Face Detection Based on Skin Color Segmentation and Neural Network

Hwei-Jen Lin, Shu-Yi Wang, Shwu-Huey Yen, and Yang-Ta Kao
Department of Computer Science and Information Engineering
Tamkang University, Taipei, Taiwan, China
E-mail: hjlin@cs.tku.edu.tw

***Abstract*—This paper proposes a human face detection system based on skin color segmentation and neural networks. The system consists of several stages. First, the system searches for the regions where faces might exist by using skin color information and forms a so-called skin map. After performing noise removal and some morphological operations on the skin map, it utilizes the aspect ratio of a face to find out possible face blocks, and then eye detection is carried out within each possible face block. If an eye pair is detected in a possible face block, a region is cropped according to the location of the two eyes, which is called a face candidate; otherwise it is regarded as a non-face block. Finally, each of the face candidates is verified by a 3-layer back-propagation neural network. Experimental results show that the proposed system results in better performance than the other methods, in terms of correct detection rate and capacity of coping with the problems of lighting, scaling, rotation, and multiple faces.**

I. INTRODUCTION

Face detection has received much attention and has been an extensive research topic in recent years. It is the important first step of many applications such as face recognition [20], facial expression analysis, surveillance systems, video-conferencing [4], intelligent human-computer interaction, content-based image retrieval systems, etc. Therefore, the efficiency of face detection influences the performance of these systems. There have been various approaches proposed for face detection, which could be generally classified into four categories. (i) Template matching methods (ii) Feature-based methods, (iii) Knowledge-based methods, and (iv) Machine learning methods. Template matching method means the final decision comes from the similarity between input image and template. It is scale-dependent, rotation-dependent and computational expensive. R. Feris et al. [15] verified the presence or absence of a face in each skin region by using an eye detector based on a template matching scheme. Y. Suzuki and T. Shibata [22] utilized edge distribution of face images and non-face images to generate several 64-dimensional feature vectors which could be regarded as templates. Then the feature vector of the input image is compared with all the templates and classified as face or non-face depending on the matching result. J. G. Wang and T. N. Tan [7] presented a deformable template based on the edge information to match the face contour. Feature-based methods use low-level features such as gray [2], [11], color [5, 6, 7, 8, 9, 14, 15], edge [20], shape [6], [7], and texture to locate facial features, and further, find out the face location. During recent years, face detection algorithms based on skin color information has attracted more attention of many researches. Therefore, the accuracy of skin color detection is very important to a face detection system. In [12], they proposed an efficient color compensation scheme for skin color segmentation. In [10], an adaptive skin color filter is proposed for detecting skin color regions in a color image. However, it failed to detect the skin color regions when an input image is composed of several different races. V. Vezhnevets et al. [21] surveyed numerous pixel-based skin color detection techniques. Knowledge-based methods [3] detected an isosceles triangle (for frontal view) or a right triangle (for side view). Machine learning methods use a lot of training samples to make the machine to be capable of judging face or non-face. In [13], eigenface for face recognition is presented. S. Karungaru et al. [18] used a fixed window to scan whole image then the input image is verified by a back-propagation neural network.

In this paper, we propose a novel approach combining feature-based, knowledge-based, and machine learning methods for detecting human faces in color images under different illumination condition, scale, rotation, wearing glasses, etc. Firstly, skin color segmentation is performed to find skin color regions. Secondly, possible face blocks are located by using some restrictions on these regions. Thirdly, eye detection and matching are carried out within each possible face blocks, and then face candidates will be obtained according to the locations of the detected eye pairs. Finally, each of the face candidates is verified by a 3-layer back-propagation neural network.

The rest of this paper is organized as follows. In Section 2, we discribe how to obtain the face candidates. The process of face canidade verification is presented in Section 3. Section 4 provides some experimental results and comparison with other systems. Finally, conclusions are given in Section 5.

II. FACE CANDIDATE SEARCHING

0-7803-9422-4/05/$20.00 ©2005 IEEE

In this section, the main purpose is to obtain face candidates. There are three steps to achieve this goal. First, skin color regions are located by performing skin color segmentation. Second, some restrictions to these regions are used to locate possible face blocks. Finally, eye detection and matching are implemented within each possible face blocks, and then face candidates are determined according to the locations of the detected eye pairs.

A. Skin color segmentation

Skin color is a very important feature of human faces. The distribution of skin colors clusters in a small region of the chromatic color space [9]. Processing color is faster than processing other facial features. Therefore, skin color detection is firstly performed on the input color image to reduce the computational complexity. Because of the accuracy of skin color detection affects the result of face detection system, choosing a suitable color space for skin color detection is very important. Among numerous color spaces, RGB color space is sensitive to the variation of intensity, and thus it is not sufficient to use only RGB color space to detect skin color.

In this paper, we combine RGB, Normalized RGB, and HSV color spaces to detect skin pixels. This is due to the fact that both normalized RGB and HSV color space can reduce the effect of lighting to an image. In Normalized RGB color space, the color values (r, g) are defined in Equations (1) and (2), where color blue is redundant because $r+g+b=1$.

$$r = R/(R+G+B), \tag{1}$$

$$g = G/(R+G+B). \tag{2}$$

The HSV (hue, saturation, value) color space [23] can be represented by the hex-cone model as shown in Figure 1. The hue (H) varies from 0 to 360°. The saturation (S) varies from 0 to 1. The value (V) ranges between 0 and 1. The HSV color space can be converted from the RGB color space using Equations (3) to (7).

$$H1 = \cos^{-1}\left\{\frac{0.5[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right\}, \tag{3}$$

$$H = H1 \quad \text{if } B \le G, \tag{4}$$

$$H = 360^\circ - H1 \quad \text{if } B > G, \tag{5}$$

$$S = \frac{\mathrm{Max}(R,G,B) - \mathrm{Min}(R,G,B)}{\mathrm{Max}(R,G,B)}, \tag{6}$$

$$V = \frac{\mathrm{Max}(R,G,B)}{255} \tag{7}$$

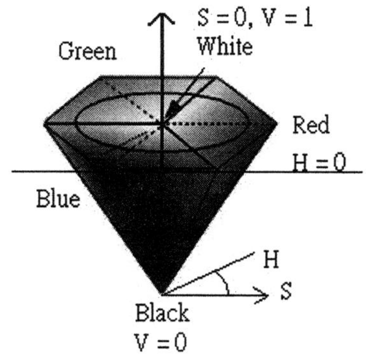


Figure 1. HSV color space

At first, RGB values of pixels in the input image are transformed into Normalized RGB and HSV color space. A pixel is labeled as a skin pixel if its color values conform to Inequalities (8), (9), (10), and (11). As a result, we can generate a binary skin map where the white points represent the skin pixels and the black points represent the non-skin pixels. Then, we apply median filter to the skin map for noise removing and perform morphological opening operation [16] with structuring element of size 3×3 to eliminate small skin blocks. Afterward, utilizing connected component operation to find out all connected skin regions and each of the skin regions is labeled by a bounding box. The process is depicted in Figure 2.

$$R > G, |R-G| \ge 11, \tag{8}$$

$$0.33 \le r \le 0.6, 0.25 \le g \le 0.37, \tag{9}$$

$$340 \le H \le 359 \vee 0 \le H \le 50, \tag{10}$$

$$0.12 \le S \le 0.7, 0.3 \le V \le 1.0 \tag{11}$$

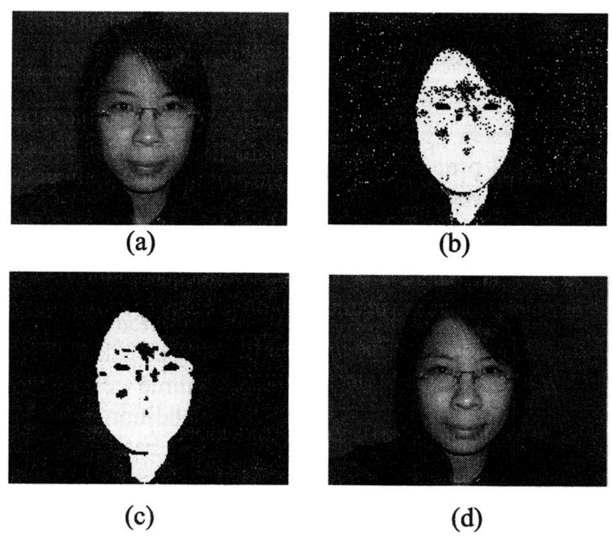


Figure 2. (a) original image, (b) skin map, (c) skin map performed median filter followed by opening, (d) result of skin region detection.

B. Locating possible face blocks

To check if a skin region contains a face, we use three constraints: *Area size*, *Aspect Ratio*, and *Occupancy*. A skin region that satisfies the three constraints is regarded as a possible face block; otherwise, a non-face region.

The first constraint is that the size of the bounding box surrounding the skin region, denoted by *Area size*, is greater than 30×30. This is due to the fact that in a skin region of too small size, facial features might be eliminated during the pre-processing.

The second constraint is that the ratio of the height to the width of the bounding box, denoted by *Aspect Ratio,* is between 0.8 and 2.6. This rule is based on observations on various face blocks.

Occupancy denotes the ratio of the amount of the skin pixels to the size of the bounding box. If the bounding box contains a face, the number of skin pixels in the bounding box must be large enough. We relax the restriction to avoid removing a real face and set the third constraint as "*Occupancy* is greater or equal to 40%".

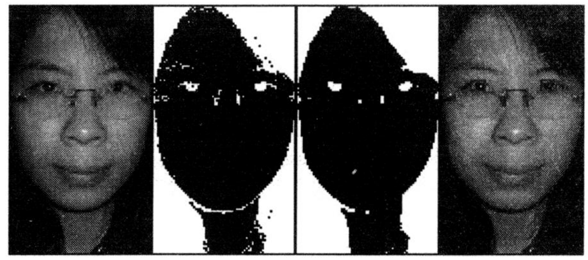
The skin regions that do not obey all the three constraints are removed, and the remaining are considered as possible face blocks and need further verification.

C. Determining face candidates

Eye is the most stable facial feature of human face. Suppose face is in frontal view, there must be two eyes in the possible face block. Hence, eye detection and matching is carried out to detect any eye pair existing in the possible face blocks. During the procedure, a possible face block is removed if it contains no eye or only one eye; otherwise, according to the location of the detected eye pair, a face candidate is located.

1) Eye detection:

In general, the intensity of eye is darker than that of other facial features in a face and it does not belong to skin region. Utilizing this property of eyes, we can find out some eye pixels and mark them by white points in the possible face block, as shown in Figure 3(b). Since some noises are simultaneously produced during this process, we apply median filter to remove them (as shown in Figure 3(c)) and then perform connected component operation to find all eye-like blocks. Each of the eye-like blocks is labeled by a bounding box, and then examined by three conditions to verify if it contains an eye. The first condition is that the *Aspect Ratio* of the eye-like block must be between 0.2 and 1.67. The second condition is that the *Occupancy* must be greater or equal to 30%. The third condition is that the ratio of the width of the eye-like block to the width of the possible face block is between 0.028 and 0.4. These parameters came from experimental results. Examining these conditions, eye blocks can be detected, as shown in Figure 3(d).



(a) (b) (c) (d)

Figure 3. (a) possible face block, (b) eye pixels, (c) (b) eye pixels after performing median filter, (d) result of eye detection.

2) Eye matching:

Continuously, we match every two eye blocks that form an eye pair for frontal view based on the geometrical relation of facial features. For each pair of eye blocks, we first locate the centroid of each of the two eye blocks and calculate the horizontal distance, *Dist*, between two centroids. Second, we match the two eye blocks based on the following rules: "*Dist* is between T_1 and T_2 times of the width of the face," "The eye is located at upper portion of the face," and "The sizes of the two eyes are nearly equal." In our experiments, we set $T_1 = 0.2$, $T_2 = 0.65$. As soon as an eye pair is located, we can clip a face candidate based on the face model as shown in Figure 4, where D is the distance of the centroids of the two eyes.

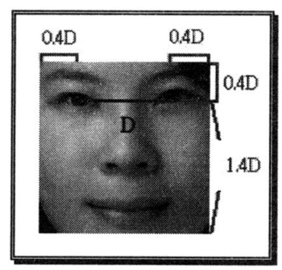


Figure 4. Face model

III. FACE CANDIDATE VERIFICATION

The previous section shows how to clip the face candidates according to the detected eye pairs. In this section, we use a 3-layer back-propagation neural network (BPNN) to verify each face candidate. The advantage of using BPNN is that the process of classification is very rapid. In our approach, the BPNN contains 400 (20×20) input units, 20 hidden units, and one output unit. The *transfer function* used here is the *sigmoid* function. There are 300 face and 300 non-face gray-level training samples. The size of each training sample is 20×20 pixels. They are manually clipped from photographs, digital cameras, or internet. Before training, each of the training samples is intensity-normalized by utilizing Equations (12) and (13), where *N* is the number of pixels in a training sample, $\bar{I}$ is the average gray value, I_i is the gray value of the ith pixel, and $I_i^{'}$ is the normalized gray value.

$$\bar{I} = \frac{1}{N}\sum_{i=1}^{N} I_i \qquad (12)$$

$$I_i^{'} = (I_i - \bar{I}) + 128 \qquad (13)$$

Figure 5 shows the intensity-normalization results of some training samples. The original images with different intensity and the resulting intensity-normalized images are listed in the upper row and the lower row, respectively. As a result, we can observe that the intensity of training samples would become uniform after intensity-normalization.

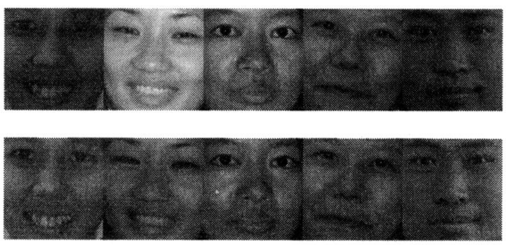


Figure 5. (a) images with different intensity, (b) results of intensity-normalization

After the 3-layer BPNN is trained, a face candidate can be verified through the following steps:

I) Normalize the size of the face candidate to size of 20×20 by the Nearest Neighbor method [16].

II) Apply intensity-normalization to the face candidate by using Equations (12) and (13).

III) Input face candidate into the BPNN to produce the output.

IV) If the output is larger than a threshold *T*, the face candidate is regarded as a human face; otherwise it is not. We set $T = 0.95$ for our network.

IV. EXPERIMENTAL RESULTS

In this section, we show a set of experimental results to demonstrate the performance of the proposed system. Our experiments were implemented in a circumstance using the AMD Athlon 2200+ 1.8 GHz CPU, with 512 MB memory and the Windows XP operating system.

A. Performance of BPNN

There are 300 face and 300 non-face testing samples used in our experiment to measure the performance of BPNN. Among the face testing samples, number of false rejection is 13. Among the non-face testing samples, number of false acceptance is 38. Therefore, false *rejection rate* is 4.33%; *false acceptance rate* is 12.67%. The performance of the BPNN is shown in Table 1.

Table 1. Classification result of BPNN

classification results \ testing samples	face	non-face
face	287	38
non-face	13	262

B. Face detection results

We tested our system on 1817 still color images, taken from digital cameras, scanners, the World Wide Web, and the Champion dataset [19]. They consist of both indoor and outdoor scenes under different lighting conditions and backgrounds. The image sizes vary from 105×158 to 640×577. Among them, there are totally 2615 faces of sizes varying from 35×30 to 370×275 pixels. Faces vary in lighting, scale, position, rotation, race, color, and facial expression.

Figure 6 shows some face detection results of our system. The man wearing glasses in Figure 6(a), two faces with different facial expression in Figure 6(b), and the faces under non-uniform lighting condition in Figure 6(c) are all successively detected. Moreover, our system also detects the faces with rotation in Figure 6(d), the face with partial occlusion in Figure 6(e), the girl in the left side turning her head left about 45° in Figure 6(f), and the face under complex background in Figure 6(g).

In the experiments, we also compare our method with the method proposed by Fröba and Küblbeck [1], which detects faces based on edge orientation matching. The comparison is given in Table 2.

Table 2. Comparison of our method and the method proposed by Fröba and Küblbeck [1]

methods \ results	test images include 2615 faces	
	true detection	false detection
method proposed by Fröba and Küblbeck	2299	352
our method	2308	152

As the results shown in Table 2, the method proposed by Fröba and Küblbeck [1] and our method have true detection rates 87.92% and 88.26%, and have numbers of false detection 352 and 152, respectively. Although the detectoin rates of these two methods are nearly equal to each other, our method detects much less false faces. Figure 7 shows the detection results of the method proposed by Fröba and Küblbeck, testing on the same images as those in Figure 6.

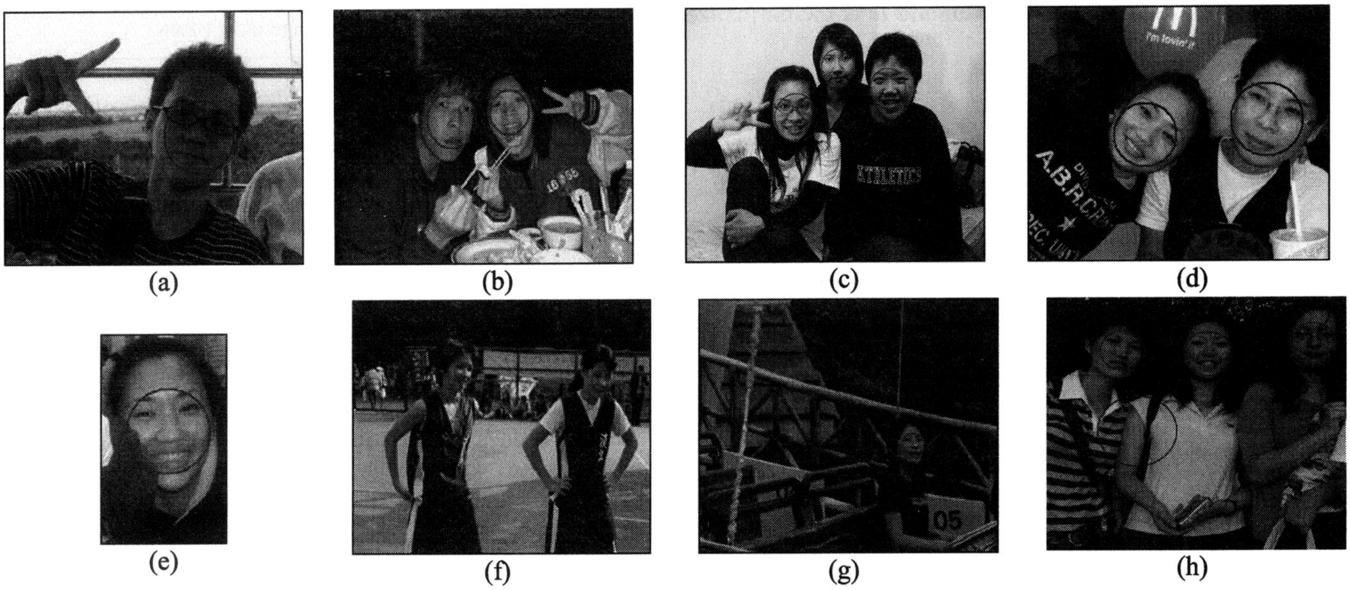


Figure 6. Some face detection results of the proposed method

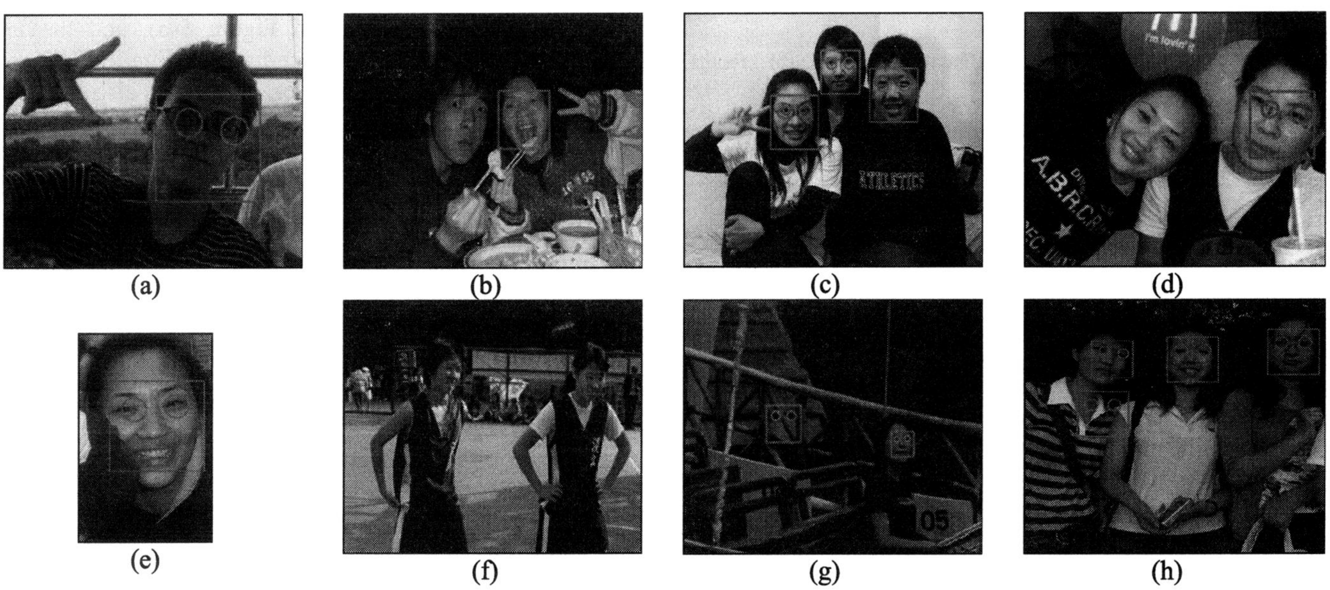


Figure 7. some detection results of the method proposed by Fröba and Küblbeck

V. DISCUSSIONS AND CONCLUSIONS

This paper proposes a human face detection system based on skin color segmentation and neural networks. Experimental results show that the proposed system results in better performance than some other methods, in terms of correct detection rate and capacity of coping with the problems of lighting, scaling, rotation, and multiple faces.

Although the proposed method shows high detection rate, it still has some problems as stated in the following:

I) Since skin color information is used for face detection, if the illumination is too bright or too dark the system would fail in skin color detection.

II) Since we determine a face candidate according to the location of two eyes, when eyes are not successfully detected, the system would fail in face detection.

In our future work, we would like to solve these problems. For the first problem, we will try to find a more robust skin color detection algorithm. For the second problem, we might try to relax the rules of detecting eye pairs.

REFERENCES

[1]. Bernhard Fröba and Christian Küblbeck, "Face Detection and Tracking using Edge Orientation Information," SPIE Visual Communications and Image Processing, 2001, pp. 583-594.

[2]. Chin-Chuan Han, Hong-Yuan Mark Liao, Gwo-Jong Yu, and Liang-Hua Chen, "Fast Face Detection via Morphology-based Pre-processing," Pattern Recognition, Vol. 33, No. 10, 2000, pp. 1701-1712.

[3]. Chiunhsiun Lin and Kuo-Chin Fan, "Triangle-based Approach to the Detection of Human Face," Pattern Recognition, Vol. 34, No. 6, 2001, pp. 1271-1284.

[4]. Cuizhu Shi, Keman Yu, Jiang Li, and Shipeng Li, "Automatic Image Quality Improvement for Videoconferencing," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04), May 2004, pp. III - 701-704.

[5]. F. Solina, P. Peer, B. Batagelj, S. Juvan, and J. Kovaè, "Color-based Face Detection in the "15 Seconds of Fame" Art Installation," Proceedings of the Computer Vision/Computer Graphics Collaboration for Model based Imaging, Rendering, image Analysis and Graphical special Effects (MIRAGE 2003), INRIA Rocquencourt, France, 2003, pp. 38-47.

[6]. Ing-Sheen Hsieh, Kuo-Chin Fan, and Chiunhsiun Lin, "A Statistic Approach to the Detection of Human Faces in Color Nature Scene," Pattern Recognition, Vol. 35 , 2002, pp. 1583-1596.

[7]. J. G. Wang and T. N. Tan, "A New Face Detection Method Based on Shape Information," Pattern Recognition Letters, No. 21, 2000, pp. 463-471.

[8]. J. Kovaè, P. Peer, and F. Solina, "Human Skin Colour Clustering for Face Detection," Proceedings of the International Conference on Computer as a Tool (EUROCON 2003), Vol. 2, Ljubljana, Slovenia, September 2003, pp. 144-148.

[9]. Jie Yang and A. Waibel, "A Real-time Face Tracker," Applications of Computer Vision (WACV'96), Proceedings of the 3rd IEEE Workshop, December 1996, pp. 142-147.

[10]. K. M. Cho, J. H. Jang, and K. S. Hong, "Adaptive Skin-color Filter," Pattern Recognition, Vol. 34, No. 5, 2001, pp. 1067-1073.

[11]. Kwok-Wai Wong, Kin-Man Lam, and Wan-Chi Siu, "An Efficient Algorithm for Human Face Detection and Facial Feature Extraction under Different Conditions," Pattern Recognition, Vol. 34, 2001, pp. 1993-2004.

[12]. Kwok-Wai Wong, Kin-Man Lam, and Wan-Chi Siu, "An Efficient Color Compensation Scheme for Skin Color Segmentation," Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS2003), Vol. II, May 2003, pp. 676-679.

[13]. M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, Vol. 3, No. 1, March 1991, pp 71-86.

[14]. P. Peer and F. Solina, "An Automatic Human Face Detection Method," Proceedings of the 4th Computer Vision Winter Workshop (CVWW'99), Rastenfeld, Austria, 1999, pp. 122-130.

[15]. R. Feris, T. Campos, and R. Cesar, "Detection and Tracking of Facial Features in Video Sequences," Lecture Notes in Artificial Intelligence, Springer-Verlag, Vol. 1793, April 2000, pp. 197-206.

[16]. Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing (2nd Edition), Prentice-Hall, Inc. 2001.

[17]. Rein-Lien Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face Detection in Color Images," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 24, No. 5, May 2002, pp. 696-706.

[18]. S. Karungaru, M. Fukumi, and N. Akamatsu, "Human Face Detection in Visual Scenes using Neural Networks," Transaction of Institute of Electrical Engineers of Japan (IEEJ), Vol. 122c, June 2002, pp. 995-1000.

[19]. The Champion dataset, http://www.libfind.unl.edu/alumni/events/champions.

[20]. Toshiaki Kondo and Hong Yan, "Automatic Human Face Detection and Recognition under Non-uniform Illumination," Vol. 32, Pattern Recognition, 1999, pp. 1707-1718.

[21]. V. Vezhnevets, V. Sazonov, and A. Andreeva, "A Survey on Pixel-Based Skin Color Detection Techniques," Proceedings of Graphicon-2003, 2003, pp. 85-92.

[22]. Y. Suzuki and T. Shibata, "Multiple-clue Face Detection Algorithm using Edge-based Feature Vectors," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004 (ICASSP '04), Vol. 5, 17-21 May 2004, pp. V - 737-740.

[23]. Yanjiang Wang and Baozong Yuan, "A Novel Approach for Human Face Detection from Color Images under Complex Background," Pattern Recognition, Vol. 34, No. 10, 2001, pp. 1983-1992.