

# An adult image identification system employing image retrieval technique

Jau-Ling Shih <sup>\*</sup>, Chang-Hsing Lee, Chang-Shen Yang

*Department of Computer Science and Information Engineering, Chung Hua University, No. 707, Section 2, WuFu Road, Hsinchu, Taiwan, ROC*

Received 7 February 2007; received in revised form 13 July 2007

Available online 15 August 2007

Communicated by G. Sanniti di Baja

## Abstract

As the development of the Internet, children are easily exposed on the pornography through web browsers. To block adult images, content-based image retrieval technique is employed for adult image identification. First, the background is removed to obtain the rectangular region of interesting based on the detection of skin-like pixels. For each input image, the MPEG-7's color, texture, and the proposed shape feature is used to retrieve 100 most similar images from the image database which contains both adult and non-adult images. If the retrieved images contains more than  $T_{ad}$  adult images, the input one is identified as an adult image. Otherwise, it is identified as a non-adult image. Experiment results have shown the effectiveness of the proposed method.

© 2007 Elsevier B.V. All rights reserved.

**Keywords:** Adult image identification; Scalable color descriptor; Edge histogram descriptor

## 1. Introduction

With the rapid growth of the Internet, any user can access and browse a fruitful of multimedia contents on the web. However, some contents, such as pornographic images, are not appropriate for all users, particularly the children. How to prevent people from accessing this type of harmful material becomes an important issue. There are a number of commercial products and research solutions developed to filter and block adult contents from web browsing. Typically, current adult contents filtering approaches can be classified into three categories: IP-based black-list blocking, textual content-based filtering, and visual content-based filtering. The IP-based black-list blocking approach first builds a set of URLs of objectionable web sites. Access to a requested web page is prohibited if its URL is contained in the black-list. However, the con-

tents in the Internet is highly dynamic because many web sites and web pages will appear everyday whereas many will disappear. In fact, it is hard to keep the black-list of all objectionable web sites up to date. Thus, the IP-based black-list filtering approach is inefficient and impractical. The textual content-based filtering approach attempts to block adult web sites based on the analysis of the textual contents. A number of offensive keywords or phrases which occur most frequently on the adult web sites will be identified manually or through machine learning or data mining techniques. Each word or phrase in a requested web page is compared with those in a keyword dictionary containing prohibited keywords or phrases. Access to a requested web page is prohibited if a lot of offensive keywords or phrases occur in the requested web page. However, the textual content-based filtering approach suffers from the well-known “over-blocking” phenomenon which blocks access to educational web sites such as health or sexology. In addition, many adult web sites with text incorporated in elaborate images cannot be blocked by textual content analysis. Therefore, many researchers investigate visual

<sup>\*</sup> Corresponding author. Tel.: +886 3 5186407; fax: +886 3 5186416.  
E-mail address: [sjl@chu.edu.tw](mailto:sjl@chu.edu.tw) (J.-L. Shih).

content-based filtering approach to analyze the image contents in a web page.

The identification of adult images is typically treated as an image classification problem. In general, the skin regions are first segmented by color and texture information. The features extracted from the detected skin regions, including color, texture, and shape features, are then used to discriminate benign images from adult images. Some researches try to employ content-based image retrieval approach for identifying adult images. Given an image, a number of similar images are retrieved from an image database consisting of both adult and non-adult images. If most of the retrieved images are adult images, the input image is identified as an adult image. Otherwise, it is regarded as a non-adult image.

The pioneer work for identifying adult images by analyzing the image contents is proposed by Forsyth et al. (Fleck et al., 1996; Forsyth and Fleck, 1996, 1997). Their approach employs a skin filter and a human figure grouper to find naked people in an image. Both color and texture information were exploited to detect skin regions. Geometric analysis was then conducted to assemble together the detected skin regions into the human body shape.

Wang et al. developed the WIPE (wavelet image pornography elimination) system (Wang et al., 1997, 1998) for screening objectionable images. The system successively eliminate pornographic images by using a combination of an icon filter, a graph-photo detector, a color histogram filter, a texture filter, and a wavelet-based shape matching algorithm.

Jones and Rehg developed a statistical color model for detecting skin and non-skin regions (Jones and Rehg, 1998). The set of aggregated features used for adult image identification includes percentage of detected skin pixels, average probability of the skin pixels, size in pixels of the largest connected component in skin regions, number of connected components in skin regions, percent of colors with no entries in the skin and non-skin histograms, and the height and width of the image. Both adult and non-adult images are used to train a neural network classifier for adult image identification.

Zeng et al. developed the image guarder system to detect adult images by analyzing the image contents (Zeng et al., 2004). An illumination adaptive statistical color model (IASCM) is proposed to detect the skin pixels under variant illumination. To characterize the smoothness property of the skin regions, a simple texture validation approach is applied to remove spurious skin regions that violate the texture distribution of the skin regions. Color, texture, and shape features are then extracted from the detected skin regions. Support vector machine (SVM) is finally used to identify adult images according to the extracted features.

Yang et al. analyzed the shape of human body trunk to identify adult images (Yang et al., 2004). First, a non-linear skin classifier is used to discriminate between skin and non-skin pixels. Each ROI (region of interest) which contains a lot of skin pixels is then extracted by image partitioning

and region growing approach. Based on the ROIs, the contour of body trunk (CBT) is segmented and a number of shape features are extracted for image classification.

Zheng et al. used an adult image filter and a harmful symbol filter to block objectionable images (Zheng et al., 2004). The adult image filter first detects the skin region using an improved statistical color model based on the maximum entropy of the probability distribution. A multi-layer perceptron neural network is employed for adult image classification. To detect harmful symbols, an edge-based Zernike moment method is used to represent the shape of symbol objects.

Zhu et al. proposed an adaptive skin color detection model to filter objectionable images (Zhu et al., 2004). First, the skin-similar pixels are identified using a generic skin model. These pixels are used to train a Gaussian mixture model (GMM) with several Gaussian kernels using the standard expectation–maximization (EM) algorithm. SVM, which uses the spatial and shape features as well as the Gaussian parameters, is then used to identify the skin component using the trained GMM.

The WebGuard system used machine learning techniques for pornographic web site classification and filtering (Hammami et al., 2006). Besides the textual and structural analysis, visual analysis based on skin colors is employed for adult content detection and filtering. A skin color related visual feature, which represents the percentage of skin pixels within a web page, is used to identify pornographic web sites. A learning data set consisting of 2000 pornographic web sites and 2000 non-pornographic ones is employed to obtain the classification threshold of the percentage of skin pixels.

Hu et al. proposed a framework for recognizing pornographic web pages (Hu et al., 2007). The C4.5 decision tree was used to classify the web pages into three categories: continuous text pages, discrete text pages, and image pages. To determine whether a web page is pornographic one or not, these three categories were distinctly classified by a continuous text classifier, a discrete text classifier, and a combination of the image classifier and the discrete text classifier. The image classifier, which is similar to the Yang's algorithm (Yang et al., 2004), used five kinds of image features extracted from the contour shape to distinguish pornographic images from normal images.

Yoo employed the MPEG-7 visual descriptors, including edge histogram descriptor (EHD), color layout descriptor (CLD), and homogeneous texture descriptor (HTD), to identify and rate adult images (Yoo, 2004). Given an input image, ten most similar images are retrieved from a database containing both adult and non-adult images. If most of the retrieved images are adult images, the input one is regarded as an adult image. Otherwise, it is identified as a non-adult one. However, the background color information will disturb the features extracted from the whole image and thus will degrade the identification accuracy.

Kuan and Hsieh used image retrieval technique for pornography image detection (Kuan and Hsieh, 2004). First,

each skin region is detected. The color histogram, intensity moments, and bit-slice edge moments within each skin region are calculated to form the feature vector. The adult image is identified by comparing its feature vector with those in the database. In this system, the features are extracted from the skin regions and thus will not be affected by the background regions.

In general, treating the identification of adult images as an image classification problem faces two major problems: (1) selection of the appropriate training set; (2) determination of the classification parameters/thresholds. Generally, the training set is formed by selecting as many representative images as possible. The classification parameters/thresholds are usually determined by some experimental results or through a machine learning approach. These classification parameters/thresholds are fixed without considering the characteristics of different web users. Thus, an adult image identification system based on content-based image retrieval approach will be developed. In the proposed system, the level whether an image will be identified as an adult image or not can be adjusted to meet the appropriate usage of different web users. In summary, the proposed system servers three main characteristics: (1) background is removed to find the region of interest, (2) a new shape descriptor is proposed and combined with the color and texture descriptors to improve the identification accuracy, and (3) the level whether an image will be identified as an adult image or not can be easily adjusted to meet the appropriate usage of different web users.

The rest of the paper is organized as follows: in Section 2, a preprocessing method is introduced to remove the background region. Then, the MPEG-7's color descriptor, texture descriptor, and the proposed compactness descriptor will be employed for adult image identification. Section 3 gives the experimental results to show the effectiveness of the proposed method. Finally, conclusions are given in Section 4.

## 2. The proposed adult image identification approach

The flowchart of our proposed system is shown in Fig. 1. First, the background region is removed to obtain the region of interest (ROI) using a generic skin-like detection method. A set of visual features, including the MPEG-7's color descriptor, texture descriptor, and a new proposed shape feature called compactness descriptor, is extracted from the ROI for similar image retrieval. Given an input image, 100 most similar images are retrieved from the image database which contains both adult and non-adult images. If more than  $T_{ad}$  of the retrieved images are adult images, the input image is identified as an adult image. Otherwise, it is identified as a non-adult one.

### 2.1. Background removal

To obtain the region of interest (ROI), the background region which contains definitely non-skin-like pixels will be

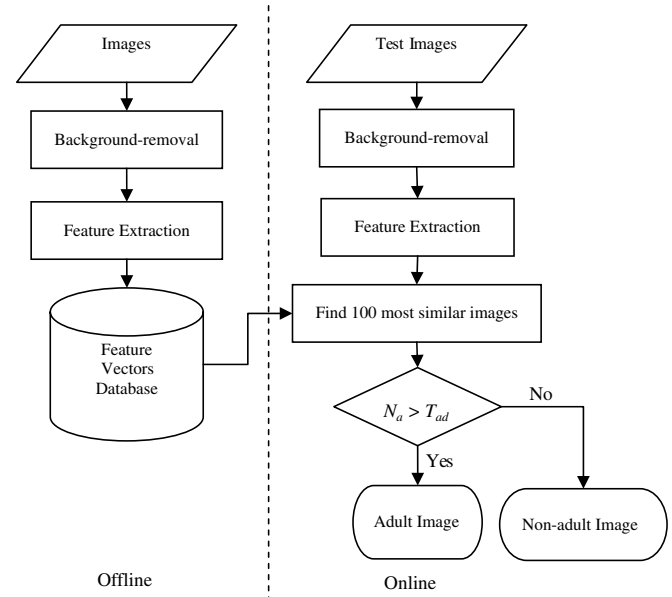


Fig. 1. The architecture of the proposed system.

first removed. In this paper, color information is employed to detect skin-like pixels. Garcia and Tziritas have shown that skin-like pixels are more correlated with Cr and Cb components than Y component (Garcia and Tziritas, 1999). Therefore, the input image is first transformed from the RGB color model to the YCbCr color space. A pixel is considered to be skin-like if its Cb and Cr components meet the following constraints

$$Cr \geq \max\{-2(Cb + 24), -(Cb + 17), -4(Cb + 32), 2.5(Cb + \theta_1), \theta_3, 0.5(\theta_4 - Cb)\}$$

and

$$Cr \leq \min\{(220 - Cb)/6, 4(\theta_2 - Cb)/3\},$$

where  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ , and  $\theta_4$ , are constants given by

$$\theta_1 = \begin{cases} -2 + (256 - Y)/16, & \text{if } Y > 128, \\ 6, & \text{otherwise,} \end{cases}$$

$$\theta_2 = \begin{cases} 20 - (256 - Y)/16, & \text{if } Y > 128, \\ 12, & \text{otherwise,} \end{cases}$$

$$\theta_3 = \begin{cases} 6, & \text{if } Y > 128, \\ 20 - (256 - Y)/16, & \text{otherwise,} \end{cases}$$

and

$$\theta_4 = \begin{cases} -8, & \text{if } Y > 128, \\ -16 + Y/16, & \text{otherwise.} \end{cases}$$

As a result, a binary image can be used to indicate whether a pixel is skin-like or not (see Fig. 2) where the skin-like pixels be represented as white pixels and non-skin-like pixels are represented as black pixels. The binary image is then decomposed into a number of equal-sized ( $32 \times 32$  pixels) non-overlapping blocks. For each block, if more than half of the pixels are detected as skin-like pixels, the block is regarded as a skin-like block; otherwise, it is regarded as a non-skin-like block (see Fig. 2d).

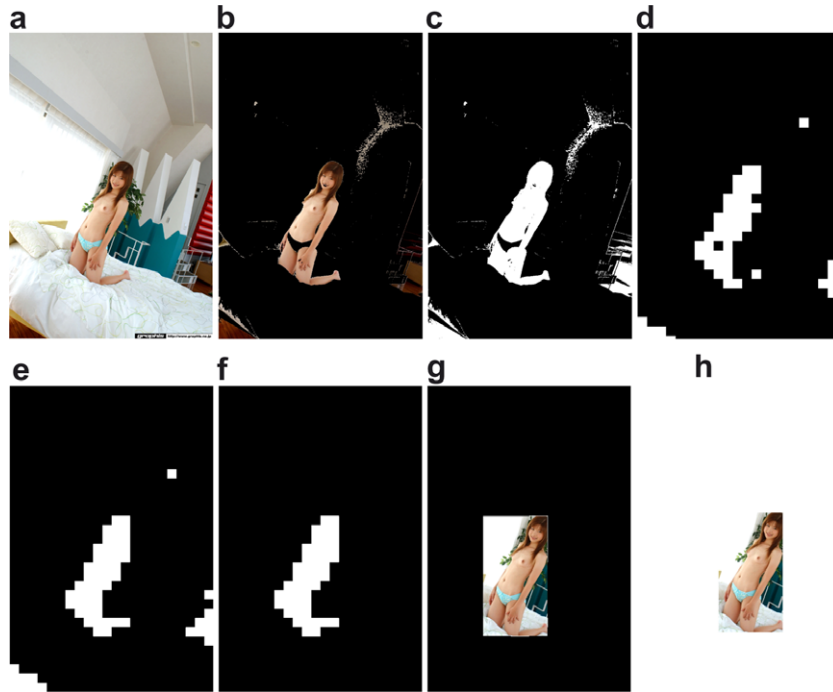


Fig. 2. An example to illustrate the background removal process. (a) The original image, (b) the image after skin-like detection, (c) the binary image used to indicate skin-like pixels, (d) the skin-like blocks, (e) the blocks after closing processing, (f) remove the noise blocks, (g) the minimum bounding box for the detected skin region and (h) the background-removed image.

The morphology closing operation with a  $3 \times 3$  structure element (Gonzalez and Woods, 2002) is employed to group skin-like blocks together (see Fig. 2e). As a result, a number of connected regions will be obtained. To remove spurious skin-like regions, the central connected component is found and regarded as the skin region of an image (see Fig. 2f). The central connected component is then circumscribed by a minimum rectangular bounding box (see Fig. 2g). This bounding box is defined as the region of interest (ROI) (see Fig. 2h) and the other area is regarded as the background. If no skin-like region is detected or the width/height of the rectangular ROI is smaller than 50 pixels, the whole image is regarded as the ROI.

## 2.2. Feature extraction

After the ROI is segmented from an input image, some features will be extracted from the ROI to search similar images from a pre-designed image database. In general, color is an important feature of the skin-like region. In addition, the texture distributions can be used to discriminate between adult and non-adult images since most non-adult images have sharper edges whereas the adult images will exhibit smooth color textures. Furthermore, the shape of the skin region is an important feature for adult images detection. Hence, in our system, a combination of color, texture, and shape feature is used for adult image identification.

Two MPEG-7 visual descriptors: scalable color descriptor (SCD) and edge histogram descriptor (EHD) (Yamada et al., 2001) are used as the color and texture features,

respectively. A new descriptor called compactness descriptor (CD) is used to represent the shape of a ROI.

### 2.2.1. Scalable color descriptor (SCD)

SCD describes the color histogram of an image in the HSV color space. The transformation from RGB color space to HSV color space is defined as follows:

$$H = \cos^{-1} \left\{ \frac{[(R - G) + (R - B)]/2}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right\},$$

$$S = \frac{\max(R, G, B) - \min(R, G, B)}{\max(R, G, B)},$$

$$V = \frac{\max(R, G, B)}{255}.$$

The HSV color space is first uniformly quantized into 256 color bins with 16 hues ( $H$ ), four saturations ( $S$ ), and four values ( $V$ ). For each image, the color histogram is formed by counting the number of pixels for each color bin. Therefore, the scalable color descriptor, *scd*, of an image is defined as

$$\text{scd} = [s[1], s[2], \dots, s[256]],$$

where  $s[i]$ ,  $1 \leq i \leq 256$ , is the probability of  $i$ th bin.

### 2.2.2. Edge histogram descriptor (EHD)

In MPEG-7, EHD is provided to represent the local edge distribution of an image. The image is first decomposed into 16 non-overlapping sub-images as shown in Fig. 3. An edge



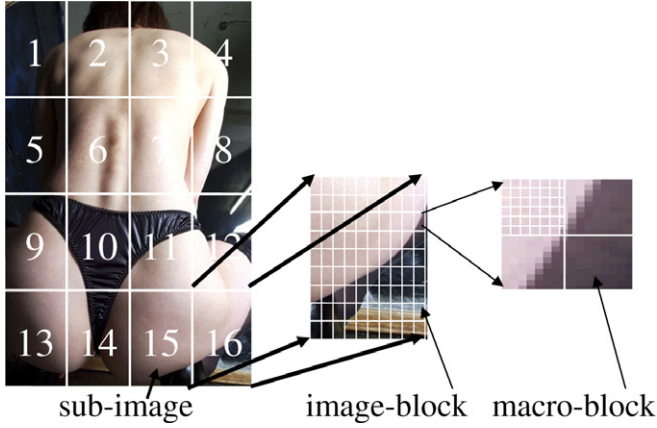


Fig. 3. Definition of 16 non-overlapping sub-images and image-blocks.

histogram is used to represent the local edge distribution of each sub-image. To derive the edge histogram, each sub-image is further divided into a predefined number of non-overlapping square image-blocks. For each image, it is shown that the total number of image-blocks around 1100 seems to get good directional edge features. Each image-block is then classified into one of the five edge categories: vertical, horizontal, 45° diagonal, 135° diagonal, and non-directional edges. First, each image-block is treated as a  $2 \times 2$  super-pixel image-block. Appropriate edge detectors are then used to compute the corresponding edge strengths. If the maximum edge strength is larger than a given threshold, the corresponding edge orientation of the image-block is determined. Otherwise, the block is classified as a non-edge block. For each sub-image, the edge category of each image-block will be accumulated in the edge histogram bins. Since these are 16 sub-images, a total of 80 ( $16 \times 5$ ) histogram bins can be obtained. In addition to the local edge histograms, the global edge histogram is calculated to capture the global edge distribution of an image. The global edge histogram is formed by accumulating the same type of edge distributions of all sub-images. Combining together the local and the global histograms, a total of 85 histogram bins are used for similarity matching. In summary, the edge histogram descriptor, *ehd*, is defined as

$$\mathbf{ehd} = [e[1], e[2], \dots, e[80], ge[1], ge[2], \dots, ge[5]],$$

where  $e[i]$ ,  $1 \leq i \leq 80$ , and  $ge[j]$ ,  $1 \leq j \leq 5$ , represent respectively the local edge histogram and the global edge histogram.

### 2.2.3. Compactness descriptor (CD)

In this study, CD is proposed to describe the shape of the ROI where skin pixels are represented by white pixels and background pixels are represented by black pixels. The binary ROI is decomposed into a number of blocks of different sizes (see Fig. 4). For each block, the proportion of the skin pixels, PSP, is used to represent the compactness of the block

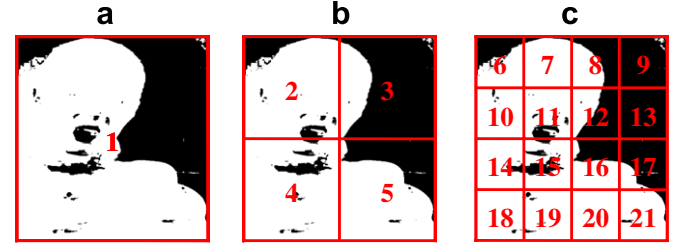


Fig. 4. Three different sized blocks. The number in each block is the block index. (a) The global block, (b) four medium-sized sub-blocks and (c) 16 smallest sub-blocks.

$$\text{PSP}_j = \frac{N_j}{W_j \times H_j},$$

where  $N_j$  is the number of skin pixels in block  $j$ ,  $W_j$  and  $H_j$  are the width and height of the block, respectively.  $\text{PSP}_1$  represents the compactness of the global block,  $\text{PSP}_j$ , for  $2 \leq j \leq 5$ , represents the compactness of the four medium-sized blocks (see Fig. 4b), and  $\text{PSP}_j$ , for  $6 \leq j \leq 21$  represents the compactness of the 16 smallest-sized blocks (see Fig. 4c).

The global compactness value,  $\text{PSP}_1$ , is set as the first feature value,  $cd_1$ , of the compactness descriptor, *cd*. In general, an adult image is more relevant to the compactness than the position of each block. Therefore, the compactness values of the medium-sized blocks are sorted to obtain the feature value  $cd_j$  for  $2 \leq j \leq 5$ . Similarly, the feature value  $cd_j$  for  $6 \leq j \leq 21$ , can be obtained by sorting the compactness values of the smallest-sized blocks. In summary, a total of 21 feature values can be obtained. That is, *cd* is represented as

$$\mathbf{cd} = [cd[1], cd[2], \dots, cd[21]].$$

### 2.2.4. Identification of adult images using image retrieval technique

The identification of adult images is based on the image retrieval technique. Given an input image, 100 most similar images are first retrieved from a pre-designed image database which consists of both adult and non-adult images. Based on the feature vectors, *scd*, *ehd*, and *cd*, the distance between the input image  $t$  and every matching image  $s$  are respectively calculated and denoted as  $d_{\text{SCD}}$ ,  $d_{\text{EHD}}$ , and  $d_{\text{CD}}$ :

$$d_{\text{SCD}}(t, s) = \|\mathbf{scd}_t - \mathbf{scd}_s\| = \sum_{i=1}^{256} |\text{scd}_t[i] - \text{scd}_s[i]|,$$

$$d_{\text{EHD}}(t, s) = \|\mathbf{ehd}_t - \mathbf{ehd}_s\| = \sum_{i=1}^{80} |e_t[i] - e_s[i]| + 5 \times \sum_{i=1}^5 |ge_t[i] - ge_s[i]|,$$

$$d_{\text{CD}}(t, s) = \|\mathbf{cd}_t - \mathbf{cd}_s\| = \sum_{i=1}^{21} |cd_t[i] - cd_s[i]|.$$

According to  $d_{\text{SCD}}(t, s)$ , the  $g$  images ( $g = 100$  in this paper) with the smallest distance values are first found and sorted

in an ascending order of the distance value. The  $i$ th sorted image will be assigned a grade of  $g - i + 1$ ,  $1 \leq i \leq g$ . In addition, the grades of the other non-similar images are set as zero. In general, the image with the smallest distance value will get the highest grade and vice versa. For the test image  $t$ , each matching image  $s$  will be assigned a grade,  $G_{\text{SCD}}(t, s)$ , according to the SCD feature. Similar grade assignment process will be applied to each matching image according to the distance calculated by using EHD and CD, denoted as  $G_{\text{EHD}}(t, s)$  and  $G_{\text{CD}}(t, s)$ . As a result, for each matching image  $s$ , the overall grade is calculated by the summation of  $G_{\text{SCD}}(t, s)$ ,  $G_{\text{EHD}}(t, s)$  and  $G_{\text{CD}}(t, s)$

$$G(t, s) = G_{\text{SCD}}(t, s) + G_{\text{EHD}}(t, s) + G_{\text{CD}}(t, s).$$

Based on these grades, the images with the largest grades will be regarded as the most similar ones. In this paper, a total of 100 most similar images will be found. Let  $N_a$  denotes the number of adult images in the retrieval set. If  $N_a$  is larger than a threshold  $T_{\text{ad}}$ , the input image is identified as an adult image. Otherwise, it is identified as a non-adult image. Note that  $T_{\text{ad}}$  indicates the level a test image to be identified as an adult image. For instance, when the kids browse web pages,  $T_{\text{ad}}$  can be set smaller such that most adult images will be filtered out. Otherwise, for generic search engines, the proper  $T_{\text{ad}}$  can be set to reduce the number of non-adult images which are mis-identified as adult ones.

### 3. Experimental results

To form the image database, we randomly select 1000 adult images as positive examples and 10,039 non-adult images as negative ones. These images are gathered from the Internet. The set of non-adult images contains 1039 bikini and 9000 non-bikini images. These test images consist of 10,000 adult images and 10,000 non-adult images. The distributions of sizes of these images are shown in Fig. 5.

Two experiments are implemented to test the performance. The image database in the first experiment contains

1000 adult images and 9000 non-adult images, whereas the image database in the second experiment contains 1000 adult images, 9000 non-adult images, and 1039 bikini images.

Given an input image, if its width or height is less than 50 pixels, this image is directly identified as a non-adult image, since it is very likely an icon image. Otherwise, the 100 most similar images are retrieved from the image database. Let  $N_a$  denotes the number of adult images in the retrieval set. If  $N_a$  is larger than a threshold  $T_{\text{ad}}$ , the query one is identified as an adult image. Otherwise, it is identified as a non-adult image. The performance is measured by the detection rate (DR)

$$\text{DR} = \frac{|Q|}{|B|},$$

where  $B$  is the set of manually labeled adult/non-adult images and  $Q$  is the set of identified adult/non-adult images.

#### 3.1. Experiment 1

Table 1 shows the DR for different feature combinations without background removal. As mentioned previously,  $T_{\text{ad}}$  controls the level an input image will be identified as an adult image. When  $T_{\text{ad}}$  is 10, the DRs of adult and non-adult images are 96.43% and 79.27% by using SCD to retrieve similar images. Meanwhile, the DRs can be raised to 99.47% and 79.46% for adult and non-adult images by using the feature combination of SCD, EHD, and CD.

Table 2 shows the DRs with background removal. When  $T_{\text{ad}}$  is 10, the DRs of adult and non-adult images are 99.72% and 80.85% by using the feature combination of SCD, EHD, and CD. It can be seen that, almost all adult images can be efficiently blocked whereas about 20% of non-adult images is mis-identified as adult ones. Moreover, if  $T_{\text{ad}}$  is 30, the DR of adult images is slightly decreased to be 97.77% while the DR of non-adult images is greatly increased to be 90.69%. From Tables 1 and 2,

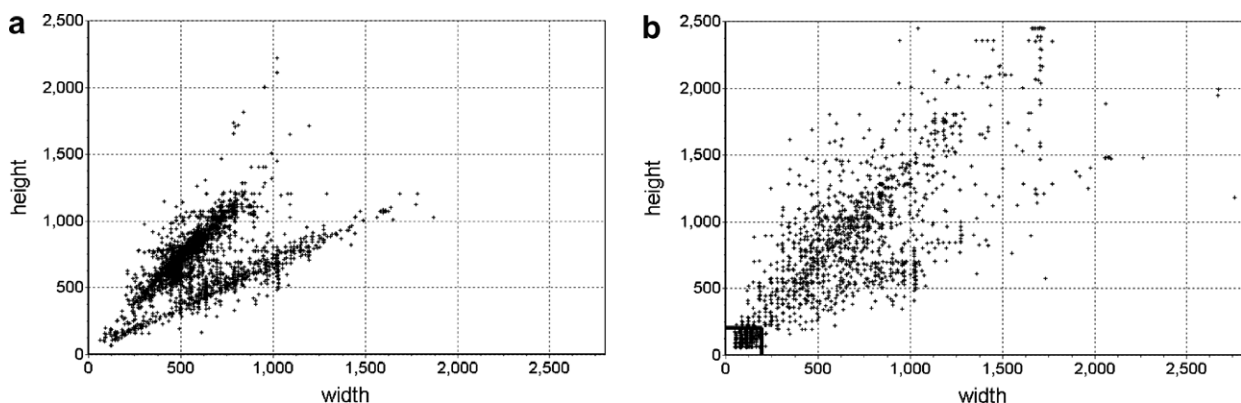


Fig. 5. Distributions of images sizes. (a) Adult images and (b) non-adult images.

Table 1

Detection rate (DR) on the first experiment without background removal

$T_{ad}$		SCD	EHD	CD	SCD + EHD	SCD + EHD + CD
10	Adult	96.43	94.66	92.51	99.10	99.47
	Non-adult	79.27	88.82	82.49	82.16	79.46
20	Adult	92.35	88.90	83.15	97.19	97.74
	Non-adult	86.79	94.13	88.21	90.13	87.60
30	Adult	87.23	83.09	74.64	93.29	94.01
	Non-adult	90.21	96.13	90.94	94.83	92.05
40	Adult	81.08	76.54	66.15	85.62	87.40
	Non-adult	92.50	97.21	92.49	98.38	95.32
50	Adult	73.49	68.49	55.88	73.60	77.50
	Non-adult	94.21	98.27	94.38	99.74	97.32

Table 2

Detection rate (DR) on the first experiment with background removal

$T_{ad}$		SCD	EHD	CD	SCD + EHD	SCD + EHD + CD
10	Adult	98.14	93.81	97.2	99.34	99.72
	Non-adult	82.16	88.79	84.94	82.32	80.85
20	Adult	96.20	88.63	95.28	98.21	99.22
	Non-adult	86.93	93.96	86.38	89.19	86.43
30	Adult	93.69	83.01	92.42	95.41	97.77
	Non-adult	89.49	95.99	87.58	94.65	90.69
40	Adult	89.57	76.57	81.97	88.80	94.82
	Non-adult	91.74	97.16	90.19	98.47	94.32
50	Adult	83.22	69.41	68.36	78.03	87.84
	Non-adult	93.60	98.16	92.66	99.66	97.35

when  $T_{ad}$  is 50, the DR of adult images will increase from 77.5% to 87.84% by using a feature combination of SCD, EHD, and CD if background region is removed.

### 3.2. Experiment 2

Table 3 shows the DRs without background removal by using different feature combinations. It can be seen that when  $T_{ad}$  is 10, the DRs of adult and non-adult images are 92.34% and 81.94% by using SCD to retrieve similar images. Meanwhile, the DRs can be raised to 98.99% and 83.49% for adult and non-adult images by using the feature combination of SCD, EHD, and CD.

Table 3

Detection rate (DR) on our proposed method of the second experiment without background removal

$T_{ad}$		SCD	EHD	CD	SCD + EHD	SCD + EHD + CD
10	Adult	92.34	93.53	87.59	98.08	98.99
	Non-adult	81.94	90.24	86.13	85.08	83.49
20	Adult	83.76	84.30	73.70	91.25	95.26
	Non-adult	88.80	95.88	90.82	92.58	90.54
30	Adult	75.26	70.01	63.17	77.59	87.66
	Non-adult	91.64	98.13	92.91	96.41	94.19
40	Adult	65.19	45.25	53.68	61.29	77.09
	Non-adult	93.74	99.24	94.55	99.13	96.69
50	Adult	54.47	16.19	45.09	42.53	64.89
	Non-adult	95.35	99.82	95.85	99.94	98.44

Table 4

Detection rate (DR) on our proposed method on the second experiment with background removal

$T_{ad}$		SCD	EHD	CD	SCD + EHD	SCD + EHD + CD
10	Adult	95.53	93.21	93.89	98.67	99.54
	Non-adult	83.41	89.82	86.43	84.93	83.49
20	Adult	87.31	85.60	83.53	92.99	97.74
	Non-adult	88.24	95.30	88.80	92.24	89.66
30	Adult	77.33	69.67	72.14	77.80	93.02
	Non-adult	91.14	97.71	90.74	96.72	93.85
40	Adult	65.28	34.27	53.29	57.72	83.78
	Non-adult	93.52	99.30	93.41	99.21	96.94
50	Adult	51.84	7.20	38.06	35.64	68.64
	Non-adult	95.35	99.85	95.79	99.91	99.11

Table 5

Comparison of detection rate (DR)

Methods	Adult	Non-adult
Our proposed method	99.54	83.49
Forsyth and Fleck (1997)	79.30	88.70
WebGuard (Hammami et al., 2006)	97.40	<sup>a</sup>
Hu et al. (2007)	92.80 <sup>b</sup>	94.00 <sup>b</sup>
Jones and Rehg (1998)	85.80	92.50
Kuan and Hsieh (2004)	90.80	90.55
WIPE (Wang et al., 1998)	96.00	91.00
Yang et al. (2004)	95.45	91.50
AIRS (Yoo, 2004)	99.25	77.00
Zeng et al. (2004)	76.50	95.00
Zhu et al. (2004)	94.67	94.67

<sup>a</sup> No corresponding data.<sup>b</sup> Using image classifier only.

Table 4 shows the DRs with background removal by using different feature combinations. When  $T_{ad}$  is 10, the DRs of adult and non-adult images are 99.54% and 83.49% by using the feature combination of SCD, EHD, and CD. It can be seen that, almost all adult images can be efficiently blocked whereas about 16% of non-adult images is mis-identified as adult ones. Moreover, when  $T_{ad}$  is 30, the DR of adult images slightly decreases to be 93.02% while the DR of non-adult images greatly increases to be 93.85%. Finally, Table 5 shows the performance comparison among the proposed method and other methods. We can see that, our proposed method is superior to others.

## 4. Conclusions

In this paper, an adult image identification approach is proposed based on content-based image retrieval technique. First, the background is removed to obtain the region of interest where non-skin-like pixels are removed. Based on the MPEG-7's SCD, EHD, and the proposed CD feature, for each input image, 100 most similar images are retrieved from a pre-designed image database. If more than  $T_{ad}$  retrieved images are adult images, the query one is identified as an adult image. Otherwise, it is identified as a

non-adult image. Experiment results have shown the effectiveness of the proposed method.

### Acknowledgement

This research was supported in part by the National Science Council, Taiwan under Contract NSC 95-2221-E-216-016.

### References

- Fleck, M.M., Forsyth, D.A., Bregler, C., 1996. Finding naked people. In: 4th European Conf. on Computer Vision 2, pp. 592–602.
- Forsyth, D.A., Fleck, M.M., 1996. Identifying nude pictures. In: 3rd IEEE Workshop on Applications of Computer Vision, pp. 103–108.
- Forsyth, D.A., Fleck, M.M., 1997. Body plans. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 678–683.
- Garcia, C., Tziritas, G., 1999. Face detection using quantized skin color regions merging and wavelet packet analysis. *IEEE Trans. Multimed.* 1, 264–277.
- Gonzalez, R.C., Woods, R.E., 2002. *Digital Image Processing*, second ed. Prentice Hall.
- Hammami, M., Chahir, Y., Chen, L., 2006. WebGuard: A web filtering engine combining textual, structural, and visual content-based analysis. *IEEE Trans. Knowled. Data Eng.* 18 (2), 272–284.
- Hu, W., Wu, O., Chen, Z., Fu, Z., Maybank, S., 2007. Recognition of pornographic web pages by classifying texts and images. *IEEE Trans. Pattern Anal. Machine Intell.* 29 (6), 1019–1034.
- Jones, M.J., Rehg, J.M., 1998. Statistical color models with application to skin detection. Technical Report Series, Cambridge Research Laboratory.
- Kuan, Y.H., Hsieh, C.H., 2004. Content-based pornography image detection. In: International Conf. on Imaging Science, Systems, and Technology.
- Wang, J.Z., Wiederhold, G., Firschein, O., 1997. System for screening objectionable images using Daubechies' wavelets and color histograms. *IDMS'97* 1309, 20–30.
- Wang, J.Z., Li, J., Wiederhold, G., Firschein, O., 1998. System for screening objectionable images. *Comput. Comm.*, 1355–1360.
- Yamada, A., Pickering, M., Jeannin, S., Cieplinski, L., Ohm, J.R., Kim, M., 2001. MPEG-7 information technology-multimedia content descriptor interface – Part 3: Visual. Final Committee Draft, ISO/IEC/JTC1/SC29/WG11/N4062.
- Yang, J., Fu, Z., Tan, T., Hu, W., 2004. A novel approach to detecting adult images. In: 17th Internat. Conf. on Pattern Recognition 4, pp. 479–482.
- Yoo, S.J., 2004. Intelligent multimedia information retrieval for identifying and rating adult images. *KES 2004*, 164–170.
- Zeng, Wei, Gao, Wen, Zhang, Tao, Liu, Yang, 2004. Image Guarder: An Intelligent Detector for Adult. In: Asian Conf. on Computer Vision, pp. 198–203.
- Zheng, H., Liu, H., Daoudi, M., 2004. Blocking objectionable images: Adult images and harmful symbols. *IEEE Internat. Conf. Multimed. Expo 2*, 27–30.
- Zhu, Q., Wu, C.T., Cheng, K.T., Wu, Y.L., 2004. An Adaptive Skin Model and Its Application to Objectionable Image Filtering. *ACM Multimedia*, New York, USA, pp. 56–63.