Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 40 (2007) 2261–2270
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Naked image detection based on adaptive and extensible skin color model

Jiann-Shu Lee[a,*], Yung-Ming Kuo[b], Pau-Choo Chung[b], E-Liang Chen[c]

[a]*Department of Information and Learning Technology, National University of Tainan, 33, Sec. 2, Shu-Lin St. Tainan 700, Taiwan*
[b]*Department of Electrical Engineering, National Cheng Kung University, Taiwan*
[c]*Department of Computer Science and Information Engineering, Leader University, Taiwan*

Received 18 January 2006; received in revised form 14 November 2006; accepted 16 November 2006

Abstract

The paper presents a new naked image detection algorithm. A learning-based chromatic distribution-matching scheme is proposed to determine the image's skin chroma distribution online such that it can tolerate the chromatic deviation coming from special lighting without increasing false alarm. The texture feature, namely coarseness, is used to acquire accurate skin segmentation. The low-level but reliable geometrical constraints and the mug shot exclusion procedure are employed to further examine the skin regions. Experimental results show our method can achieve satisfactory performance for detecting naked images under special lighting conditions.

Keywords: Skin detection; Naked image detection; Adult images; Chromatic distribution-matching scheme

1. Introduction

In a relatively short period of time, the Internet has become readily accessible in most organizations, schools and homes. Meanwhile, however, the problem of pornography through the Internet access in the workplace, at home and in education has considerably escalated. In the workplace, the pornography related access not only costs companies millions in non-business Internet activities, but it also has led to shattering business reputations and harassment cases. Being anonymous and often anarchic, images that would be illegal to sell even in adult bookstores can be easily transferred to home through the Internet, causing juveniles to see those obscene images intentionally or unintentionally. Therefore, how to effectively block or filter out pornography has been arousing a serious concern in related research areas.

The mostly used approach to blocking smut from the Internet is based on contextual keyword pattern matching technology that categorizes URLs by means of checking contexts of web pages and then traps the websites assorted as the obscene. Although this method can successfully filter out a mass of obscene websites, it is unable to deal with images, leading to its failure to detect those obscene web sites containing naked images instead of smut texts. Besides the threat coming from the web sites, a lot of the e-mail image attachments are naked. Hence, the development of naked image detection technology is urgently desired to prevent juveniles from getting access to pornographic contents from the Internet more thoroughly.

There is an extensive literature [1–6] on the detection of images by using features such as color histograms, texture measures and shape measures. On a whole, studies in this area consider whole image matches instead of semantic matches. However, naked image detection using semantic matches is a hard task because it has to deal with jointed objects of highly variable shapes, in a diverse range of poses, seen from many different views. Furthermore, lighting and background are uncontrolled, making segmentation very difficult. In Refs. [7,8] Forsyth and Fleck proposed an automatic system that marks skin-like pixels using color and texture properties to tell whether there are naked humans present in an image. These skin regions are then fed to a specialized grouper, which attempts to group a human figure using geometric constraints on human structure. If the grouper finds a predefined structure, the system decides a naked image is present. Jiao et al. [9] presented an adult

* Corresponding author. Tel.: +886 6 2606123x7717; fax: +886 6 2144409.
E-mail address: cslee@mail.nutn.edu.tw (J.-S. Lee).

0031-3203/$30.00
doi:10.1016/j.patcog.2006.11.016

image detection method. They first use the skin color model to detect naked skin areas roughly. Then the Sobel operator and the Gabor filter are applied to remove those non-skin pixels. Finally, the color coherence vector and the color histogram are employed to determine which image contains naked people. Wang et al. [10] combined an icon filter, a graph-photo detector, a color histogram filter, a texture filter and a wavelet-based shape-matching algorithm to detect the objectionable images. For the color histogram filter, they use a total of 512 bins to compute the histogram. Jones and Rehg [11] proposed a skin color detection technique by estimating the distribution of skin and non-skin color in the color space using labeled training data. Some simple features are extracted to detect adult images. Bosson et al. [12] presented a method to block the pornographic images. The authors compute the likelihood ratio for a quantized color space. Then, the blobs' features of the image, detected by likelihood histogram with 25^3 bins in RGB space, are computed and presented as a vector. Finally, the artificial neural network is utilized to classify the image is pornographic or not. Jedynak et al. [13] proposed a statistical model for skin detection. The maximum entropy model is used to infer the skin models from the data set. Then, the Bethe Tree approximation and Belief Propagation algorithm are utilized to approximate the probability for skin at pixel locations. Cao et al. [14] proposed a learning scheme for judging if there are any naked people in an image. The Learning Vector Quantization is first utilized to build several classifiers based on low-level features, such as color histogram, texture etc. Subsequently, these classifiers are combined to detect naked images. Zheng et al. [15] proposed a method to detect the adult images. The architecture is divided-into two parts. The skin detecting model, similar to [13], is applied to detect the skin blocks in the image first. Next, the features of skin blocks in the testing image are fed into the Multi-Layer Perceptron Classifier to identify that is an adult image or not.

As can be seen in these methods, none of them consider the inference coming from special lighting and color altering. There exist a large number of naked pictures taken under special lighting. 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. If the skin color model cannot tolerate the deviation, it will tend to miss a lot of naked pictures. On the contrary, if the skin color model accommodates the deviation, an abundance of non-skin objects like wood, desert sand, rock, foods, and the skin or fur of animals would be detected in the skin detection phase and deteriorates the system performance. Accordingly, the above mentioned approaches suffer from the skin color deviation resulting from special lighting, which is often seen in the naked images. Dealing with the special lighting effect in the naked images is a difficult task. If the skin color model tolerates the deviation, lots of non-skin objects would be detected simultaneously. A feasible solution for the problem is to adapt the adopted skin chroma distribution to the lighting of the input image. Based on this concept, a new naked image detection system is proposed. We develop a learning-based chromatic distribution-matching scheme that consists of the online sampling mechanism and the one-class-one-net neural network. Based on this approach, the object's chroma distribution can be online determined so that the skin color deviation coming from lighting can be accommodated without sacrificing the accuracy. The roughness feature is further applied to reject confusion coming from non-skin objects, so the skin area can be more effectively detected. Several representative features induced from the naked images are used to verify these skin areas. Subsequently, the face detection process is employed to filter out those false candidates coming from mug shots. The function blocks of the proposed naked image detection algorithm are demonstrated in Fig. 1.

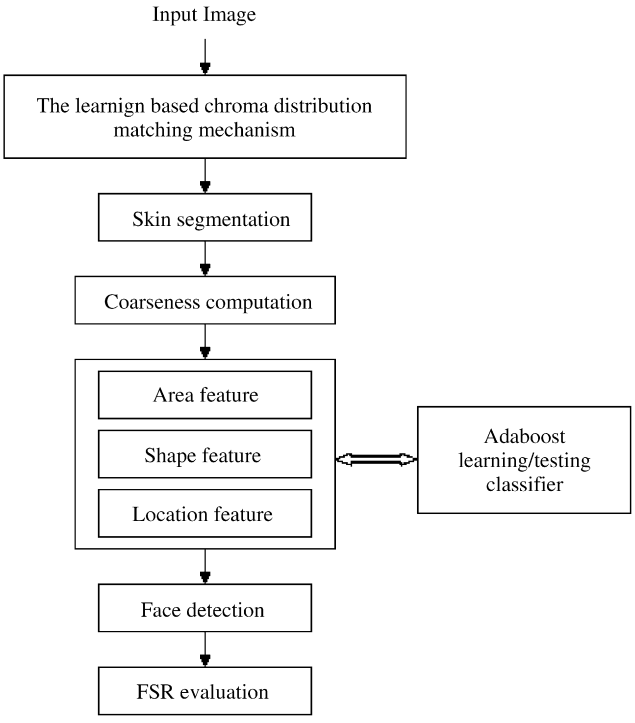

Fig. 1. The function block diagram of the proposed algorithm.

The paper is structured as follows. In Section 2, the adaptive skin segmentation procedure is introduced. The post-processing employed to enhance the system accuracy is presented in Section 3. Section 4 shows the naked image detection process. Section 5 demonstrates the experimental results. Finally, we make conclusions in Section 6.

2. Adaptive skin segmentation

The skin tone is formed by the interaction between skin and light. Therefore, the captured skin color in an image depends on the surrounding light in addition to the intrinsic skin tone. To make naked images look more attractive, photographers usually apply special lighting, thereby altering the chroma distribution of the skin tone. Hence, if the referenced skin chroma distribution is gathered from the normal conditions beforehand, the corresponding skin detection performance will be dramatically degenerated. A popular skin color model proposed by Sobottka and Pitas [16] defines the fleshtone by

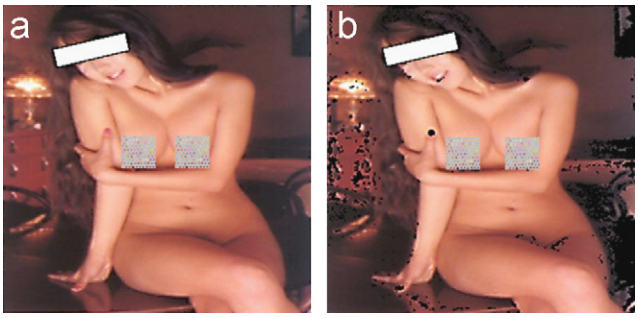

Fig. 2. (a) A special lighting image. (b) The skin segmentation result by applying the flesh tone model proposed by Sobottka and Pitas [16].

ranges in hue and saturation. Fig. 2 demonstrates that the fleshtone model is not capable of successful detecting skin under the special lighting conditions. A practicable approach to overcoming this predicament is to adaptively determine the corresponding skin chroma distribution from the input image itself.

In the past, a lot of skin color detection approaches have been studied. In Ref. [17], skin color was modeled based on a reflectance model of the skin, knowledge about the camera parameters, and the spectrum of the light source. It has application in skin color segmentation and may have application in the estimation of the color temperature in camera image containing skin color. Brand and Mason [18] assessed the merits of three different approaches to pixel-level human skin detection. The first two approaches used simple ratios and color space transforms, respectively, whereas the third was a numerically efficient approach based on a 3D RGB probability map. The lowest of these false acceptance rates is found to be about 20% given by the 3D probability map. Albiol et al. [19] showed that for every color space there exists an optimum skin detector scheme such that the performance of all these schemes is the same. Gomez [20] used a data analysis approach for selecting color components for skin detection. After evaluating each component of several color models, he found that a mixture of components can cope well with such requirements. Jayaram et al. [21] evaluated the effect of three steps on the skin detection performance, i.e. color transformation, illumination component dropping, and skin color distribution modeling. The importance of this study is proposing a new comprehensive color space and color modeling testing methodology. Sigal et al. [22] proposed a real-time skin segmentation method in video sequences. An explicit second order Markov model was used to predict evolution of the skin-color histogram over time. Histograms were dynamically updated based on feedback from the current segmentation and predictions of the Markov model. In these approaches, only the methods in Refs. [17,22] adopted adaptive skin chroma distributions concept. Nevertheless, Ref. [17] suffers from the knowledge about the camera parameters that are usually absent in the naked image detection application. As to Ref. [22], the present skin chroma distribution is obtained by updating the one in the previous frame. These distributions are temporally correlated such that this approach cannot be applied to the naked image detection task either.

Adapting the skin chroma distribution to lighting is difficult if without any priori information. Here, we propose an adaptive scheme, which can adapt the skin color model to the corresponding lighting condition of the input naked image. From the viewpoint of photography, the shot subject should be positioned around the center of an image. Based on this property, we propose an on-line sampling strategy, i.e. the skin chroma distribution of the input image is decided by directly sampling around the image center for skin areas. Subsequently, a neural network is used to determine the corresponding chroma distribution that is then applied to detect the skin area.

Selecting an appropriate color space is required for modeling skin color. In Ref. [23], Vezhnevets et al. surveyed the pixel-based skin color detection techniques. They pointed out that the performance of parametric skin classifiers depends heavily on the color space choice that can be observed by the results in Refs. [24,25]. On the contrary, the non-parametric methods are almost independent to the color space choice [19,26,27]. The proposed skin classifier belongs to the non-parametric one hence the color space selection is not so critical. For considering what color space should be adopted, we aim at two factors i.e. the computational cost and the compactness of the skin chroma distribution. Chai and Ngan [28,29] and Phung et al. [30] in their studies pointed out that pixels belonging to skin region exhibit similar *Cb* and *Cr* values. And, the skin color model based on *Cb* and *Cr* values can provide good coverage of all human races. Despite their different appearances, these color types belong to the same small cluster in *Cb–Cr* plane. The apparent difference in skin colors perceived by viewers mainly comes from the darkness or fairness of the skin. These features are reflected on the difference in the brightness of the color, which is governed by *Y* component rather than *Cb* and *Cr* components. The removal of luminance component helps constructing skin classifier that will work well for images with different skin brightness. Also, the reduction of space dimensionality profits decreasing the computational complexity. Because it provides an effective separation into luminance and chrominance channel and generates a compact skin chroma distribution, we adopt the *Cb–Cr* space as the skin color space. We classify the popular skin colors of naked images, including various lighting conditions, into several categories. The corresponding compactly chroma histogram of each category is compiled in advance. The existing skin detection methods usually generate the corresponding chroma histogram by straightforward compiling the whole skin samples. If we follow that approach, we need to manually categorize the skin samples in advance. It is difficult to do accurate categorization because of limited chromatic sensitivity and subjective bias. To avoid this drawback, we design a systematic approach to the automatic and objective classification of the whole skin samples into several categories. This approach consists of two major steps: (1) merging and (2) grouping.

(1) At the merging step, every sample skin image is first transformed from the RGB color space to the *YCbCr* space by

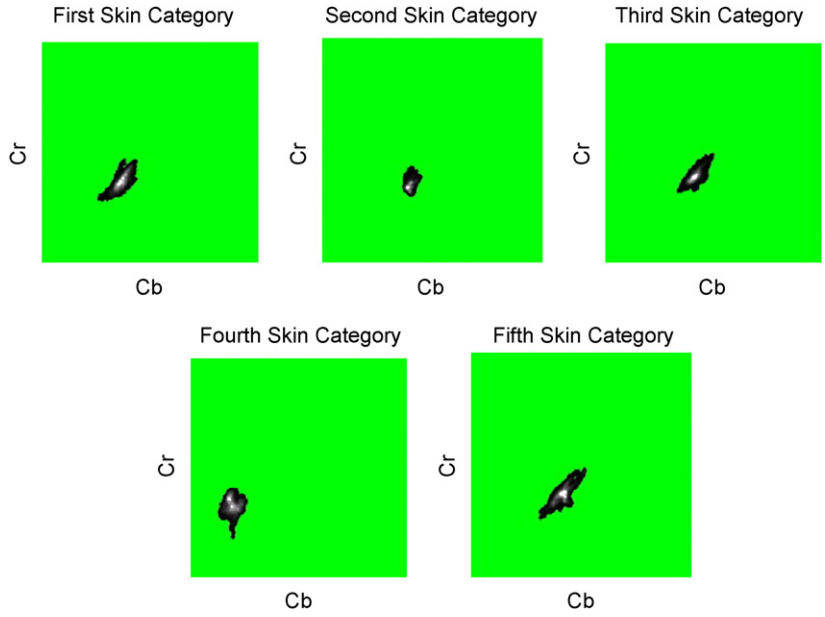

Fig. 3. The five major skin chroma clusters. The intensity represents the appearance frequency of a specific chroma.

using the following formula:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.533 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -84.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (1)$$

Each transformed result forms a chroma histogram denoted as CH. Then, dichotomizing the CH followed by the closing and the filling operations to get the solid skin chroma region, denoted as CR. For two CHs, if the overlapping part exceeds 80% of the individual CH, they are merged into a larger CH. For more explicit expression, let CH_1 and CH_2 represent the two CHs, CR_1 and CR_2 denote the corresponding CRs, and Z represent the overlapping area, i.e. $CR_1 \cap CR_2$. Once the two CHs satisfy both Eqs. (2) and (3), they are merged into a single CH.

$$\sum_{Cb\in Z}\sum_{Cr\in Z} CH_1(Cb,Cr) \Big/ \sum_{Cb\in CR_1}\sum_{Cr\in CR_1} CH_1(Cb,Cr) \geqslant 0.8, \quad (2)$$

$$\sum_{Cb\in Z}\sum_{Cr\in Z} CH_2(Cb,Cr) \Big/ \sum_{Cb\in CR_2}\sum_{Cr\in CR_2} CH_2(Cb,Cr) \geqslant 0.8. \quad (3)$$

(2) After the merging step, the merged CHs can be further merged again. For example, if CH_1 and CH_2 are merged into $CH_{1,2}$ and CH_2 and CH_3 are merged into $CH_{2,3}$, then $CH_{1,2}$ and $CH_{2,3}$ can be further merged into $CH_{1,2,3}$. Repeat this process until no more mergence occurs. Finally, we acquire five major skin chroma clusters as shown in Fig. 3. The intensity in Fig. 3 represents the appearance frequency of a specific chroma and can be viewed as the corresponding probability if normalization is performed. That is, each cluster represents the corresponding chroma histogram for a specific skin category.

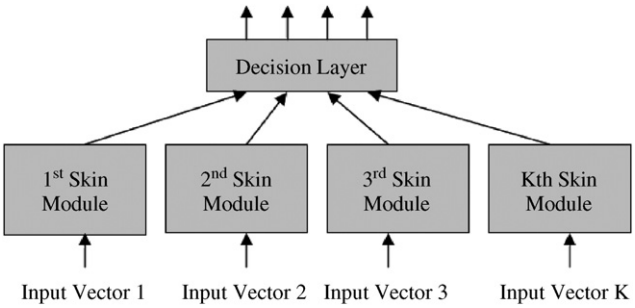

Fig. 4. The learning-based chroma distribution matching mechanism comprises K multilayer feed-forward neural networks.

The chroma histogram is denoted as H_x, where the subscript x means the xth skin chroma cluster. The corresponding element at position (i, j) is represented as $H_x^{(i,j)}$. The learning-based chroma distribution matching mechanism comprises a multilayer feed-forward neural network used to learn and judge whether the input image contains skin, as shown in Fig. 4. Each skin module, as illustrated in Fig. 5, consists of a three-layer feed-forward neural network and is responsible for learning and classifying the corresponding skin chroma. The advantage of this architecture is that the system can adjust the number of skin modules to meet the practical requirement. For example,

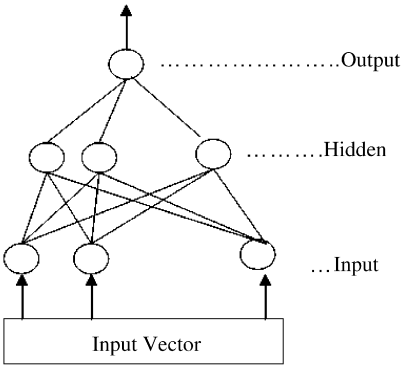

Fig. 5. The architecture of a skin module.

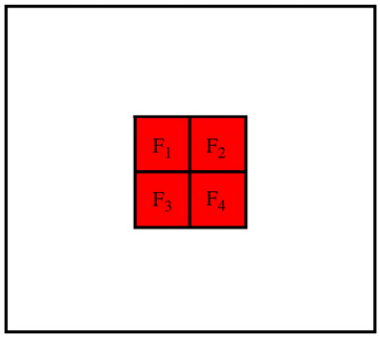

Fig. 6. The foreground blocks chosen for the learning phase.

once a new skin chroma, not included in the existing skin modules, frequently appears in the naked images we can establish the corresponding new skin module easily.

The macro block is the basic unit for matching mechanism. The macro block size depends on the size of the input image. For the real naked images, the sizes are not uniform and usually larger than 256 by 256. To save computational time, the input image is first normalized before further processing. Let the width and the height of the input image is W and H, respectively. The bigger one is scaled to 256 and the other dimension is scaled proportionably to retain the aspect ratio. The normalized image is divided into 16 by 16 blocks. Each block is called a macro block. The matching mechanism can be divided into the training phase and the testing phase. In the training phase, the K skin modules are trained, respectively. For each skin module, both the positive and the negative training samples are interlaced fed into the network until convergence. The input vector is composed of the foreground feature vector derived from the foreground blocks (FB). And, four macro blocks construct these foreground blocks. Each macro block comprises M by N pixels. To ensure the training samples can be captured from the right places, the configuration of the macro blocks for the positive training sample is distinct from that for the negative training sample. In our investigation, the skin region in a naked image is usually so close to the image center that it can easily catch the viewer's eye. Hence, for the case of positive training, a central area (CA) with size $4M$ by $4N$ pixels is first selected from the training image. In other words, the CA comprises 16 macro blocks arranged to a macro block array with four rows by four columns. The larger number of macro block is used to ensure the coverage of the skin region. Four most representative macro blocks are chosen from the 16 blocks. For example, we want to train the xth skin module. A positive sample image P belonging to the xth skin chroma cluster is fed to train this module. First, the CA is picked out from P. Let $MB^{(i,j)}$ denote the macro block at the intersection of the ith raw and the jth column macro block in CA. Let $MB^{(i,j)}_{(u,v)}$ represent the position (u, v) of $MB^{(i,j)}$. The (Cb, Cr) components at $MB^{(i,j)}_{(u,v)}$ are denoted as $Cb_{MB^{(i,j)}_{(u,v)}}$ and $Cr_{MB^{(i,j)}_{(u,v)}}$, respectively. Let $MS^{(i,j)}_x$ mean the membership strength of $MB^{(i,j)}$ to the chroma histogram H_x and is defined as

$$MS^{(i,j)}_x = \sum_{u=1}^{M}\sum_{v=1}^{N} H_x^{(Cb_{MB^{(i,j)}_{(u,v)}}, Cr_{MB^{(i,j)}_{(u,v)}})}. \quad (4)$$

The four macro-blocks with the top four biggest values of $\{MS^{(i,j)}_x | 1 \leqslant i, j \leqslant 16, i, j \in N\}$ are chosen as the FB. As to the case of negative training, a central area CA with size $2M$ by $2N$ pixels, i.e. four macro blocks, is selected as the FB. Fig. 6 is illustrated to explain how to get the K input vectors from the FB. Let F_i, where $i = 1, 2, 3, 4$, denote the chroma distribution of each macro block of FB. By accumulating those F_i, we can get the chroma distribution of the FB, that is, F.

$$F = \sum_{i=1}^{4} F_i. \quad (5)$$

The input feature is defined as the product between the pre-categorized skin chroma distribution H_x and F. Assume P_x represents the product between F and H_x. Then, we can get

$$P_x = F \times H_x, \quad (6)$$

where '$\times$' means the element-to-element product. Because the size of P_x is 256×256, it is not practical for neural network training. We reduce P_x to $\tilde{P}_x$ by down sampling, i.e. the original 32 by 32 elements are summed to generate a new element. Then, $\tilde{P}_x$ is treated as the input vector for the xth skin module.

In the testing phase, the uncertainty of the skin region location is higher than that in the training phase. To increase the hit rate of capturing the skin region, the online sampling strategy is proposed. We utilize five FBs arranged as Fig. 7 to sample the skin region. Every FB generates K input vectors. These vectors are fed into the corresponding skin modules. Subsequently, the outputs of the K skin modules are fed into the decision layer. The decision layer in Fig. 8 is a competitive network, with function

$$Y_k = \begin{cases} f_k & \text{if } f_k \geqslant f_j \text{ and } f_k > 0.5, \ \forall j \neq k, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

If $Y_k, k \in \{1, 2, 3, \ldots, K\}$, is the winner, it means that the skin chroma distribution of the test image may belong to the kth skin category. Because there are five FBs, the above procedures are executed five times. If there is no winner, it implies that

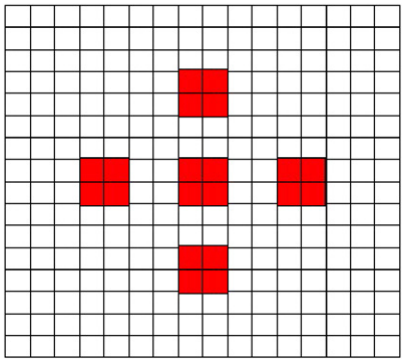

Fig. 7. The foreground blocks chosen for the testing phase.

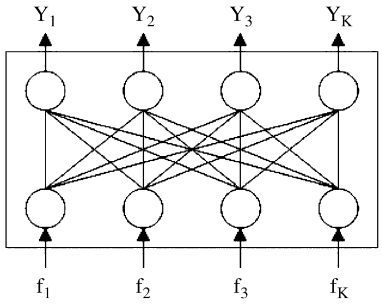

Fig. 8. The decision layer is a competitive network.

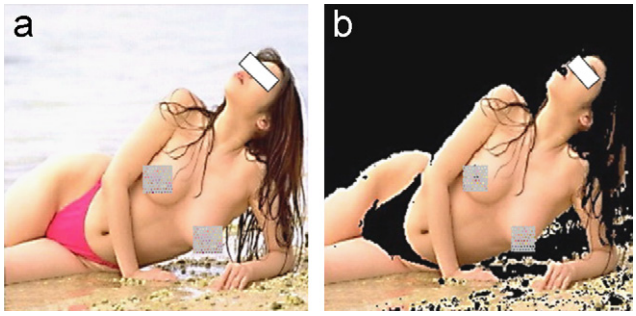

Fig. 9. (a) A real image used to illustrate the process effect. (b) The skin segmentation result of (a) by using our adaptive skin segmentation method.

the input image does not contain naked people. If the winner is not unique, further competition is needed. Assume Y_a and Y_b are the winners generated by the m and n FB, respectively. Then, the bigger one is viewed as the real winner. Assuming Y_a is the winner, the test image is then segmented by setting the pixels with chroma belonging to the ath skin chroma cluster to one and others to zero. And, the acquired binary image, i.e. skin image, is denoted as S. A real image in Fig. 9(a) is used to illustrate the process effect. The skin segmentation result is shown in Fig. 9(b). We can find that the skin region is successfully detected. As to the special lighting case shown in Fig. 2(a), our adaptive skin segmentation method can also successfully detect the skin area shown in Fig. 10. This means that our adaptive skin segmentation method can effectively deal with the special lighting conditions.

Fig. 10. The skin segmentation result of Fig. 2(a) by using our adaptive skin segmentation method.

3. The post-processing

The performance of the naked image detection is extremely dependent on the accurate skin segmentation. There exist a lot of objects, possessing skin like chroma in an image, e.g. wood, foods, rock, desert sand, and animal fur. It is almost impossible to distinguish them from the human skin solely by using the chroma property. According to our observation, smoothness is a very important feature for skin. Therefore, we utilize the roughness feature to further reject the confusion coming from the non-skin objects with skin like chroma. There exist many approaches to measuring the texture roughness in image processing, e.g. Fourier power spectrum [31], fractal dimension [32], co-occurrence gray-level matrix [33], neighborhood gray-tone difference matrix [34], auto-covariance function and edge density [35] etc. Intuitively, the rougher the object's surface the more its extreme number in image intensity. Based on the concern about computational complexity, the window based extreme density approach is proposed to quantify the roughness. First, the input image is transformed to an intensity image I. And, we create an extreme image X with the same size of I and each element is set to zero. A pixel $I(x, y)$ of I at position (x, y), if it is an extreme, i.e. $I(x, y) > I(x+\Delta x, y+\Delta y)$ where both Δx and $\Delta y \in \{-1, 0, 1\}$ and $|\Delta x| + |\Delta y| \neq 0$, $X(x, y)$ is set to one. Subsequently, the X is divided into 8×8 blocks. If the extreme number of a block is less than a threshold value t_e, that block is set to one. Otherwise, that block is reset to zero. Therefore, the zero-block in X represents the rough region, whereas the one-block in X corresponds to the smooth region. The coarseness image corresponding to Fig. 9(a) is demonstrated in Fig. 11(a). The smooth skin region SS can be obtained by multiplying S and X.

4. Naked image detection

We induce three common properties for naked images through investigating a mass of naked images. First, for delighting viewers, the naked body usually occupies a significant portion of an image. Second, the aspect ratio of the naked body is usually in a reasonable range. Third, the position of the naked body is close to the image center to harmonize with the frame. The features based on these properties are

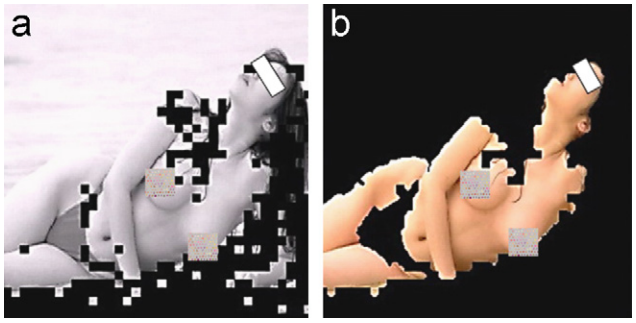

Fig. 11. (a) The coarseness image corresponding to Fig. 9(a). (b) The maximum object corresponding to Fig. 9(a).

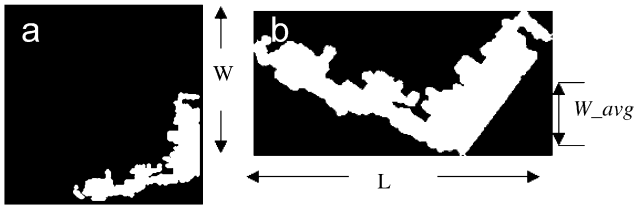

Fig. 12. (a) An *MO* used to explain how to quantify the shape feature. (b) The minimum rectangle containing the maximum object in (a).

extracted to judge if the smooth skin region *SS* contains naked bodies.

4.1. Feature 1. Area

The area feature is used to quantify the size of the examined area. To further separate the body from the skin like background in the *SS* we apply the opening operation to the *SS* and get the outcome *OSS*. We locate the body area by picking out the maximum object, denoted as *MO*, from the *OSS*. The area of *MO* is denoted as *AMO*. The occupation ratio of *MO*, i.e. $AMO/Area(I)$, represented as *OR* is viewed as the first feature. The maximum object area *MO* corresponding to Fig. 9(a) is shown in Fig. 11(b).

4.2. Feature 2. Shape

The shape feature is used to quantify how the profile of the examined area is like the naked body. For the convenience of explanation we use two figures as examples. Fig. 12(a) shows an *MO*. By applying the principal component analysis (PCA) to *MO*, we can obtain two orthogonal eigenvectors. By projecting *MO* to the two eigenvectors, respectively, we can obtain a minimum rectangle *R* containing *MO*. Assume the lengths of the two sides of the rectangle are *L* and *W* (and $L \geqq W$). If we directly adopt the ratio of *L* and *W* as the aspect ratio, an elongate and bended area may correspond to a significant aspect ratio like the case in Fig. 12(b). To avoid this problem, we take the equivalent aspect ratio as the shape feature. The equivalent aspect ratio *Q* is defined as the ratio of *L* to the equivalent width W_{eq}.

$$Q = L/W_{eq}, \tag{8}$$

$$W_{eq} = AMO/L. \tag{9}$$

4.3. Feature 3. Location

The location feature is used to quantify if the examined area is close to the image center. The normalized horizontal and vertical distances from the gravity center (*GC*) of *MO* to the image center (*IC*) are adopted as the location features *NHD* and *NVD*, respectively. The normalization is used because the naked image's size is not always square. The location features are defined as the following equations.

$$NHD = |IC_x - GC_x|/IW, \tag{10}$$

$$NVD = |IC_y - GC_y|/IL, \tag{11}$$

$$GC_x = \frac{\sum_{(x_i,y_i)\in MO} x_i}{AMO}, \tag{12}$$

$$GC_y = \frac{\sum_{(x_i,y_i)\in MO} y_i}{AMO}, \tag{13}$$

$$(IC_x, IC_y) = \left(\frac{IW}{2}, \frac{IH}{2}\right), \tag{14}$$

where *IW* and *IH* represent the width and the height of the input image, respectively. Subsequently, these features are applied to judge if the *MO* corresponds to the naked body. However, the effective combination of these features into one classifier must be taken into serious account. Boosting is a powerful method to combine a collection of simple classification functions to form a stronger one. The smart combination often outperforms most monolithic strong classifiers such as SVMs and Neural Networks. Adaboost, making the simple classification functions work, is a kind of large margin classifiers [36,37]. It is utilized to serve as the feature classifier in our work. If the *MO* is classified as no naked bodies, our system will conclude that the examined image is not a naked image. On the contrary, if the *MO* is classified as a naked body, further examination is needed because most mug shots also meet the above properties of a naked body. The most important feature for a mug shot is the face area occupies a significant portion of the skin area. So we utilize the face-to-skin ratio (*FSR*) to exclude the false positives coming from mug shots. There exist a lot of face detection algorithms in the literature. For practical applications, we need a face detection algorithm in which faces can be detected effectively as well as efficiently. Recently, Lienhart et al. [38] proposed a new face detection algorithm that minimizes computation time while achieving high detection accuracy. We apply this approach to detecting faces in the so-called *ROI*, obtained by referring to the minimal rectangle containing the *MO*. The *FSR* is derived from calculating the ratio of the predicted face area to the *AMO*. If the *FSR* is larger than a threshold value, the test image is viewed as a mug shot. Otherwise, the test image is considered a naked image. In terms of computational complexity, the face detection stage indeed increases the

system's computation load. However, it can effectively filter out those false positives resulting from mug shots in the classification stage. In addition, the face detection routine is invoked only when the *MO* is classified as a naked body that only occupies a little portion in real applications. Therefore, the face detection stage is necessary and worthwhile when the system robustness and complexity are both taken into consideration.

5. Experimental results

The training images and the testing ones are collected from the Internet and the album. There are 508 naked images, including 312 fully naked and upper-body naked images and 196 intercourse images. The former are categorized as naked-1 images while the latter are categorized as naked-2 images. There are 482 non-naked images, including 127 clothed people and people wearing swimming suits, 41 mug shots and 314 other miscellaneous images (nature scenes, buildings, wood, foods, rock, desert sand, and animals), categorized as non-naked-1 images, non-naked-2 images, and non-naked-3 images, respectively. The naked roles comprise Caucasians, Blacks and Asians. The training set contains 50 naked images and 40 non-naked images.

A serial of experiments are conducted to test the performance of each stage in our system, including the adaptive skin segmentation, the smooth skin detection, the geometric assessment and the mug shot exclusion. First, we test the effectiveness of the adaptive skin segmentation approach. Each image is decided if contain skin color object and which color category it belongs to by the learning-based chroma distribution matching mechanism. The segmentation results are evaluated by the ratio of the skin area to the whole image area. The first part of Table 1 shows the rate that every image category can be correctly classified by using the adaptive skin segmentation outcomes only. We can find that both the two naked categories can be well recognized. However, those non-naked images fail to be well recognized except for the non-naked-3 category. After investigating those failure cases, we find that those missed naked images correspond to bodies with shadow or specular light. When the body intensity is too dark or too bright, the corresponding skin chroma will move away from the actual chroma so that it is missed in the chroma distribution matching process. As those missed non-naked images correspond to skin-tone scenes or objects such as sunset scenes, wood, rock, desert sand and animals, proceeding to the following stage can further retrieve them. Subsequently, we examine the efficacy of the smooth skin detection stage. In this stage, the smoothness constraint is employed to the adaptive skin segmentation outcomes to reduce the false alarms coming from the skin-tone scenes and objects. The second part of Table 1 shows the correct classification rate for each category by utilizing the smooth skin detection results. Those non-naked categories can be recognized better. Unavoidably, however, the performance of the naked categories is sacrificed a little for improving the false positives. The third stage performs geometric verification, including size, shape and location. First, the related threshold values of these geometric parameters are determined by empiricism. The corresponding classification results are demonstrated in the third part of Table 1. We can find that the false positives improve a little, whereas the false negatives worsen a lot. Contrasting with the empirical approach, the fourth part of Table 1 shows the classification outcomes employing the Adaboost learning algorithm to determine the corresponding optimal threshold values. We find that the false positives improve a lot except for the non-naked-2 category. It is worth noting that the false negatives almost keep at the same level at the same time. This means that these geometric features indeed can effectively distinguish the naked images from the non-naked ones. The final stage filters out the mug shots by applying the *FSR*. The classification results are depicted in the last part of Table 1. The correct classification rates of the naked images do not decrease in this process. And, the correct classification rate of the non-naked-2 category dramatically improves. The confusion matrix of the

Table 1
The effectiveness of different stages

	Naked-1	Naked-2	Non-naked-1	Non-naked-2	Non-naked-3
Image no.	312	196	127	41	314
The effectiveness of the adaptive skin segmentation stage					
Correct classification rate (%)	91.7	85.2	69.3	43.9	86.3
The effectiveness of the smooth skin detection stage					
Correct classification rate (%)	88.1	84.2	77.2	48.8	88.2
The effectiveness of the geometric verification stage using empirical approach to determining the corresponding threshold values					
Correct classification rate (%)	84.3	78.6	80.3	48.8	93.9
The effectiveness of the geometric verification stage employing the adaboost learning algorithm to determine the corresponding optimal threshold values					
Correct classification rate (%)	87.8	84.2	85.8	48.8	96.8
The effectiveness of the mug shot exclusion stage					
Correct classification rate (%)	87.8	84.2	89.8	95.1	96.8

Table 2
The confusion matrix of the proposed method

Classification results	Test images	
	Naked image	Non-naked image
Naked image	439 (86.4%)	25 (5.2%)
Non-naked image	69 (13.6%)	457 (94.8%)

Table 3
The confusion matrix of the compared method

Classification results	Test images	
	Naked image	Non-naked image
Naked image	364 (71.7%)	76 (15.8%)
Non-naked image	144 (28.3%)	406 (84.2%)

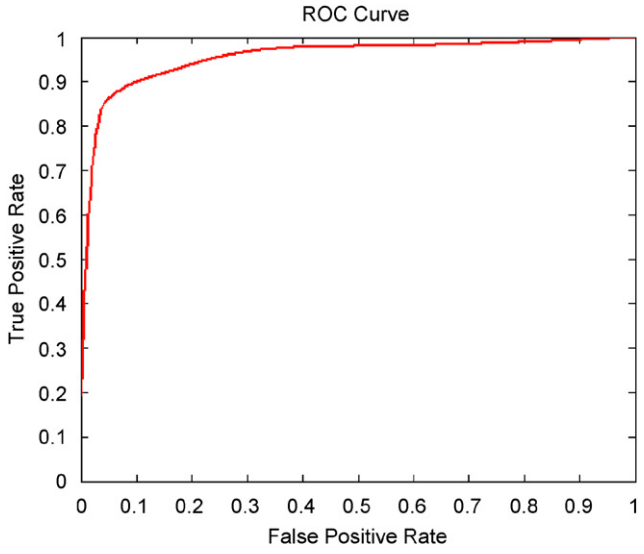

Fig. 13. The ROC curve of the proposed system.

experimental results is shown in Table 2. The detection rates of the naked images and the non-naked images are 86.4% and 94.8%, respectively. The receiver operating characteristics (ROC) curve of the proposed system is shown in Fig. 13. It is noted that the results meet the requirements for the practical application of a naked image detection system, i.e. the moderate detection rate and low false alarm. From these experiments, our method proves to be able to correctly online determine the skin chroma and effectively distinguish naked images from non-naked images by integrating texture, geometric features and face detection.

To test the superior performance of our method, an existing method proposed by Cao et al. [14] was implemented to compare with ours. This choice was made on the basis of fair and objective considerations: it also adopts *YCbCr* color space and integrates features by utilizing the Adaboost learning algorithm. The same test images are employed to perform comparison. The confusion matrix of the compared method is shown in Table 3. The detection rates of the naked images and the non-naked images are 71.7% and 84.2%, respectively. These results show our method can achieve superior performance for naked images detection.

6. Conclusions

Because of the popularization of the Internet access, governments and parents have been facing a serious problem that juveniles can access naked images from the Internet intentionally or unintentionally in effortless ways. Thus, effectively blocking or filtering out pornography has become a critical issue. In this paper, a new naked image detection algorithm is proposed. We develop a learning-based chromatic distribution-matching scheme that consists of the online sampling mechanism and the one-class-one-net neural network. Based on this scheme, we can deal with the chromatic deviation coming from special lighting without increasing false alarm. The utilization of the roughness feature further rejects confusion coming from flesh tone objects such that our skin detection method can get not only high detection rate but also high detection precision. Low-level while reliable geometric constraints of naked bodies are effectively combined by utilizing the Adaboost algorithm. Finally, the mug shot exclusion process makes the overall system achieve satisfactory detection performance. Experiments show our method can correctly online determine the skin chroma and effectively distinguish naked images from non-naked images by integrating texture, geometric features and face detection. Because of the moderate detection rate and low false alarm, our method lends itself to the practical application of naked image detection.

Acknowledgements

Authors would like to express their deep appreciation for the support of the National Science Council under Grant number NSC-92-2213-E-150-022.

References

[1] S. Aksoy, K. Koperski, C. Tusk, G. Marchisio, J.C. Tilton, Learning Bayesian classifiers for scene classification with a visual grammar, IEEE Trans. Geosci. Remote Sensing 43 (3) (2005) 581–589.
[2] J. Luo, M. Boutell, R.T. Gray, C. Brown, Image transform bootstrapping and its applications to semantic scene classification systems, IEEE Trans. Man Cybern. Part B 35 (3) (2005) 563–570.
[3] S. Liapis, G. Tziritas, Color and texture image retrieval using chromaticity histograms and wavelet frames, IEEE Trans. Multimedia 6 (5) (2004) 676–686.
[4] J.-W. Hsieh, W.E.L. Grimson, Spatial template extraction for image retrieval by region matching, IEEE Trans. Image Process. 12 (11) (2003) 1404–1415.
[5] T.P. Minka, R.W. Picard, Interactive learning with a society of models, Pattern Recognition 30 (1997) 465–481.
[6] J. Han, K.N. Ngan, L. Mingjing, H.-J. Zhang, A memory learning framework for effective image retrieval, IEEE Trans. Image Process. 14 (4) (2005) 511–524.
[7] D.A. Forsyth, M.M. Fleck, Body plans, in: Conference on Computer Vision and Pattern Recognition, 1997, pp. 678–683.
[8] D. Forsyth, M. Fleck, Automatic detection of human nudes, Int. J. Comput. Vision 32 (1) (1999) 63–77.

[9] F. Jiao, W. Gao, L. Duan, G. Cui, Detecting adult image using multiple features, in: Proceedings of IEEE International Conference on Info-tech and Info-net, vol. 3, 2001, pp. 378–383.
[10] J.Z. Wang, J. Li, G. Wiederhold, O. Firschein, System for screening objectionable images, Comput. Commun. 15 (21) (1998) 1355–1360.
[11] M.J. Jones, J.M. Rehg, Statistical color models with application to skin detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, 1999, pp. 274–280.
[12] A. Bosson, G.C. Cawley, Y. Chian, R. Harvey, Non-retrieval: blocking pornographic images, in: Proceedings of the International Conference on the Challenge of Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, Springer, London, 2002, pp. 50–60.
[13] B. Jedynak, H. Zheng, M. Daoudi, Statistical models for skin detection, IEEE Workshop on Statistical Analysis in Computer Vision, in conjunction with CVPR 2003 Madison, Wisconsin, June 16–22, 2003.
[14] L.L. Cao, X.L. Li, N.H. Yu, Z.K. Liu, Naked people retrieval based on adaboost learning, in: Proceedings of the First International Conference on Machine Learning and Cybernetics, 2002, pp. 1133–1138.
[15] H. Zheng, M. Daoudi, B. Jedynak, Blocking adult images based on statistical skin, Electron. Lett. Comput. Vision Image Anal. 4 (2) (2004) 1–14.
[16] K. Sobottka, I. Pitas, Looking for faces and facial features in color images, Pattern Recognition Image Anal. 7 (1) (1997).
[17] M. Storring, H.J. Andersen, E. Granum, Skin colour detection under changing lighting conditions, in: Seventh Symposium on Intelligent Robotics Systems, 1999, pp. 187–195.
[18] J. Brand, J.S. Mason, A comparative assessment of three approaches to pixel-level human skin-detection, in: Proceedings of the 15th International Conference on Pattern Recognition, 2000, pp. 1056–1059.
[19] A. Albiol, L. Torres, E.J. Delp, Optimum color spaces for skin detection, in: Proceedings of the IEEE International Conference on Image Processing, 2001, pp. 122–124.
[20] G. Gomez, On selecting colour components for skin detection, in: Proceedings of 16th International Conference on Pattern Recognition, vol. 2, 2002, pp. 961–964.
[21] S. Jayaram, S. Schmugge, M.C. Shin, L.V. Tsap, Effect of color space transformation, the illuminance component, and color modeling on skin detection, in: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2004, pp. II-813–II-818.
[22] L. Sigal, S. Sclaroff, V. Athitsos, Skin color-based video segmentation under time-varying illumination, IEEE Trans. Pattern Anal. Mach. Intell. 26 (7) (2004) 863–877.
[23] V. Vezhnevets, V. Sazonov, A. Andreeva, A survey on pixel-based skin color detection techniques, in: Proceedings of the International Conference on Computer Graphics, 2003, pp. 85–92.
[24] J.C. Terrillon, M.N. Shirazi, H. Fukamachi, S. Akamatsu, Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images, in: Proceedings of the IEEE International Conference on Face and Gesture Recognition, 2000, pp. 54–61.
[25] J.Y. Lee, S.I. Yoo, An elliptical boundary model for skin color detection, in: Proceedings of the 2002 International Conference on Imaging Science, Systems, and Technology, 2002.
[26] B.D. Zarit, B.J. Super, F.K.H. Quek, Comparison of five color models in skin pixel classification, in: International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems, 1999, pp. 58–63.
[27] D. Brown, I. Craw, J. Lewthwaite, A SOM based approach to skin detection with application in real time systems, in: Proceedings of the British Machine Vision Conference, 2001.
[28] D. Chai, K.N. Ngan, Locating facial region of a head-and-shoulders color image, in: Proceedings of Third IEEE International Conference on Automatic Face and Gesture Recognition (FG '98), Nara, Japan, April 1998, pp. 124–129.
[29] D. Chai, K.N. Ngan, Face segmentation using skin-color map in videophone applications, IEEE Trans. Circuits Syst. for Video Technol. 9 (4) (1999) 551–564.
[30] S.L. Phung, A. Bouzerdoum, D. Chai, A novel skin color model in YCBCR color space and its application to human face detection, in: Proceedings of IEEE International Conference on Image Processing, vol. I, 2002, pp. 289–292.
[31] K.R. Castleman, Digital Image Processing, Prentice-Hall, Englewood Cliffs, NJ, 1996.
[32] B. Chaudhuri, N. Sarkar, Texture segmentation using fractal dimension, IEEE Trans. Pattern Anal. Mach. Intell. 17 (1) (1995) 72–77.
[33] Haralick et al., Texture features for image classification, IEEE Trans. SMC (3) (1973) 610–621.
[34] M. Amadasun, R. King, Textural features corresponding to textural properties, IEEE Trans. SMC 19 (5) (1989) 1264–1274.
[35] F. Tomita, S. Tsuji, Computer Analysis of Visual Textures, Kluwer, Norwell, MA, 1990.
[36] K. Tieu, P. Viola, Boosting image retrieval, in: IEEE Conference on ICCV, 2000.
[37] G.D. Guo, H.J. Zhang, S.Z. Li, Pairwise face recognition, in: IEEE Conference on ICCV, Vancouver, Canada, 2001.
[38] R. Lienhart, A. Kuranov, V. Pisarevsky, Empirical analysis of detection cascades of boosted classifiers for rapid object detection, Technical Report, MRL, Intel Labs, 2002.

About the Author—JIANN-SHU LEE received his Ph.D. in electrical engineering from National Cheng Kung University. His experience includes assistant researcher at the Institute for Information Industry (III), and management experience at the information center at Chung Shan Medical University and in the Department of Computer Science and Information Engineering at Da-Yeh University. He is specialized in digital content analysis, image processing, medical image processing, watermarking and e-learning and he is also the author of many academic articles, including publications in journals such as IEEE_Transaction on Image Processing and Pattern Recognition. He has served as section chairs in many domestic and international conferences. He has also received numerous awards and certificates from organizations such as Cisco, Acer, and Xerox. Currently, he is an associate professor at the Department of Information and Learning Technology, National University of Tainan. And, he is also the academic director of the Biomedical Engineering Society of the ROC.

About the Author—YUNG-MING KUO received the B.S. degree in Computer Science and Information Engineering from Fu Jen Catholic University, Taiwan, in 1999, and the M.S. degree in Electrical Engineering from National Cheng Kung University, Taiwan, in 2001. He is currently a Ph.D. candidate in the Department of Electrical Engineering, National Cheng Kung University, Taiwan. His research interests include image processing and pattern recognition, video-based behavior analysis and neural networks.

About the Author—PAU-CHOO CHUNG received the B.S. and M.S. degrees in electrical engineering from National Cheng Kung University, Taiwan, Republic of China, in 1981 and 1983, respectively, and the Ph.D. degree in electrical engineering from Texas Tech University in 1991. In 1991, she joined the Department of Electrical Engineering, National Cheng Kung University, and has become a full professor since 1996. Since 2001, she has served as the Vice Director, and currently the Director, of the Center for Research of E-life Digital Technology, National Cheng Kung University. She was selected as Distinguished Professor of National Cheng Kung University, 2005. She is also serving as the Director of Electrical Laboratory, National Cheng Kung University.

About the Author—E-LIANG CHEN received the Ph.D. degree from the Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan, in 1999. He is an assistant professor in the Computer Science and Information Engineering Department of Leader University, Tainan, Taiwan. His current research includes face recognition and applications of neural networks.