areas over the entire image are then segmented with the generated SCM. Figure 1 shows the overall flow of the proposed adult image identification algorithm.

As shown in Fig. 1, the suggested method consists of two main parts: a learning part and an identification part. The learning part robustly segments skin areas and extracts a set of visual features from the segmented skin regions which characterize naked bodies. It then trains the features by using a MLP neural network. The identification part performs skin area detection and feature extraction from a test image similarly to the learning part and optimally distinguishes adult images from nonadult images while fusing the different types of extracted features with the MLP.

DETECTING REPRESENTATIVE FACIAL COMPONENTS

The purpose of this step is to accurately detect human eye areas in a color image, which will provide significant base positions of the search areas for selecting skin samples reliably. In general, among various facial features, eyes are the most prominent features that can be used for face detection

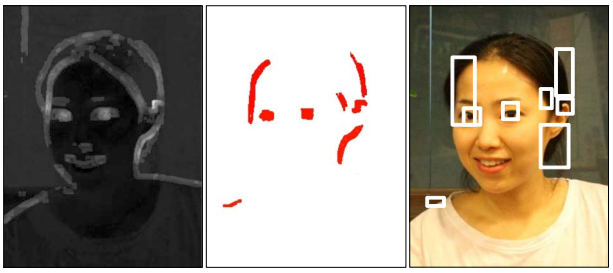

Figure 2. Example of detecting eyes: (a) eye map image; (b) segmented regions; and (c) eye candidates.

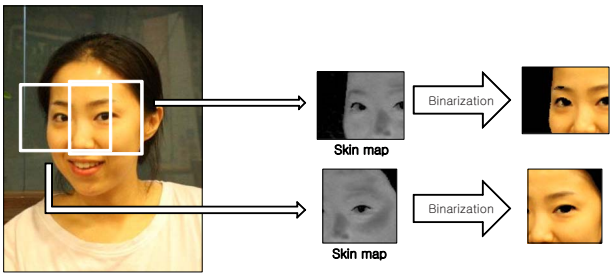

Figure 3. Constructing and binarizing a skin map.

and recognition.[16,17] In this article, we first transform red-green-blue (RGB) color space to YCbCr color space because detecting human skin color using RGB color values does not seem feasible and thus different color spaces are needed to reduce the variance of skin colors. Since the cluster of skin color pixels is more compact in YCbCr space than hue-saturation-value (HSV) space,[18] the red, green, and blue color components are nonlinearly transformed into YCbCr color space components.

We then use *EyeMap* to extract human eyes in a color image, as proposed by Hsu et al.[19] Two separate eye maps are first constructed by using the chrominance and luminance components, and these two maps are then combined into a single eye map. The eye map from the chroma is based on the observation that high C_b and low C_r values are found around the eyes and is defined as in Eq. (1),

$$EyeMapC = \frac{C_b^2 + (255 - C_r)^2 + (C_b/C_r)}{3}, \qquad (1)$$

where C_b^2, $(255-C_r)^2$ and C_b/C_r are all normalized to the range from 0 to 255. Since the eyes usually contain both dark and bright pixels in the luma component, gray-scale morphological operators can be designed to emphasize brighter and darker pixels around eye regions.[20,21] Thus, gray-scale dilation and erosion operations with a hemispheric structuring element are used to build the eye map from the luma as in Eq. (2),

$$EyeMapL = \frac{Y(x,y) \otimes g_\sigma(x,y)}{Y(x,y) \oplus g_\sigma(x,y) + 1}, \qquad (2)$$

where $\otimes$ and $\oplus$ represent dilation and erosion operations, respectively. The eye map from the chroma is enhanced by histogram equalization and then combined with the eye map from the luma by an *AND* (multiplication) operation as in Eq. (3),

$$EyeMap = (EyeMapC) \ \ AND \ \ (EyeMapL). \qquad (3)$$

The resulting eye map is further dilated, masked with the area that is built by enclosing the skin-tone regions with a peudoconvex hull, and normalized to brighten the eyes and suppress other facial areas. The locations of eye candidates are initially estimated from the pyramid decomposition of the eye map and then refined using iterative thresholding and binary morphological closing on this eye map. Figure 2 shows an example of detecting candidate eye regions by using the *AND*ed *EyeMap*. Fig. 2(a) represents the resulting image of applying *EyeMap* to a test image, Fig. 2(b) displays segmented and binarized candidate eye regions with *EyeMap* values greater than a predefined threshold value, and Fig. 2(c) shows the resulting image overlaid with minimum enclosing rectangles corresponding to the detected eye candidates.

EyeMapC seems to be more or less sensitive to the variance of eye colors due to different races and color contact lenses since it is defined with chrominance components. For *EyeMapL*, we can find that the overall result is better when the boundaries of eyelashes or pupils of the eyes are clear. In order to select real eyes among the detected candidate eye regions, we utilized both photometric and geometrical features such as color distribution of the candidate region, the elongatedness and compactness of the region, and the distance between two eye candidates.

In the present work, we detected human eyes by using the *EyeMap* in YCbCr color space, but any other eye detection methods can be used if their accuracy is acceptable. In addition, other approaches to obtaining representative facial components such as noses and lips apart from eyes may be applied similarly to define the search area for choosing human skin samples.

EXTRACTING HUMAN SKIN AREAS

The SCM is one of the most important factors in segmenting human skin regions from a color image. That is, unless the skin model is made to effectively reflect the distribution of human skin color, accurate extraction of skin regions is difficult to achieve. To robustly extract skin areas from color images which are taken under various conditions or some constraints are not imposed on, in this work an image-adapted SCM for each test image is constructed by using the color distribution of the skin regions near the detected eyes. Instead of making use of some predefined or learned form of skin color distribution model as in most of the conventional skin detection methods, we adaptively build our color model from each test image online while selecting reliable skin samples with the aid of an eye extracting facility, so that the resulting skin segmentation may be very successful.

Therefore, we obtain human skin samples within the fivefold enlarged areas of the corresponding maximum en-

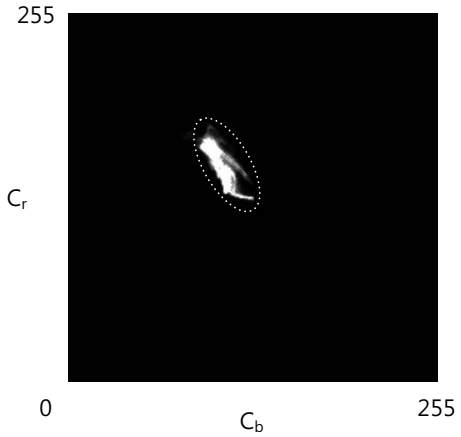

Figure 4. Distribution of skin samples in C_bC_r subspace.

closing rectangles for the detected eye regions. In general, though there is a high possibility of human skin pixels existing in the selected regions near the eyes, nonskin areas such as glasses or other background regions may also be present, therefore we need to classify the chosen regions into skin and nonskin pixel regions. For this purpose, we suggest building a skin map by calculating the distance between a selected skin sample and an average value of skin color as in Eq. (4),

$$SkinMap = 255 - \frac{|\hat{C}_r - C_r| + |\hat{C}_b - C_b|}{2}. \tag{4}$$

In Eq. (4), $\hat{C}_r$ and $\hat{C}_b$ represent the general C_r and C_b skin color values, respectively. *SkinMap* has values between 0 and 255 and is so constructed that it has a value close to 255 as the selected sample has color similar to the average skin color. Figure 3 illustrates an example of building a skin map from a test color image. As the color of the skin map is close to white, it means that the map is similar to human skin.

In order to choose only true skin sample pixels in the skin map, the histogram binarization suggested by Otsu is then performed.[22] Otsu's method statistically selects an optimal threshold that binarizes a gray-level histogram, without *a priori* knowledge, by using a discriminating criterion to maximize the separability of the resultant classes. This method is well known to show excellent performance when the histogram has two distributions of probability density. Fig. 3 also shows an example of displaying subimages of skin regions after binarizing the histogram of skin maps, where nonskin regions are shown in black and skin regions are in the corresponding original colors of the test image. When the color of hair is analogous to skin color owing to hair dyeing, the corresponding hair regions may not be eliminated completely. However, since the hair color is similar to skin color, the overall result of constructing a skin model is not affected very much, and such cases are very rare.

Modeling skin color requires choice of an appropriate color space and identifying a cluster associated with skin color in this space.[4] In general, the distribution of skin color samples in the C_bC_r subspace takes the form of an elliptical shape, as shown in Figure 4, but the positions of corresponding ellipses for test images are different. As a result, we build our image-adapted SCM online by extracting skin samples directly from each test image itself. Our SCM is generated by using the human skin samples selected near the detected eyes of the test image and is specified by the center and spread of the elliptical cluster. The elliptical model for the skin tones is described in Eqs. (5) and (6),

$$\frac{(x - ec_x)^2}{a^2} + \frac{(y - ec_y)^2}{b^2} \leq 1, \tag{5}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} C_b - C_b' \\ C_r - C_r' \end{bmatrix}, \tag{6}$$

where (C_r, C_b) denotes the C_r and C_b values of the test image and θ represents the angle of rotation of the ellipse. (a, b) means the length of the major and minor axes of the ellipse and (ec_x, ec_y) denotes compensation values for the rotation error of the ellipse. In this article, the values of θ, (a, b), and (ec_x, ec_y) are set to 2.53 (in radians), (25.39,14.03), and (1.0,2.0), respectively. These values are determined through repetitive experiments with various test images. The center of our skin model (C_b', C_r') is formalized in the C_bC_r subspace as in Eq. (7),

$$C_b' = \frac{1}{K}\sum_{t=0}^{K-1} C_b(t), \quad C_r' = \frac{1}{K}\sum_{t=0}^{K-1} C_r(t). \tag{7}$$

In Eq. (7), t denotes the index of representing each skin sample, $C_b(t)$ and $C_r(t)$ are the C_b and C_r chrominance values at the tth sample, respectively, and K represents the number of chosen skin samples. C_b' and C_r' denote the average values of C_r and C_b of the samples, and they are utilized as standard values for later skin segmentation.

After making the elliptical model of skin color distribution, we perform skin segmentation for the test image with the model. Once our SCM is generated, the skin region segmentation is rather simple. In other words, if the color of a pixel in the image is within the range of the elliptical model, we define the pixel as a skin color pixel. Otherwise, it is classified as a nonskin color pixel. After detecting all skin color pixels in the image, we can obtain individual skin regions by applying the labeling process to them. In this article, since the image-adapted skin model is constructed directly from each test image instead of using a predefined or learned SCM, robust segmentation of skin regions regardless of various conditions can be achieved.

FEATURES CHARACTERIZING NAKED BODIES

After skin areas are robustly obtained from an input image by using the skin color distribution model, a set of visual features that characterize adult images is extracted to judge if the skin regions contain naked bodies. In general, there may exist a lot of nonskin objects possessing skinlike chroma in an image, e.g., wood, desert sand, rocks, food, and the skin or fur of animals, and thus the selection of major features

representing adult images is very import for naked image identification. For this purpose, we investigated a mass of naked images and use the following five types of image features: location, size, elongatedness, compactness, and texture smoothness. For the sake of practicability, the features should be designed to be simple and effective.

For the location feature, we compute the normalized distance between the center of gravity of each skin region and the center of the image as in Eq. (8). The normalization is used because the size of the image is not always square. This feature is to measure if the segmented skin area is close to the image center, and it is based on the observation that the position of a naked body is usually near the center of an adult image in order to harmonize with the frame. Here,

$$F_{loc}^{i} = \frac{1}{\sqrt{2}} \times \sqrt{\left(\frac{|\mathrm{CG}_x^i - \mathrm{IC}_x|}{W}\right)^2 + \left(\frac{|\mathrm{CG}_y^i - \mathrm{IC}_y|}{H}\right)^2},$$

$$\mathrm{CG}_x^i = \frac{1}{N^i}\sum_{j=1}^{N^i} x_j^i, \quad \mathrm{CG}_y^i = \frac{1}{N^i}\sum_{j=1}^{N^i} y_j^i, \quad (x_j^i, y_j^i) \in R^i,$$

$$\mathrm{IC}_x = \tfrac{1}{2}W, \quad \mathrm{IC}_y = \tfrac{1}{2}H. \tag{8}$$

In Eq. (8), R^i denotes the ith segmented skin region, W and H represent the width and height of the input image, and IC_x and IC_y are the x and y coordinates of the image center, respectively. CG_x^i and CG_y^i are the coordinates of the gravity center of the region R^i and (x,y) denotes the x and y coordinates of the R^i. N^i represents the number of pixels in the R^i. The location feature F_{loc}^i is designed to have a value between 0 and 1. If the feature has a value close to 0, the corresponding skin region lies near the center of the image. On the other hand, if it has a value close to unity, the region is located far away from the image center.

For the size feature, we use the ratio of the number of pixels in the skin region to those of the whole image. For delighting viewers, the naked body usually occupies a significant portion of an image. Therefore, the size feature F_{size}^i should have a value greater than some predefined threshold value so that the segmented region can be judged to contain a naked body,

$$F_{size}^{i} = \frac{N^i}{W \times H}. \tag{9}$$

For the elongatedness feature, we calculate the aspect ratio of the segmented skin area. This feature is used to quantify how the profile of the segmented region is like the naked body. To implement the elongatedness feature, we apply principal component analysis to the segmented region and obtain two orthogonal eigenvectors. By projecting the region to the two orthogonal eigenvectors, we then get a minimum rectangle which barely encloses the region. Our elongatedness feature F_{elon}^i is defined with the ratio between the horizontal and vertical lengths of the rectangle. In Eq. (10), LER^i denotes the least enclosing rectangle of the R^i and

-1	-1	-1
-1	8	-1
-1	-1	-1

Figure 5. Laplacian mask for computing edgeness.

$L_{hor}(\mathrm{LER}^i)$ and $L_{ver}(\mathrm{LER}^i)$ represent the horizontal and vertical lengths of LER^i, respectively,

$$F_{elon}^{i} = \frac{L_{hor}(\mathrm{LER}^i)}{L_{ver}(\mathrm{LER}^i)}. \tag{10}$$

For the compactness feature, we compute the ratio of the area of the extracted skin region to the area of its LER. Compactness represents the denseness of a region. Usually, since a naked body region consists of a set of connected pixels that has similar skin color, the region is dense rather than sparse. This feature is very useful in differentiating real naked body areas from non-naked ones. The compactness feature F_{comp}^i is defined as in Eq. (11), and it has a value close to unity when a region is fully compact,

$$F_{comp}^{i} = \frac{N^i}{L_{hor}(\mathrm{LER}^i) \times L_{ver}(\mathrm{LER}^i)}. \tag{11}$$

For the texture smoothness feature, we characterize the edgeness of the segmented skin region. This feature is employed to discriminate between adult and nonadult images since most nonadult images may have sharper edges whereas the adult ones will exhibit smooth textures. In order to compute the smoothness feature, we first apply the Laplacian edge operator to the segmented area so that we can calculate the magnitude of its edgeness. The Laplacian operator is well known to be a two-dimensional isotropic measure of the second derivative of an image, and it is fast and can effectively detect edges in all directions.[23] We use the Laplacian convolution mask as illustrated in Figure 5.

We then count the number of pixels in the region whose edge magnitude value is similar to the mean edge value of the whole skin region. The smoothness feature F_{text}^i is defined as in Eq. (12), and it has a value close to 1 as the edgeness of the skin region is smooth. In Eq. (12), $E(x_j^i, y_j^i)$ denotes the edge magnitude of the pixel at the position (x,y) of the region R^i, $\hat{E}^i$ is the mean edge value of the R^i, and E_{th} represents a threshold edgeness value determined through experiments,

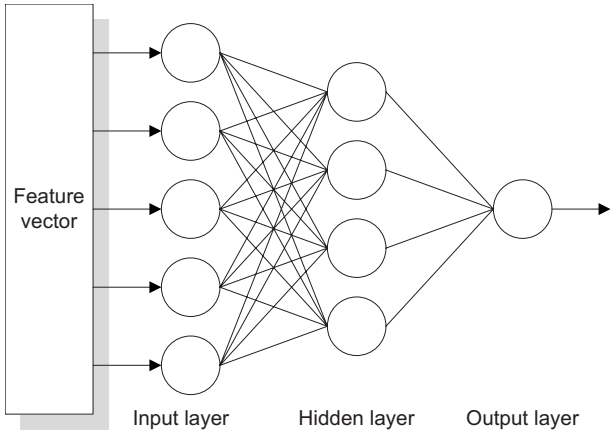

Figure 6. Structure of multilayer perceptron.

$$F_{text}^{i} = \frac{1}{N_i} \times \sum_{j=1}^{N^i} \Phi(E(x_j^i, y_j^i)),$$

$$\Phi(E(x_j^i, y_j^i)) = \begin{cases} 1 & \text{if } \hat{E}^i - E_{th} \le E(x_j^i, y_j^i) \le \hat{E}^i + E_{th} \\ 0 & \text{else,} \end{cases}$$

$$\hat{E}^i = \frac{E(x_j^i, y_j^i)}{N^i}. \tag{12}$$

The computed five types of normalized features characterizing each segmented skin region can be expressed in the form of a feature vector as in Eq. (13), which will be used for classifying images. Every feature in the vector is normalized to have a value from 0 to 1 for effective processing in later steps,

$$F(R^i) = [F_{loc}^i, F_{size}^i, F_{elon}^i, F_{comp}^i, F_{text}^i]. \tag{13}$$

IDENTIFICATION OF ADULT IMAGES

In order to optimally distinguish adult images from nonadult images while effectively fusing the five types of features produced in the feature extraction step, we employ the MLP classifier. The MLP is well known to be a feed-forward artificial neural network that has been extensively used in classification and regression. Evidence from the references shows that the MLP classifier offers a statistically significant performance improvement over other classification approaches such as the generalized linear model, the k-nearest neighbor classifier, the support vector machine, and so forth.[7,24] Our MLP is a three-layer feed-forward neural network with one hidden layer, as illustrated in Figure 6.

The five types of normalized features extracted from each skin region are put into the input layer, and a sigmoid function is used as the nonlinear activation function. The output of the MLP has a value between 0 and 1, which represents the degree of likeliness that the segmented skin region contains naked bodies. In other words, the closer the value is to 1, the more likely the input image corresponds to an adult image. We then set a threshold value T, $0 < T < 1$, to get a binary decision. In this work, the threshold T plays the role of adaptively controlling the level at which a test image is identified as an adult image. For instance, when children browse web pages, T can be increased such that most adult images will be filtered out. Otherwise, a proper T can be set to minimize the number of nonadult images which are misidentified as adult images. If there is more than one skin region in the input image, we use the maximum output value among them.

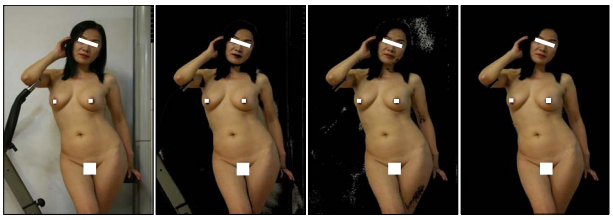

Figure 7. Extracting skin color regions in a naked body image: (a) input image; (b) method of Hsu *et al.*; (c) method of Lee *et al.*; and (d) proposed method.

Although we attempt to detect adult images with skin color distribution and a set of visual features, many false positives may still exist. For example, mug shots often include a large portion of skin area in the image, so that they tend to be frequently recognized as adult images. To exclude such false positives coming from mug shots, most existing adult image detection approaches employ some type of face detector. In general, the mug shot has an important characteristic in that its face area occupies a significant portion of skin area. As a result, it can easily be identified by utilizing the ratio of the face region to the whole skin region. Unfortunately, the face detection module increases the system's computational load. On the other hand, in this work we have already extracted eyes when generating the skin color distribution model and thus can effectively distinguish mug shots from adult images through the geometric relationships between the eyes and the face region.

EXPERIMENTAL RESULTS

The proposed adult image detection algorithm was implemented using VISUAL C++ 2005 and tested in WINDOWS 7 on an Intel Core2 Quad Q9400 2.66 GHz processor with 4 GB memory. In order to construct our image database, we collected a variety of adult and nonadult images which are distributed for commercial and noncommercial use on the web. Since the images were taken and digitized under various conditions, it can be said that no special illumination or other constraints are imposed on our test images. A total of 2400 images were collected to form our database and manually categorized into four groups, including 1200 naked body images, 400 bikini images, 400 portrait images, and 400 other miscellaneous images. The naked body images contain fully naked and seminaked (upper body only) people; the bikini images are non-naked images including people wearing revealing clothing such as swimming suits or bikinis; the portrait images involve single near-frontal faces or mug

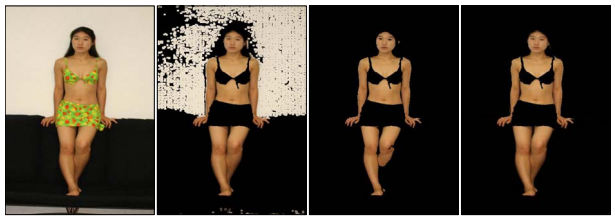

Figure 8. Extracting skin color regions in a bikini image: (a) input image; (b) method of Hsu *et al.*; (c) method of Lee *et al.*; and (d) proposed method.

Figure 9. Extracting skin color regions in a portrait image: (a) input image; (b) method of Hsu *et al.*; (c) method of Lee *et al.*; and (d) proposed method.

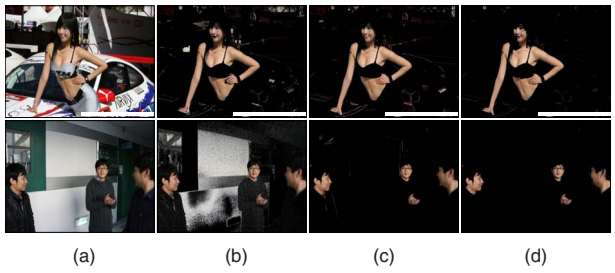

Figure 10. Extracting skin regions in a image with complex backgrounds: (a) input image; (b) method of Hsu *et al.*; (c) method of Lee *et al.*; and (d) proposed method.

shots but differ in terms of backgrounds and illuminations; and other miscellaneous images are the ones with complex backgrounds and multiple human faces, respectively. Among 2400 database images, 1200 images including half of each group were used for training and the remaining 1200 images for testing.

In this article, we evaluate the performance of our suggested method in terms of two aspects: skin color segmentation and adult image identification. To compare the performance of the skin segmentation aspect, we also implemented two existing algorithms proposed by Hsu et al.[19] and Lee et al.[9] which seem to be the most representative skin detection methods, and the same test images were applied to the two conventional methods. Some experimental results of the suggested and existing methods are shown in Figures 7–10, where the resulting image of skin detection is a semantically binary image consisting of skin color pixels and nonskin black color pixels.

Fig. 7 shows some results of extracting skin areas in a naked body image. As can be seen in Fig. 7, the existing methods seem to have difficulty accurately obtaining the boundaries of skin color areas and tend to detect many unnecessary regions because the background includes colors similar to skin color. On the other hand, the proposed method is very successful.

The method of Hsu et al. is mainly for the detection of human faces, not for naked bodies that normally occupy a large portion of the image. Furthermore, there exist a large number of naked body pictures taken under special illumination conditions. Usually, warm lighting is applied to make skin tone look more attractive, while human skin color deviates from the normal case at the same time. Skin extraction method of Hsu et al. seems to be inadequate to cope well with these environments. The method proposed by Lee et al. may detect skin areas successfully with the aid of different patterns of learned skin color, but it tends to extract redundant nonskin background regions as human skin. That is, their method often fails to distinguish skinlike background areas from real skin areas owing to somewhat detailed skin clusters. Meanwhile, the suggested method shows promising results because it does not depend on some predefined SCMs or learning scheme and constructs skin color distribution adaptively from each test image itself.

Fig. 8 illustrates some results of obtaining skin color areas from a bikini image. We can notice that the overall results of skin detection for naked and bikini images are normally similar since naked human body regions occupy a large portion of both types of images. Fig. 9 demonstrates the results for a portrait image, where the image is a nonnaked image that contains skin regions corresponding to the parts of a human body such as faces, hands, and legs. In portrait images, the skin areas usually do not occupy a large portion of the image compared to the naked body and bikini images. Also, the portrait images are mostly captured in a somewhat good indoor or outdoor environment so that they do not involve serious distortions, noise, or reflections. Therefore, both the existing and proposed skin extraction methods produce satisfactory results. In Fig. 10, the upper and lower images show some results for an image with complex backgrounds of varying illuminations and different patterns of textures and for an image with multiple faces and hands, respectively. For both cases, we can clearly see that our approach outperforms others.

To evaluate the performance of skin detection quantitatively, we define the root mean square error (RMSE) measure as in Eq. (14), which deals with both general and specific aspects of image quality.[25,26] RMSE is a frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed from the thing being modeled or estimated, and it is a good measure of accuracy,

$$\mathrm{RMSE} = \sqrt{\frac{1}{M \times N} \sum_{i=m}^{M} \sum_{j=n}^{N} |\mathrm{OB}(i,j) - \mathrm{RB}(i,j)|^2}, \quad (14)$$

where M and N represent the width and height of the image and (i,j) is the horizontal and vertical indices representing

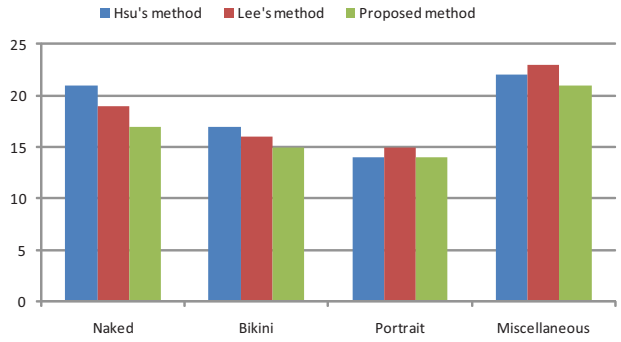

Figure 11. Root mean square error.

Table I. Correct classification rates with different features.

Features	Naked body	Bikini	Portrait	Miscellaneous
F_{loca}	33.64	36.22	38.34	37.76
$F_{loca}+F_{size}$	89.47	42.38	56.25	98.67
$F_{loca}+F_{size}+F_{elon}$	91.67	79.97	62.50	99.91
$F_{loca}+F_{size}+F_{elon}+F_{comp}$	91.68	82.21	93.75	87.65
$F_{loca}+F_{size}+F_{elon}+F_{comp}+F_{text}$	94.33	84.39	93.76	85.16

Table II. Confusion matrix of the suggested method.

	Test images	
Classification results	Naked body images	Non-naked images
Naked images	566 (94.33%)	44 (7.32%)
Non-naked images	34 (5.56%)	556 (92.68%)

positions of the image. OB(i,j) denotes the ground truth binary image and RB(i,j) is a resulting binary image of skin detection. For the accurate evaluation of performance, we manually converted the original test images into binary images consisting of skin color and nonskin color pixels.

The computed root mean square errors for the existing and proposed skin extraction methods are illustrated in Figure 11. We can clearly notice that our approach shows better results compared to other approaches. From these experiments, our skin detection method proves to be able to robustly obtain human skin regions by building image-adapted SCMs from the test image itself.

To verify the performance of our adult image identification method, true and false detection rates are measured. The true detection rate is the percentage of adult images correctly classified and the false detection rate denotes the percentage of nonadult images incorrectly classified. In this article, we first measure the correct classification rate of our method while increasing the features used to characterize naked images. We can easily find from Table I that as more features are considered, better classification results are generally achieved. For example, the compactness feature shows

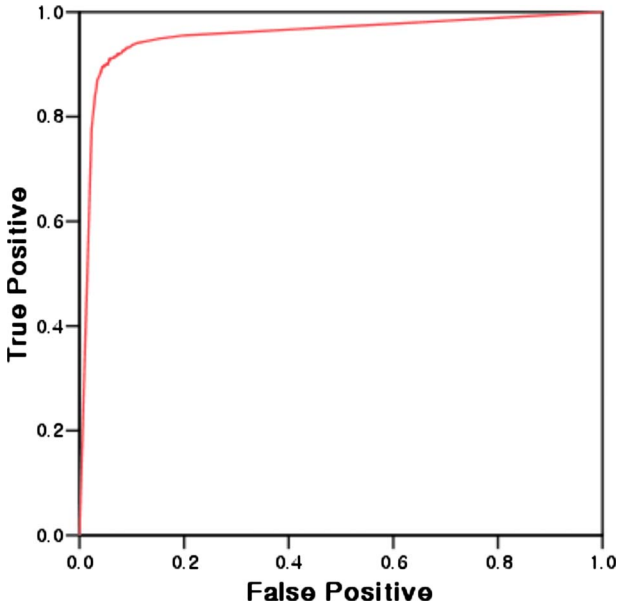

Figure 12. ROC curve of the proposed system.

Table III. Confusion matrix in the method of Lee *et al.*

	Test images	
Classification results	Naked body images	Non-naked images
Naked images	537 (89.5%)	53 (8.83%)
Non-naked images	63 (10.5%)	547 (91.17%)

Table IV. Confusion matrix in the method of Yang *et al.*

	Test images	
Classification results	Naked body images	Non-naked images
Naked images	512 (85.33%)	79 (13.17%)
Non-naked images	88 (14.67%)	521 (86.83%)

better classification results for the portrait images since they are mostly captured in somewhat good indoor or outdoor environments, and their skin areas do not occupy a big portion of the image. On the other hand, the texture smoothness feature shows better classification results for the naked body and bikini images because most nonadult images may have sharper edges whereas the adult images will exhibit smooth textures. However, for the miscellaneous images, we get classification results that do not seem to be dependent on the number of features used. This result is probably because our method uses several features that effectively represent naked body images, but the miscellaneous images contain various types of non-naked bodies or scenes.

Table II shows the confusion matrix of our experimental results. The identification rates of the naked body and non-naked images are 94.33% and 92.68%, respectively.

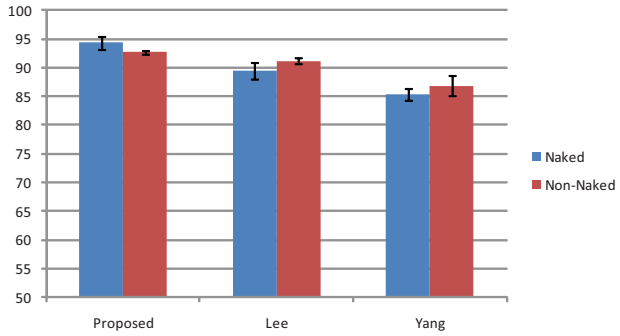

Figure 13. Error bars for the correct proportions.

The receiver operating characteristic (ROC) curve of the proposed adult image detection method is illustrated in Figure 12. As can be seen in Fig. 12, the suggested system meets the requirements for the practical application of an adult image identification system, that is, the reasonable detection rate and low false alarm. From these experiments, our method proves to be able to effectively distinguish naked body images from non-naked ones by adopting the input image-adapted SCM as well as integrating a set of representative skin area features through the multilayer perceptron neural network scheme.

To compare the performance of our adult image detection algorithm with those of other approaches, we also implemented the method of Lee et al. and the method of Yang et al.[9,11] This choice was made based on the fact that these two existing methods were introduced recently, and their results are known to be relatively good as well. The same test images are employed to derive the performance comparison. The confusion matrices of the two experimental results are shown in Tables III and IV, respectively. These results show that our method can detect adult images more accurately than the other approaches.

Figure 13 shows the error bars for the correct proportions of the three confusion matrices. As can be seen from Fig. 13, for the case of recognizing adult images correctly, our method has a similar standard deviation to the method of Yang et al., and the method of Lee et al. shows a larger one. For the case of recognizing nonadult images correctly, the suggested method shows the smallest variation and the method of Yang et al. has the largest one. With the error bars, it is clear that the suggested method is better than prior works.

The suggested adult image detection system usually works well when it robustly detects eyes by exploiting eye maps and when the skin region extraction using the color near the detected eyes is effectively performed. Also, the adult image identification performance is good when the features of extracted skin areas are similar to the five features we defined. On the other hand, our system often fails when nonskin objects possessing skinlike chroma, such as wood and desert sand, could be detected as skin and the detected regions may have the characteristics analogous to those of skin regions. In addition, bikini images containing a large portion of skin areas tend to be falsely detected as adult images.

In our system, about 20% of eyes are incorrectly detected. This happens when there are people with their eyes closed, eyes are partially or completely covered by hair, or the image contains some captions since *EyeMap* has difficulty distinguishing captions from eyes. If the eyes are incorrectly detected, the statistical distribution of our image-adapted SCM may be far from that of a generally used SCM. In this case, we must use the general SCM instead of our skin model.

Adult image identification in more complex images has lower accuracy than those of other images since eyes may be incorrectly detected and there is a high possibility of falsely detected skin regions existing in backgrounds. However, complex images do not always make the detection accuracy low. If skin color regions extracted from complex images are cluttered with small areas, the overall performance is the same as for normal images. Also, although there may be skin-tone areas in backgrounds, they do not make an important impact on the identification results if their features such as compactness and texture smoothness are not similar to those of real skin color regions.

Our system may not distinguish between pornographic images and fine art images although it depends on the types of fine art images used. For instance, Botticelli's "Birth of Venus" includes large areas of human skin and is thus classified as an adult image. Our method is also racially robust, properly handling various skin colors because it creates and uses an image-adapted and statistical SCM appropriate to the test image, choosing reliable skin samples in facial areas near the detected eyes.

CONCLUSIONS

With the rapid development of the internet, it has become very easy to access and browse various multimedia contents on the web. Meanwhile, objectionable or illegal content such as pornographic images is widely available, causing severe social problems. Therefore, in order to filter out this type of harmful materials, different types of approaches have been explored based on visual content-based filtering; however, they still have essential problems in segmenting skin color due to a variety of reasons. In other words, the color of human skin varies because of individual skin differences or different races. Moreover, the skin regions of captured images may not have identical color due to makeup, different cameras used, various illumination conditions, and so on. Therefore, the conventional approaches that use some predefined or learned form of SCM are unable to overcome all of the above mentioned circumstances.

In this article, therefore, we have proposed a new adult image detection method which employs robust skin segmentation. Our method first detect eyes, known as the most stable component in the face, and constructs an input image-adapted skin color distribution model by choosing reliable skin samples near the detected eyes. Skin areas over the entire image are then segmented with the generated color model. Since we make the adaptive SCM online, robust skin segmentation can be achieved. Subsequently, five types

of geometric features that characterize adult images are extracted from the chosen skin region, and we finally verify if the skin region contains naked bodies by effectively fusing the multiple features through a three-layer perceptron neural network.

Our adult image identification algorithm may be potentially applicable to internet censorship, which is one of the ways to protect users from accessing violent or sexual websites that are especially harmful to teenagers. However, since it has been a rather controversial topic among most web users, internet censorship should be considered carefully.

Our future work will focus on including individual human body parts, such as female breasts and hips, which can characterize adult images more clearly. Furthermore, we will undertake various experiments with other adult images captured in more complex environments.

ACKNOWLEDGMENTS

This research was supported by The Ministry of Knowledge Economy (MKE), Korea, under the Information Technology Research Center (ITRC) support program supervised by the National IT Industry Promotion Agency (NIPA) [Grant No. NIPA-2010-(C1090-1021-0010)].

REFERENCES

[1] J.-L. Shih, C.-H. Lee, and C.-H. Yang, "An adult images identification system employing image retrieval technique", Pattern Recogn. Lett. **28**, 2367 (2007).
[2] D. A. Forsyth and M. M. Fleck, "Automatic detection of human nudes", Int. J. Comput. Vis. **32**, 63 (1999).
[3] J. Ze Wang, J. Li, G. Wiederhold, and O. Firschein, "System for screening objectionable images", Comput. Commun. **21**, 1355 (1998).
[4] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection", Int. J. Comput. Vis. **46**, 81 (2002).
[5] A. Bosson, G. C. Cawley, Y. Chian, and R. Harvey, "Non-retrieval: Blocking pornographic images", Lect. Notes Comput. Sci. **2383**, 50 (2002).
[6] B. Jedynak, H. Zheng, and M. Daoudi, "Statistical models for skin detection", *Proc. IEEE Workshop on Statistical Analysis in Computer Vision* (IEEE, Los Alamitos, 2003), Vol. **8**, p. 92.
[7] H. Zheng, M. Daoudi, and B. Jedynak, "Blocking adult images based on statistical skin", Electron. Lett. Comput. Vis. Image Anal. **4**, 1 (2004).
[8] 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. **18**, 272 (2006).
[9] J.-S. Lee, Y.-M. Kuo, P.-C. Chung, and E.-L. Chen, "Naked image detection based on adaptive and extensible skin color model", Pattern Recogn. **40**, 2261 (2007).
[10] S. Ioffe and D. A. Forsyth, "Probabilistic methods for finding people", Int. J. Comput. Vis. **43**, 45 (2001).
[11] J. Yang, Z. Fu, T. Tan, and W. Hu, "A novel approach to detecting adult images", *Proc. International Conference on Pattern Recognition* (IEEE, Los Alamitos, 2004), Vol. **4**, p. 479.
[12] Y. H. Kuan and C. H. Hsieh, "Content-based pornography image detection", *Proc. International Conference on Imaging Science, Systems, and Technology* (CSREA, Athens, 2004).
[13] 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. **29**, 1019 (2007).
[14] J. Song, Z. Chi, and J. Liu, "A robust eye detection method using combined binary edge and intensity information", Pattern Recogn. **39**, 1110 (2006).
[15] J. Wu and Z.-H. Zhou, "Efficient face candidates selector for face detection", Pattern Recogn. **36**, 1175 (2003).
[16] A. Nikolaidis and I. Pitas, "Facial feature extraction and pose determination", Pattern Recogn. **33**, 1783 (2000).
[17] F. Smeraldi, O. Carmona, and J. Bigün, "Saccadic search with Gabor features applied to eye detection and real-time head tracking", Image Vis. Comput. **18**, 323 (2000).
[18] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis", IEEE Trans. Multimedia **1**, 264 (1999).
[19] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images", IEEE Trans. Pattern Anal. Mach. Intell. **24**, 696 (2002).
[20] P. T. Jackway and M. Deriche, "Scale-space properties of the multi-scale morphological dilation-erosion", IEEE Trans. Pattern Anal. Mach. Intell. **18**, 38 (1996).
[21] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal face authentication using morphological elastic graph matching", IEEE Trans. Image Process. **9**, 555 (2000).
[22] N. Otsu, "A threshold selection method from gray-level histogram", IEEE Trans. Syst. Man Cybern. **9**, 62 (1979).
[23] H. Zhao, Q. Li, and H. Feng, "Multi-focus color image fusion in the HSI space using the sum-modified-Laplacian and a coarse edge map", Image Vis. Comput. **26**, 1285 (2008).
[24] B. Li and M. Q.-H. Meng, "Texture analysis for ulcer detection in capsule endoscopy images", Image Vis. Comput. **27**, 1336 (2009).
[25] A. S. Malik and T.-S. Choi, "A novel algorithm for estimation of depth map using image focus for 3D shape recovery in the presence of noise", Pattern Recogn. **41**, 2200 (2008).
[26] G.-J. Liu, X.-L. Tang, H.-D. Cheng, J.-H. Huang, and J.-F. Liu, "A novel approach for tracking high speed skaters in sports using a panning camera", Pattern Recogn. **42**, 2922 (2009).