Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 2190–2200
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Face detection in complicated backgrounds and different illumination conditions by using YCbCr color space and neural network

Chiunhsiun Lin *

National Taipei University, Taipei 10433, Taiwan, ROC

Received 26 December 2005; received in revised form 3 July 2007
Available online 19 July 2007

Communicated by G. Sanniti di Baja

Abstract

This investigation develops an efficient face detection scheme that can detect multiple faces in color images with complex environments and different illumination levels. The proposed scheme comprises two stages. The first stage adopts color and triangle-based segmentation to search potential face regions. The second stage involves face verification using a multilayer feedforward neural network. The system can handle various sizes of faces, different illumination conditions, diverse pose and changeable expression. In particular, the scheme significantly increases the execution speed of the face detection algorithm in the case of complex backgrounds. Results of this study demonstrate that the proposed method performs better than previous methods in terms of speed and ability to handle different illumination conditions.

Keywords: Face detection; Color segmentation; Triangle-based segmentation; Neural network

1. Introduction

One of the important research topics in color image analysis is to find specific color regions in a given color image. Such a technique is very valuable, for example, automatic detection of human faces. Automatic face detection and recognition of human faces is one of the most intricate and important problems in face image database management, computer vision and cybernetics. It can be used as the security mechanism to replace metal key, plastic card, and password or PIN number. Above and beyond, it can also be applied in the intelligent systems for human-computer interaction. However, most face detection systems require the number of faces to be known beforehand, and the input faces to be free of environment and the size to be approximately consistent. Moreover, they cannot tolerate variable expression and a range of pose. These constraints greatly hinder the usefulness of the face detection system. Because a successful face detection process is the prerequisite to facilitate later face recognition task, we should not deal with face detection problem merely as a preprocessing of a face recognition system. If we do not have successful face detection method, the goal of successful face recognition cannot be achieved. Therefore, we should deal with face detection problem as important as the face recognition one. Speedy and accurate detection of faces in images is a crucial goal to be pursued in any face detection system.

Abundant of researches have been conducted on human face detection. Some successful systems have been developed and reported in the literature. For comprehensive survey of previous techniques, see Chellappa et al. (1995), Hjelmas and Low (2001) and Yang et al. (2002). They can be categorized into the following groups:

1.1. Template matching

Bichsel and Pentland (1993) utilized motion analysis and template matching to trail head movements in image

* Tel.: +886 2 25009838; fax: +886 2 25015974.
E-mail address: lincs@mail.ntpu.edu.tw

0167-8655/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2007.07.003

sequences. The head detector returns the position and orientation parameters of the face for each frame. From the variation of these parameters, they interpret head movements, such as nodding and shaking. Sobottka and Pitas (1996) located the poses of human faces and face features from color images. They presented a system that used an ellipse in Hue-saturation-value (HSV) color space to express the shape of a face.

1.2. Elastic graph matching (EGM)

Maurer and Malsburg von der (1996) proposed a system that can find out an arbitrary pose more accurately in the image sequences within the framework of EGM and jets. In their system, the faces have to be normalized according to the size and the initial pose in the sequence has to be known beforehand. Krüger et al. (1997) described a pose estimation algorithm based on EGM. The algorithm is an extension of the face representation introduced in (Wiskott et al., 1995) to the problem of pose estimation. Wiskott et al. (1997) represented faces by labeled graphs based on a Gabor wavelet transform. New faces are extracted by an EGM process and can be compared by an uncomplicated similarity function. They used "phase information" for accurate node positioning and used "object-adapted graphs" to handle large rotations in depth. In their method, image graph extraction is implemented based on a novel data structure (the bunch graph), which is constructed from a small set of sample image graphs.

1.3. Principal component analysis (PCA)

Pentland et al. (1994) utilized PCA for face recognition. Their system uses local Eigenfeatures. Manually selected landmarks, such as eyes and mouth, are used to apply a local PCA. Currently, PCA is still utilized very frequently in the area of face recognition. Gottumukkal and Asari (2003) stated a face detection system capable of detection of faces in real time from a streaming color video. Extracting skin color regions from a color image is the first step in this system. Skin color detection is used to segment regions of the image that correspond to face regions based on pixel color. Principle component analysis (PCA) is used to classify if a particular skin region is a face or a non-face. The PCA algorithm is trained for frontal view faces only. The tested images are captured by a surveillance camera in real time. Xie et al. (2005) presented a novel illumination compensation algorithm, which can compensate for the uneven illuminations on human faces and reconstruct face images in normal lighting conditions. An effective local contrast enhancement method, namely block-based histogram equalization (BHE), is first proposed. The resulting image processed using BHE is then compared with the original face image processed using histogram equalization (HE) to guesstimate the class of its light source. In order to remove the influence of uneven illumination while retaining the shape information about a human face, a 2D face shape model is used.

1.4. Neural network

Juell and Marsh (1996) proposed a hierarchical neural network to detect human faces in gray scale images. An edge enhancing preprocessor and four backpropagation neural networks arranged in a hierarchical structure were devised to find multiple faces in the scene. Their approach was invariant with respect to translation, rotation, and scale, but they cannot classify the pose. Rowley et al. (1998) presented a neural network-based face detection system by using a retinal connected neural network to check the small windows of an image, and judge whether each window contains a face or not. They adopted a small window (20 * 20) to slide over all portions of an image at various scales and used oval mask for ignoring background pixels. However, the inefficient search is a time-consuming procedure. The problem is that their method cannot detect partially occluded face and the pose. Han et al. (2000) proposed a system using a morphology-based technique to perform eye-analogue segmentation. Then, a trained backpropagation neural network performs the face verification task. Nevertheless, they cannot deal with partially occluded faces. Stathopoulou and Tsihrintzis (2004) expounded a system consists of two modules: firstly, they used a face detection algorithm to determine whether or not there are any faces in the image and return the location and extent of each face. Secondly, they used a facial expression classification module to determine the facial expression of a person.

1.5. Color/motion/shape information

Lee et al. (1996) adopted both the motion and color information of a face. Firstly, the face region is isolated from the complex background using its motion as a visual hint. They used sequence images of the whole face as the input to the system. They assumed that the object having the largest motion in the sequence image is the face to be detected. Secondly, they segmented the eye, eyebrow, and mouth regions of the face obtained in the first stage. However, their assumption will not hold in many cases. Dai and Nakano (1996) presented a texture model combining with color information. Their system is implemented based on the space gray level dependence (SGLD) matrix. The face texture model composed by a set of condition inequalities was derived by experimental basis. However, they used a window to scan the full picture, and calculated the average intensity to decide whether the window contains a face region or not. Saber and Tekalp (1998) expressed an algorithm for detecting human faces and facial features, such as the location of the eyes, nose and mouth. Firstly, a supervised pixel-based color classifier is employed to mark all pixels that are within a prespecified distance of "skin color", which is computed from a training set of skin

patches. An ellipse model is fit to each disjoint skin region. Finally, they presented symmetry-based cost functions to search the center of the eyes, tip of nose, and center of mouth within ellipses whose aspect ratio is similar to that of a face. Sherrah and Gong (1999) proposed a framework for fusing different information sources through the estimation of covariance from observations. The framework is established in a pose tracking procedure that fuses similarity-to-prototypes measures and skin color to trace head pose and face position. The use of data fusion through covariance introduces constraints that allow the tracker to robustly estimate the head pose and track the face location. Oliver et al. (2000) illustrated an active-camera real-time system for tracking, shape description, and classification of the human face and mouth expressions. Their system is based on the use of 2D blob features, which are spatially compact clusters of pixels that are similar in terms of low-level image properties. Patterns of behavior can be classified in real-time using hidden Markov models (HMMs). Hsu et al. (2002) proposed a face detection algorithm for color images. Based on a novel lighting compensation method and a nonlinear color transformation, they can detect skin regions over the entire image and then generates face candidates based on the spatial arrangement of these skin patches. The algorithm constructs eye, mouth, and boundary maps for verifying each face candidate.

1.6. Others

Gong et al. (1997) presented a framework for modeling nonlinear face poses density distributions using Gaussian mixtures for face recognition under image sequences from large head rotations in depth adopting general and modified Hyper Basis Function networks. Brunelli (1997) demonstrated a rough model of the 3D-head structure that is generated from a single frontal view of a face under uniform illumination. The location of one eye is required to know beforehand and the pose of the face is restricted to left–right rotations. Würtz (1997) introduced a system for the recognition of human faces independent of hairstyle. They used correspondence maps between an image and a model by coarse-fine matching in a Gabor pyramid. Wu et al. (2003) proposed an efficient face candidates selector for face detection tasks in still gray level images. Eye-analogue segments at a given scale are discovered by finding regions which are roughly as large as real eyes and are darker than their neighborhoods. Then a pair of eye-analogue segments are hypothesized to be eyes in a face and combined into a face candidate if their placement is consistent with the anthropological characteristic of human eyes.

However, most of the aforementioned approaches limit themselves to deal with human faces in frontal view, and only some of them use the color information. As a matter

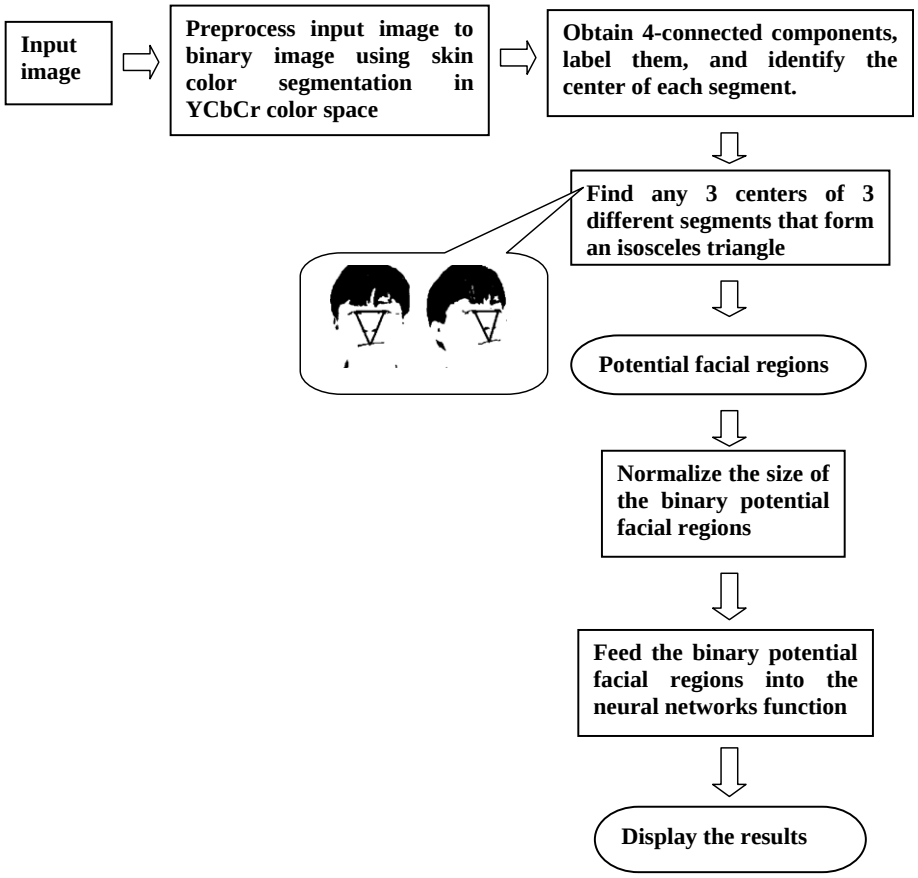

Fig. 1. Overview of the proposed system.

of fact, color is a very important and useful attribute for image matching and retrieval.

In this paper, a robust and efficient human face detection system is proposed to use this valuable component – color. Our approach can solve the problems in the aforementioned approaches, such as (1) detect the face that is smaller than 50 * 50 pixels, (2) detect multiple faces (more than 3 faces) in complex backgrounds, (3) handle varying pose and expression at the same time, and (4) detect the face that is in various illuminations. The designed system is composed of two principal parts as shown in Fig. 1.

The rest of the paper is organized as follows. In Section 2, searching the potential face regions based on skin color segmentation in YCbCr color space, and triangle-based criteria is described. In Section 3, each of the normalized potential face regions is fed to the neural network function to verify whether the potential face region really contains a face. Experimental results are demonstrated in Section 4 to verify the validity of the proposed face detection system. Finally, conclusions are given in Section 5.

2. Searching the potential face regions

This process searches the regions of an input image that potentially contain faces. The segmentation process comprises four steps. First, read in a color image. Then, transform this input color image to a binary image using skin color segmentation in the YCbCr color space. Second, label all four-connected components in the image to create blocks and identify the center of each block. Third, detect any three centers of three different blocks to form an isosceles triangle. Fourth, clip the blocks that meet the triangle criteria as the potential face region.

2.1. Skin color segmentation in YCbCr color space

Since the RGB color space is highly sensitive to intensity difference, many color spaces have been presented to improve color consistency or segmentation. The YCbCr color space was chosen for this investigation after performing many tests in the RGB, YES, HSI, and HSV color spaces. The Y in YCbCr denotes the luminance component, and Cb and Cr represent the chrominance factors. In YCbCr, the Y is the brightness (luma), Cb is blue[1] minus luma (B – Y) and Cr is red minus luma (R – Y). If R, G and B are given with 8 bit digital precision, then YCbCr (601) from "digital 8-bit RGB" can be obtained from RGB coordinate using a transformation matrix as illustrated in Fig. 2. When representing the signals in digital form, the results are scaled and rounded, and offsets are typically added. For example, the scaling and offset applied to the Y component per specification results in the value of 16 for black and the value of 235 for white when using an

$$\begin{bmatrix} Y \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256} \begin{bmatrix} 65.738 & 129.057 & 25.064 \\ -37.945 & -74.494 & 112.439 \\ 112.439 & -94.154 & -18.285 \end{bmatrix} \bullet \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Fig. 2. RGB–YCbCr conversion matrix.

8-bit representation. The standard has 8 bit digitized versions of Cb and Cr scaled to a different range of 16–240. The YCbCr space was chosen for six reasons. (1) The luminance component (Y) of YCbCr is independent of the color, so can be adopted to solve the illumination variation problem and it is easy to program. (2) According to Hsu et al. (2002), the skin color cluster is more compact in YCbCr than in other color space. (3) YCbCr has the smallest overlap between skin and non-skin data in under various illumination conditions. (4) YCbCr is broadly utilized in video compression standards (e.g., MPEG and JPEG) Garcia and Tziritas (1999). (5) YCbCr is a family of color spaces used in video systems. YCbCr was defined for standard-definition television use in the ITU-R BT.601 standard for use with digital component video. (6) YCbCr is one of two primary color spaces used to represent digital component video (the other is RGB). The difference between YCbCr and RGB is that YCbCr represents color as brightness and two color difference signals, while RGB represents color as red, green and blue. In YCbCr, the Y is the brightness (luma), Cb is blue minus luma (B – Y) and Cr is red minus luma (R – Y).

Many researchers assume that the chrominance components of the skin-tone color are independent of the luminance component (Jones and Rehg, 1998; Sobottka and Pitas, 1998; Saber and Tekalp, 1998; Menser and Brunig, 1999), however, in practice; the skin-tone color is nonlinearly dependent on luminance. Therefore, samples were taken at random from very large pixels in human faces to obtain the skin color distribution in the YCbCr color space. Human skin colors under different illumination circumstances were collected from 100 faces. One hundred pixels were collected from each face. Therefore, about 10,000 pixels of human skin colors were employed to acquire the skin color cluster. The principle applied in this investigation is completely different from that of Garcia and Tziritas (1999), who attempted to separate skin color zones from slightly different color regions. The principle in this work is that the output of human skin color segmentation must include seemingly similar skin colors. Some examples are shown in Fig. 3. Fig. 3a shows the original 32-bit-color map and the colors of human skin map. Fig. 3b illustrates the result generated by applying our skin colors segmentation to Fig. 3a. The pixels with skin color are assigned to original color, and the others are assigned to pure black color. The color set was defined as the union of seemingly similar skin colors and real human skin colors, even if some colors are not real human skin color. For instance, light yellow, light pink, and other light colors of red series were all included in the skin color set. We transfer the image from "RGB color space" to "Y, Cb,

[1] For interpretation of color in Figs. 3–8 and 10–13 the reader is referred to the web version of this article.

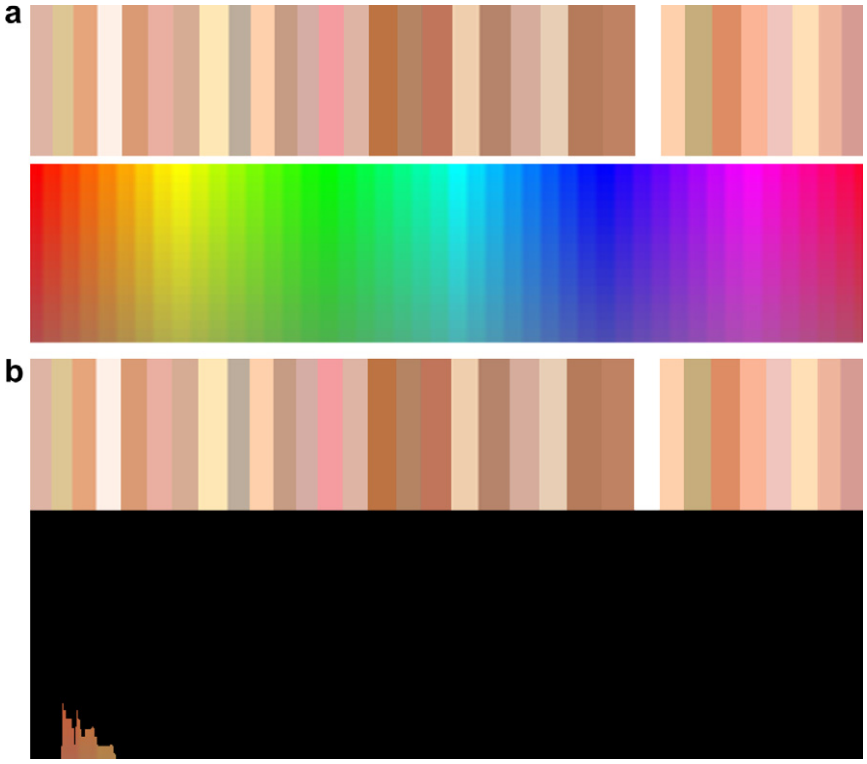

Fig. 3. (a) Shows the original 32-bit-color map and the colors of human skin map. (b) Illustrates the result generated by applying our skin colors segmentation to (a).

Cr color space'' is used only for human skin color segmentation. After ''human skin color segmentation'', we transfer the image from ''Y, Cb, Cr color space'' to ''RGB color space'' again. In other words, the apparently similar light colors should be treated as real skin colors, and should be retained in the human skin color segmentation process, since apparently similar light colors are eliminated in the following binarization process (color image in RGB color space $\rightarrow$ gray level image $\rightarrow$ binary image with threshold $= 100$). The threshold 100 is used to transfer gray level image to binary image. We binarize the gray level image to a ''binary image'' by simple global thresholding with threshold $T = 100$ because the objects of interest in our case are darker than the background. Pixels with gray level $\leqq 100$ are labeled black, and any pixel with gray level > 100 is labeled white.

From experiment, we obtain three rules to aid the task of human skin color segmentation:

1. $\mathrm{Y}(i) > \alpha$: means that the Y (luma/bright) should be larger than α.
2. $\mathrm{Cb} > \beta$: means that the Cb ($\mathrm{Cb} = \mathrm{Blue} - \mathrm{luma}$) should be larger than β.
3. $\mathrm{Cr} > \gamma$: means that the Cr ($\mathrm{Cr} = \mathrm{Red} - \mathrm{luma}$) should be larger than γ.

The first rule means that the (Y) (luma/bright) should be larger than α. The second rule means that the Cb ($\mathrm{Cb} = \mathrm{Blue} - \mathrm{luma}$) should be larger than β. The third rule means that the Cr ($\mathrm{Cr} = \mathrm{Red} - \mathrm{luma}$) should be larger than γ. In our system, the value of ''α'' $= 120$, the value of ''β'' $= 95$, and the value of ''γ'' $= 110$. In other words, if the pixels of the input image satisfy the above three rules, then the pixels are regarded as skin color and will keep the original color. Otherwise, we will treat the pixels as non-skin color and assign them to pure black color.

Human skin color segmentation task is performed first after reading in a color image. Pixels are classified in the YCbCr color space. Fig. 4 displays examples demonstrating that the proposed skin color segmentation algorithm is robust to various illumination conditions. Fig. 4a shows the original input RGB color image; Fig. 4b illustrates the result of the human skin color segmentation of Fig. 4a; Fig. 4c depicts the darker illumination input RGB color image; Fig. 4d shows the result of human skin color segmentation of Fig. 4c; Fig. 4e is the lighter illumination input RGB color image, and Fig. 4f illustrates the result of human skin color segmentation of Fig. 4e. Fig. 5 illustrates two examples illustrating that the proposed skin color segmentation algorithm is robust to dissimilar indoor illumination conditions. Fig. 5a depicts the input RGB color image; Fig. 5b displays the result of human skin color segmentation of Fig. 5a; Fig 5c denotes the input RGB color image, and Fig. 5d represents the result of human skin color segmentation of Fig. 5c. Fig. 6 depicts an example indicating that the proposed skin color segmentation approach is robust to various contrast and light sources from different directions. Fig. 6a displays the original input

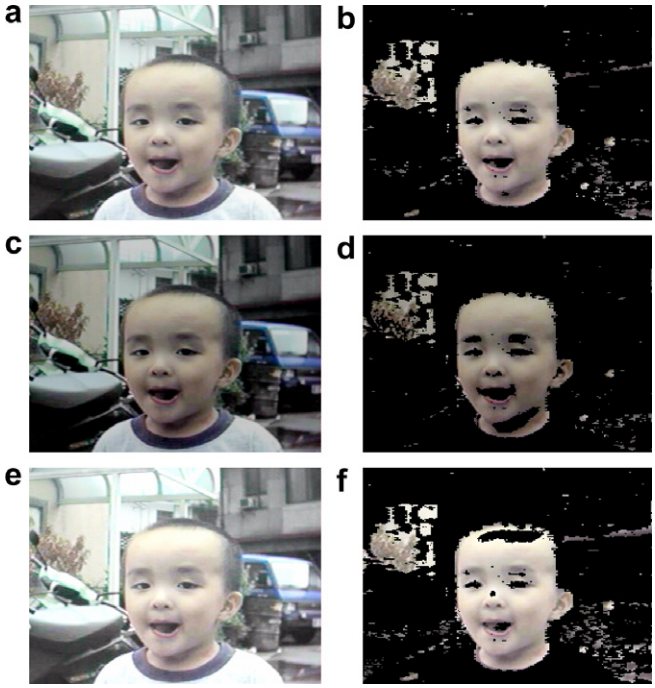

Fig. 4. Example demonstrating the effect of light on human skin color segmentation.

RGB color image; Fig. 6d illustrates the result of human skin color segmentation of Fig. 6a; Fig. 6b depicts the high-contrast input RGB color image; Fig. 6e shows the result of human skin color segmentation of Fig. 6b; Fig. 6c illustrates the low-contrast input RGB color image, and Fig. 6f displays the results of human skin color segmentation of Fig. 6c. Fig. 7 displays some examples illustrating that our skin color segmentation scheme is robust to different people/race and illumination condition.

2.2. Label all four-connected components and find the center of each block

Raster scanning (left-to-right and top-to-bottom) was adopted to obtain the four-connected components, label them, and then identify the center of each block. Raster scanning is described in detail in (Gonzalez et al., 1992).

2.3. Find any three centers of three different blocks forming an isosceles triangle

Careful observation indicates that two eyes and one mouth in the frontal view constitute an isosceles triangle. Potential face regions are discovered by finding such

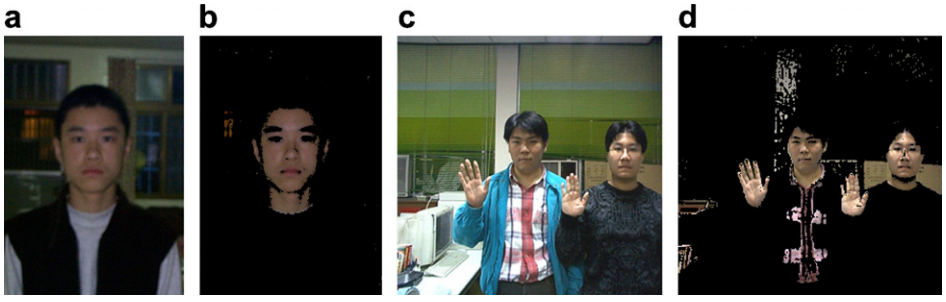

Fig. 5. Two examples of real image of an indoor environment.

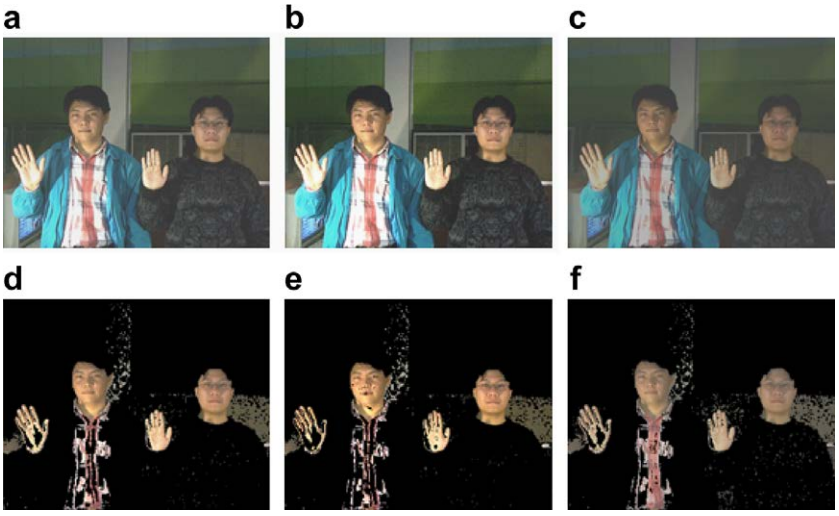

Fig. 6. Demonstrate an example illustrating that our skin color segmentation scheme is robust to different contrast and light source from different direction. (a) is the original input RGB color image, (d) is the result of human skin color segmentation of (a), (b) is the high contrast input RGB color image, (e) is the result of human skin color segmentation of (b), (c) is the low contrast input RGB color image, (f) is the result of human skin color segmentation of (c).

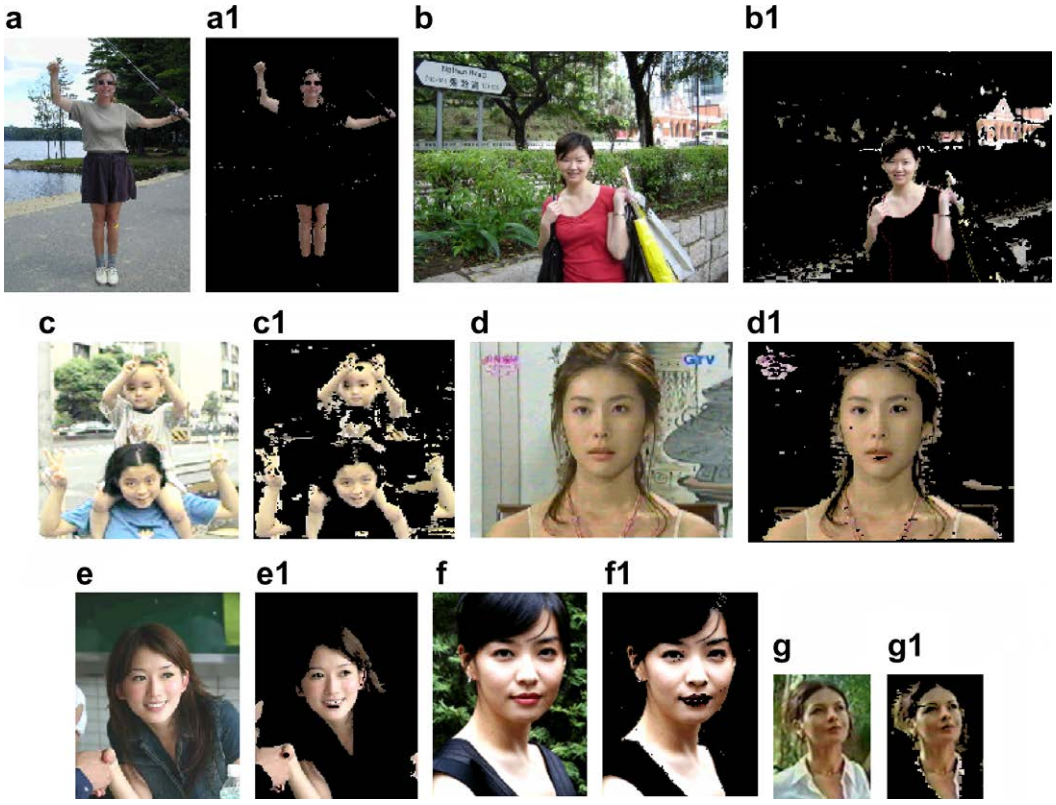

Fig. 7. Examples illustrating that our skin color segmentation scheme is robust to different people/race and illumination conditions.

triangles. Isosceles triangles obtained from the criteria of "the combination of two eyes and one mouth". Due to the imaging effect and imperfect binarization result, a 25% deviation is given to absore the tolerance. In other word, we don't try to find perfect isosceles triangles (even imperfect isosceles triangles are also allowable).

If the triangle ijk is an isosceles triangle as shown in Fig. 8, then it should possess the characteristic of "the distance of line ij = the distance of line jk". From observation, we discover that the Euclidean distance between two eyes (line ik) is about 90% to 110% of the Euclidean distance between the center of the right/left eye and the mouth. Due to the imaging effect and imperfect binarization result, a 25% deviation is given to absore the tolerance. The first matching rule can thereby be stated as $(\text{abs}(D(i,j) - D(j,k)) < 0.25 * \max(D(i,j), D(j,k)))$, and the second matching rule is $(\text{abs}(D(i,j) - D(i,k)) < 0.25 * \max(D(i,j), D(j,k)))$. Since the labeling process is operated from left to right then from top to bottom, we can get the third matching rule as "$i < j < k$". Here, "abs" means the absolute value, "$D(i,j)$" denotes the Euclidean distance between the centers of block i (right eye) and block j (mouth), "$D(j,k)$" denotes the Euclidean distance between the center of block k (left eye) and block j (mouth), "$D(i,k)$" represents the Euclidean distance between the centers of block i (right eye) and block k (left eye). When we can find two eyes and one mouth, they will form an isosceles triangle

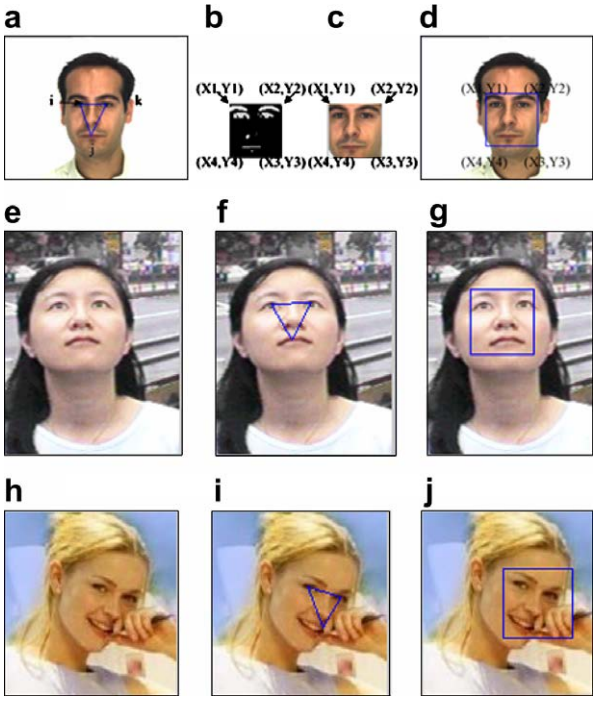

Fig. 8. Assume that (Xi, Yi), (Xj, Yj) and (Xk, Yk) are the three center points of blocks i (right eye), j (mouth), and k (right eye), respectively. The four corner points of the face region are thus (X_1, Y_1), (X_2, Y_2), (X_3, Y_3), and (X_4, Y_4). When we can find two eyes and one mouth, they will form an isosceles triangle (even imperfect isosceles triangles are also allowable).

(even imperfect isosceles triangles are also allowable). For example, as shown in Fig. 8f and Fig. 8i, if three points (i, j, and k) satisfy the matching rules, then we think that they form an isosceles triangle.

2.4. Clip the potential face regions

Assuming that the real facial region should cover the eyebrows, the eyes, the mouth and some area below the mouth, the coordinates can be determined as follows:

$$X_1 = X_4 = X_i - 1/4D(i,k); \quad (1)$$

$$X_2 = X_3 = X_k + 1/4D(i,k); \quad (2)$$

$$Y_1 = Y_2 = Y_i + 1/4D(i,k); \quad (3)$$

$$Y_3 = Y_4 = Y_j - 1/4D(i,k); \quad (4)$$

Assume that (X_i, Y_i), (X_j, Y_j) and (X_k, Y_k) denote the three center points of blocks i (right eye), j(mouth) and k(right eye), respectively, and (X_1, Y_1), (X_2, Y_2), (X_3, Y_3), and (X_4, Y_4) represent the four corner points of the face region as illustrated in Fig. 8. X_1 and X_4 locate at the same coordinates of $(X_i - 1/4D(i,k))$; X_2 and X_3 locate at the same coordinates of $(X_k + 1/4D(i,k))$; Y_1 and Y_2 locate at the same coordinates of $(Y_i + 1/4D(i,k))$, and Y_3 and Y_4 locate at the same coordinates of $(Y_j - 1/4D(i,k))$, where $D(i,k)$ represents the Euclidean distance between the centers of blocks i (right eye) and k (left eye). The detail of potential face regions segmentation can be found in (Lin and Fan, 2001).

3. Face verification

The second part of the proposed scheme is to perform face verification. In the previous section, a set of potential face regions in an image was selected. This section presents an efficient neural network function to determine whether a potential face region contains a face. This tasks has two steps. The first step is to normalize the size of each potential facial region. The second step is to feed each normalized potential facial region into the neural network function and then perform the verification.

3.1. Normalization of potential facial regions

Normalizing a potential face region can decrease the effects of variation in the distance and location. Since all potential faces are normalized to a standard size (i.e. 60×60 pixels) in this step, the potential face regions selected in the previous section can have different sizes. Herein, the potential facial region is resized by the bicubic interpolation technique as described in (Gonzalez et al., 1992).

3.2. The neural network function and the verification

The neural networks function adopted here work as follows:

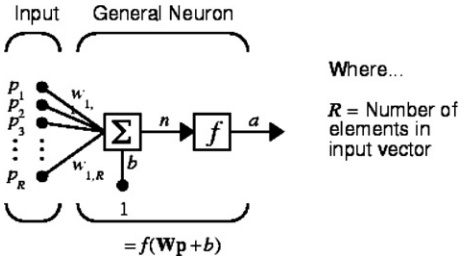

Fig. 9. Block diagram of the proposed neural network function.

Train the network: (1) Input the training data (e.g. 1200 faces and 1200 non_faces) to the network. (2) Compare the network output to the required output (If the image is a face, then the output should be 1 – is_face. If the image is not a face, then the output should be 0 – non_face.). (3) Modify the network weights to decrease error.

Use the network: (1) Input new data (e.g. a binary potential facial region) to the network. (2) The network determines output according to the training data. (If the image is a face, then the output should be 1 – is_face. If the image is not a face, then the output should be 0 – non_face.).

Because each binary potential facial region is resized to $60 * 60$ pixels as 3600 nodes, the normalized binary potential facial region contains 3600 nodes (e.g. a real binary human face) or 3600 nodes (e.g. not a real binary human face) is fed into the neural network function. The hidden layer has 15 nodes. The output unit gives a result of 0 (non_face) or 1 (is_face). Fig. 9 illustrates the block diagram of the proposed neural network function. An elementary neuron with R inputs is depicted below. Each input is weighted with a suitable value of w. The sum of the weighted inputs and the bias forms the input to the transfer function f. Each normalized binary potential facial region is fed into the neural network function to perform the verification. Once a face region has been confirmed, the final step is to remove regions that overlap with the selected face region, and then output the result.

4. Experimental results and discussion

The proposed approach was verified using 1500 face images (of 630 different persons). Some face images are taken from digital camera, some from digital video, some from scanner, and some from videotape. Moreover, we also use some parts of the "AR face database" (Martinez and Benavente, 1998). It contains over 4000 color images corresponding to 126 people's faces (70 men and 56 women) created by Aleix Martinez and Robert Benavente in the Computer Vision Center (CVC) at the UAB. Images in this database feature frontal view faces with different facial expressions, illumination conditions and occlusions (sun glasses and scarf). The pictures are taken at the CVC under strictly controlled conditions. No restrictions on wear (clothes, glasses, etc.), make-up or hair style are

imposed on participants. Each person appearing in the database participated in two sessions, two weeks (14 days) apart. We select *the parts excluding* the occlusions images (sun glasses and scarf), and examples of the detected faces with various conditions that are taken from AR database are shown in Fig. 10. Therefore, a total of 882 faces (126 persons * 7 faces = 882 faces) were selected from the "AR face database". The other face images (618 faces) are taken from digital camera, some from digital video, some from scanner, and some from videotape. Consequently, a total of 1500 faces were applied to verify the validity of the proposed system. Among these, only 27 faces could not be found correctly. Experimental results indicate that a success rate of around 98.2% (1473/1500 = 98.2%) was achieved. The false rejection rate was less than 2% (27/1500 = 1.8%), and the false acceptance rate was 0% (0/1500 = 0%).

The environments of the experimental set were as follows. Fig. 10 displays the verification of the face images with altered illumination conditions and different expression problems. The images of Fig. 10 are 192×144 pixels, and were taken with the following conditions: Fig. 10a neutral, Fig. 10b smile, Fig. 10c anger, Fig. 10d scream, Fig. 10e left light on, Fig. 10f right light on, Fig. 10g all side lights on. The execution time required to locate the precise locations of the face in the test image set depends on the image complexity, and Fig. 11 illustrates the experimental results of color images with complex backgrounds. Fig. 12 displays the experimental results of color images without skin color segmentation in complex backgrounds (our previous system). Fig. 12a illustrates the original input RGB color image; Fig. 12b depicts the result of binary image of Fig. 12a with 99 blocks; Fig. 12c displays the isosceles triangles obtained from the criteria of "the combination of two eyes and one mouth"; Fig. 12d illustrates the result of triangle-based segmentation of Fig. 12b; Fig. 12e depicts the result of triangle-based segmentation of Fig. 12a; Fig. 12f displays the result of triangle-based segmentation with a normalized size binary image, and Fig. 12g shows the final result of human face detection of Fig. 12a. Fig. 13 displays the experimental results of color images with skin color segmentation in complex backgrounds (our novel system). Fig. 13a depicts the original input RGB color image; Fig. 13a1 illustrates the result of the skin color segmentation of Fig. 13a, Fig. 13b displays the result of binary image of Fig. 13a1, where the number of blocks is only 7, Fig. 13c is the result of isosceles triangles obtained from the criteria of "the combination of two eyes and one mouth", Fig. 13d illustrates the result of triangle-based segmentation of Fig. 13b and e depicts the result of triangle-based segmentation of Fig. 13a1; Fig. 13f shows the result of triangle-based segmentation with a normalized size binary image, Fig. 13g illustrates the result of the potential human face of Fig. 13a1, and Fig. 13g1 depicts the final result of human face detection of Fig. 13a. The decrease in the number of blocks (from 99 to 7) indicates that the speed improves significantly when the background is complex. Figs. 12a/13a display the same image with size 90×161 pixels. Fig. 12a requires 387.3750 s to locate the correct face position by using a P4 CPU 3.0 GHz PC in our previous work Lin and Fan (2001), but Fig. 13a needs only 0.0625 s to locate the correct face position using a P4 CPU 3.0 GHz PC in the new system. Rowley et al. (1998) presented a neural network-based face detection system by using a retinal connected neural network to check the small windows of an image, and judge whether each window contains a face or not. Because their inefficient search [they adopted a small window (20 * 20) to slide over all portions of an image at various scales] is a time-consuming procedure, the experimental results demonstrate that our method is very fast. For example, Fig. 13a needs only 0.0625 s to locate the correct face position using a P4 CPU 3.0 GHz PC in our new system. Therefore, the proposed method is better than Rowley et al. (1998).

We consider not only Asian people, but also consider European and other people except African. Since African have dark black skin, it is hard to obtain satisfied binary

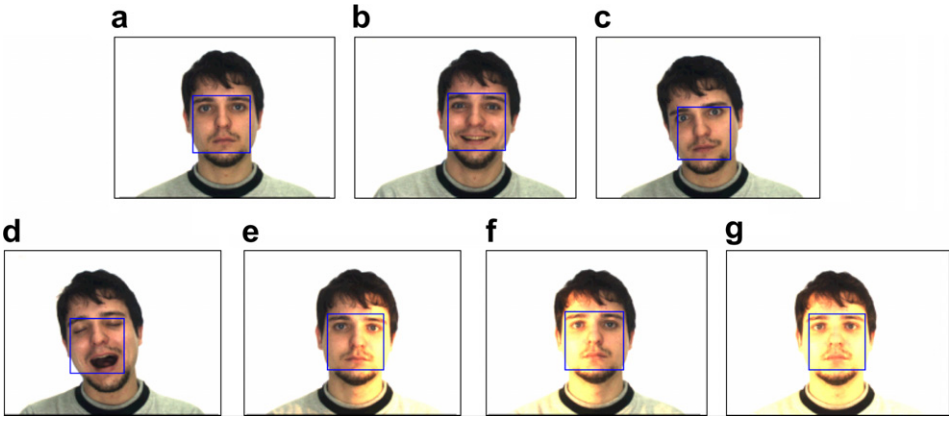

Fig. 10. Face images with altered lighting conditions and different expressions by using dynamic thresholds (a) neutral expression, (b) smile, (c) anger, (d) scream, (e) left light on, (f) right light on, (g) all side lights on, respectively.

Fig. 11. Experimental results of color images with complex backgrounds.

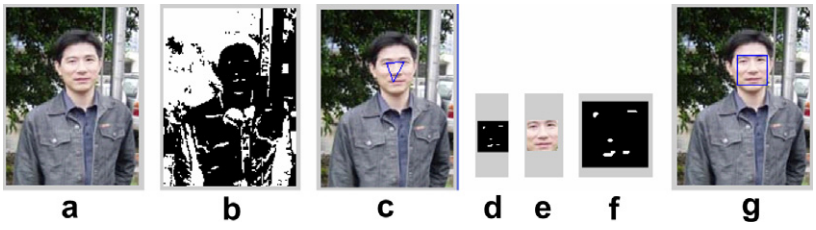

Fig. 12. Experimental results *without skin color segmentation* in complex backgrounds. (previous system).

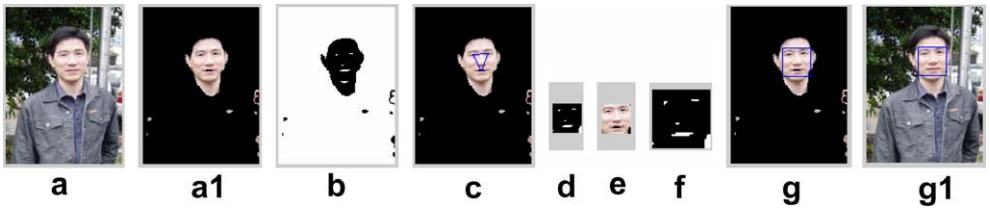

Fig. 13. Experimental results *with skin color segmentation* in complex backgrounds. (proposed system).

image with threshold = 100 in the post-process. We will try to overcome this weak point in the future.

5. Conclusions

This investigation develops a robust and effective face detection scheme to extract faces from various images. Since human skin color segmentation can remove complex backgrounds, it can significantly decrease execution time in the complex background case. The proposed system has three novel distributions. First, since in this work, the output of human skin color segmentation must include all apparently similar skin color (e.g. light yellow and light pink and other light colors of red series were all included

in the skin color set as illustrated in Fig. 3) unlike in other skin color segmentation rules, making the proposed system very robust for various lighting conditions. Second, the neural network was incorporated into the new system, and the experimental results demonstrate that the proposed method is better than Rowley et al. (1998) in terms of efficiency. Finally, using a color and triangle-based segmentation process can reduce the background part sufficiently. This process significantly speeds up the subsequent face detection procedure, since only some of the original image remains for further processing.

The proposed novel human skin color segmentation algorithm is significantly faster than our previous investigation in (Lin and Fan, 2001) in the case of complex backgrounds. In the future, we plan to use this face detection system for preprocessing to solve face recognition problems.

References

Bichsel, M., Pentland, A., 1993. Automatic Interpretation of Human Head Movements. MIT Media Laboratory, Vision and Modeling Group, Technical Report No. 186.

Brunelli, R., 1997. Estimation of pose and illuminant direction for face processing. Image Vision Comput. 15, 741–748.

Chellappa, R., Wilson, C.L., Sirohey, S., 1995. Human and machine recognition of faces: A survey. Proc. IEEE 83 (5), 705–740.

Dai, Y., Nakano, Y., 1996. Face-texture model based on SGLD and its application in face detection in a color scene. Pattern Recognition 29 (6), 1007–1017.

Garcia, C., Tziritas, G., 1999. Face detection using quantized skin color regions merging and wavelet packet analysis. IEEE Trans. Multimedia 1 (3), 264–277.

Gong, S., Ong, E.J., Loft, P.J., 1997. Appearance-based face recognition under large head rotations in depth. <http://www.dcs.qmw.ac.uk/~sgg/msc/hbt/>.

Gonzalez, Rafael C., Woods, Richard E., 1992. Digital Image Processing. Addison-Wesley Publishing Company, Inc.

Gottumukkal, R., Asari, V.K., 2003. Real time face detection from color video stream based on PCA method. Applied Imagery Pattern Recognition Workshop, 2003. In: Proc. 32nd, 15–17 October 2003, pp. 146–150.

Han, C.C., Mark Liao, H.Y., Yu, G.J., Chen, L.H., 2000. Fast face detection via morphology-based pre-processing. Pattern Recognition 33 (10), 1701–1712.

Hjelmas, E., Low, B.K., 2001. Face detection: A survey. Computer Vision and Image Understanding 83, 236–274.

Hsu, Rein-Lien, Abdel-Mottaleb, M., Jain, A.K., 2002. Face detection in color images. IEEE Trans. Pattern Anal. Machine Intell. 24 (5), 696–706.

Jones, M., Rehg, J.M., 1998. Statistical Color Models with Application to Skin Detection, Technical Report Series, Cambridge Research Laboratory, December 1998.

Juell, P., Marsh, R., 1996. A hierarchical neural network for human face detection. Pattern Recognition 29 (5), 781–787.

Krüger, N., Pötzsch, M., von der Malsburg, C., 1997. Determination of face position and pose with a learned representation based on labeled graphs. Image Vision Comput. 15 (8), 665–673.

Lee, C.H., Kim, J.S., Park, K.H., 1996. Automatic human face location in a complex background using motion and color information. Pattern Recognition 29 (11), 1877–1889.

Lin, C., Fan, K.C., 2001. Triangle-based approach to the detection of human face. Pattern Recognition 34 (6), 1271–1284.

Martinez, A.M., Benavente, R., 1998. The AR Face Database. CVC Technical Report #24, June 1998.

Maurer, T., Malsburg von der, C., 1996. Tracking and learning graphs and pose on image sequences of faces. In: Proc. 2nd Internat. Conf. on Automatic Face and Gesture Recognition, Killington, Vermont 1996, IEEE Computer Society Press.

Menser, B., Brunig, M., 1999. Locating human faces in color images with complex background. Intelligent Signal Process. Commun. Systems, 533–536, December.

Oliver, N., Pentland, A., Bérard, F., 2000. LAFTER: A real-time face and lips tracker with facial expression recognition. Pattern Recognition 33 (8), 1369–1382.

Pentland, A., Moghaddam, B., Starner, T., 1994. View-based and modular eigenspaces for face recognition. IEEE Conf. on Computer Vision and Pattern Recognition, Seattle.

Rowley, H., Baluja, A.S., Kanade, T., 1998. Neural network-based face detection. IEEE Trans. Pattern Anal. Machine Intell. 20 (1), 23–38.

Saber, E., Tekalp, A.M., 1998. Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions. Pattern Recognition Lett. 19, 669–680.

Sherrah, J., Gong, S., 1999. Fusion of 2D face alignment and 3D head pose estimation for robust and real-time performance. In: Proc. IEEE Internat. Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems, Corfu, Greece September 1999, pp. 26–27.

Sobottka, K., Pitas, I. 1996. Extraction of facial regions and features using color and shape information. In: Proc. 13th Internat. Conf. Pattern Recognition, Vienna, Austria, August 1996, pp. 421–425.

Sobottka, K., Pitas, A., 1998. Novel method for automatic face segmentation, facial feature extraction and tracking. Signal Process. Image Commun. 12, 263–281.

Stathopoulou, I.-O., Tsihrintzis, G.A., 2004. An improved neural-network-based face detection and facial expression classifleation system. IEEE Internat. Systems Man and Cybernet. 1 (10–13), 666–671.

Wiskott, L., Fellous, J.M., Krüger, N., von der Malsburg, C., 1995. Face recognition and gender determination. In: Bichsel, M. (Ed.), Proc. Internat. Workshop on Automatic Face and Gesture Recognition. Multimedia Laboratory, Zurich.

Wiskott, L., Fellous, J.M., Krüger, N., von der Malsburg, C., 1997. Face Recognition by Elastic Bunch Graph Matching. IEEE Trans. Pattern Anal. Machine Intell. 19 (7), 769–775.

Wu, Jianxin, Zhou, Zhi-Hua, 2003. Efficient face candidates selector for face detection. Pattern Recognition 36 (5), 1175–1186.

Würtz, R.P., 1997. Object recognition robust under translations, deformations, and changes in background. IEEE Trans. Pattern Anal. Machine Intell. 19 (7), 775–779.

Xie, Xudong, Lam, Kin-Man, 2005. Face recognition under varying illumination based on a 2D face shape model. Pattern Recognition 38 (2), 221–230.

Yang, Ming-Hsuan, Kriegman, D.J., Ahuja, N., 2002. Detecting faces in images: A survey. IEEE Trans. PAMI 24 (1), 34–58.