Proceedings of 2008 CACS International Automatic Control Conference
National Cheng Kung University, Tainan, Taiwan, Nov. 21-23, 2008

Design and Implementation of A Real-Time Face Recognition System for Mobile Robot

Yen-Shao Chen, Ming-Yuan Shieh*, Guei-Hua Chen

Department of Electrical Engineering,
Southern Taiwan University
Tainan County, Taiwan
**e-mail: myshieh@mail.stut.edu.tw*

***Abstract*—This paper proposes a real-time face recognition system for mobile robot which can recognize the faces of human beings instantly from the images and also identify the users. The proposed system consists of the face detection and the face recognition subsystems. In the face detection subsystem, it uses the facial color filter to partition the possible facial color areas, and then applies the connected component labeling procedures to localize the objective faces initially. After this, an ellipse mask is adopted to verify where the faces are in the image, and then normalize these face blocks for the later faces recognition procedures. In order to recognize the faces, the principal component analysis method aims to determine the eigenvector of each face. After constructing the database of the weight vector for all sample faces, one can use the magnitude of the Euclidean distance between the objective and every sample to determine which face sample is the closest and the resultant one. From experimental results, it is seen that the proposed real-time face recognition system for mobile robots have fast and precise recognition abilities. It also can be applied to access control systems, security systems, the human-machine interaction systems and so on.**

Keywords*—*Face Recognition, Face Detection, Principal Component Analysis, Eigenface, Interaction.

I. INTRODUCTION

In daily life, people always carry a lot of keys and keycards or need to memorize numerous passwords for enter, pass or login in house, office or those places need identification. However, all types of tools for passage are possibly, sometimes easily copied, recreated or cracked except distinguishing characteristics. Such biological characteristics as iris, fingerprint, voice and face have been adopted to identify the human. Among these biological authentication techniques, facial recognition is the most widely used methodology. It is because of faster growing digital signal and image processing technology.

Autonomous mobile robots have a wide variety of applications in the industry as well as our daily life. It is known that mobile robots are helpful for various applications, such as factory automation, robot arm for manufacturing aid, home or public service and many others in industry. Since the robot usually has to cope with incomplete, uncertain, or inaccurate information while it transports materials in a factory, accomplishes tasks in the office, or carrying out chores around the house. Recently, there have been many significant and novel applications using intelligent control systems with learning schemes. These advances have shown that intelligent methods have significantly improved the performance of a variety of mobile robotic systems. Besides, there are many approaches with a focus on security service applications of mobile robots; they aim to introduce some useful facial recognition techniques that may be implemented in robotic systems for human identification.

Face recognition systems usually include three modules, *i.e.* the preprocessing stage, feature extraction, and classification. Most of the approaches may be classified into two categories: one consists of geometric feature-based techniques which rely on the identification of specific components of a face such as eyes, nose, mouth, and distances among them. Another means holistic template-based techniques which are usually based on projecting the high dimensional (original) images onto lower dimensional subspaces spanned by specific basis vectors.

The ways to estimate the features of the faces have been discussed and studied in recent years. They are used to filter the background and detect the faces blocks from a digital image firstly, then to determine their features and generate characteristic vectors, to localize the faces continuously, and to recognize the main face finally. One can easily partition them into two fields as the face detection and the face recognition.

How to determine where the faces blocks is from the image with complex background is one of the most popular approach topics. In most studies, one always performs color space transform firstly, then tries to obtain eigenvectors after removing the background, and finds out the resultant position of the detected face. The necessary techniques in such processes are such as color analysis, pattern comparison, object tracking and neural networks. Hsu *et al.* proposed an algorithm of face detection in color image [1], in which the proposed detection of skin color distribution is based on the method of color analysis.

To well define the face pattern as the reference for face detection is the main idea of pattern comparison. Hongo *et al.* [2] addressed how to search the most matching face from the pattern faces by updating eigen-patterns and performing feature comparison continuously. Miao *et al.* [3] proposed a hierarchical multiscale and multiangle system for human face detection in a complex background using gravity-center template. They describe the frame of a human face based on the patterns of eyes, mouth, nose and outline of face.

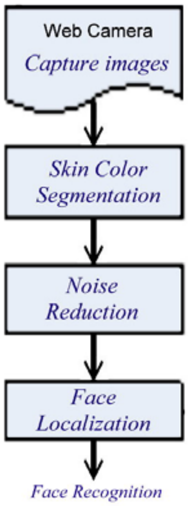


Fig. 1 The configuration of the real-time face detection system

There are a lot of studies focus on change detection and tracking problems. To adopt pyramid transformation technique that is also called as image differencing method [4], one can easily determine where the moving objects are. Besides, Rowley *et al.* [5] addressed a learning scheme of neural network-based face detection, in which the neural network learns and finds out the position of the face by well matching images' features and face distribution.

The most used methods on face recognition consist of wavelet transform, linear discriminant analysis (LDA) and principal component analysis (PCA). In which, wavelet transforms have advantages over traditional Fourier transforms for representing functions that have discontinuities and sharp peaks, and for accurately deconstructing and reconstructing finite, non-periodic and/or non-stationary signals. Generally, an approximation to digital wavelet transform (DWT) is used for data compression if signal is already sampled, and the continuous wavelet transform (CWT) for signal analysis.

Harr wavelet transform [6], a new blur detection scheme, aims to regard every pixel as an independent value firstly, then to add or to subtract these values to determine the wavelet coefficient in frequency domain. There are two standard linear subspace projections of the low-resolution facial images, PCA and LDA, used to distinguish the different styles of images. PCA is basically a compression procedure based on a linear projection technique on a subspace spanned by the principal eigenvectors (those corresponding to the largest eigenvalues) of the input covariance matrix. LDA approach was proposed by Fisher which identifies directions in space along which separation of the projections is maximized. While LDA is not always superior to PCA in terms of recognition accuracy, the PCA+LDA approach [7] has been successfully applied in face recognition applications.

This paper aims to design a real-time face detection and recognition system based on PCA. The authors intend to adopt such simple but useful methods without any complex frequency computation to analyze the detected image then to discriminate the face. It is noted that the experimental results demonstrate the feasibility of the proposed scheme.

II. Real-Time Face Detection System

Fig. 1 introduces the proposed configuration of the real-time face detection system. In the system, there consist of four procedures. The first one is to capture images continuously by using a web camera. These images will be sent to the next procedure for skin color segmentation. The next will be noise reduction. After this procedure, it has filtered out the most range of pixels except the skin color regions. The face regions will be localized in the resultant images by these steps of connected component labeling, edge detection and ellipse shape matching.

A. Skin color segmentation

The images captured by a normal level camera usually have chromatism because of different illumination of the light in every capture. In order to compensate the illumination, one can use the "reference white" technique to revise the color tones near original ones. Generally, the reference white method is to average the pixels' brightness (gray level) values whose brightness within the highest 5%. The brightness of the reference white will be redefined by this average. It is worthy noted that any counted gray level value must be larger than 100 if the gray level value of the original white is 255.

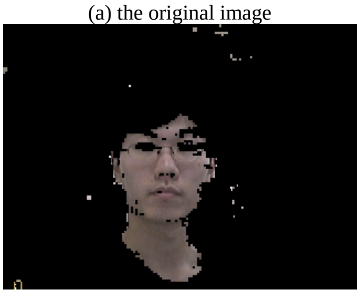
In order to extract the face regions from the image with complex background, a faster and useful way is to focus on skin color regions. However, how to define the range of skin color becomes difficult since the color values in RGB is very sensitive to illumination. That is why the color models such as HSV, YIQ, NCC, YCbCr are adopted in most such approaches. Based on this reason, in this paper, the authors adopt YCbCr color model to define the color. In which, the Y indicates luminance that is respective to value and the Cb and Cr represent blueness and redness respectively that are relative to chroma. This model have good discrimination in color value or chroma. The transform from RGB to YCbCr can be as (1).

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.229 & 0.587 & 0.114 \\ -0.1687 & -0.3313 & 0.5 \\ 0.5 & -0.4187 & -0.0813 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad (1)$$

In order to segment the skin color from the image, one can use color binarization in YCbCr space. In this paper, the Y, Cb and Cr color values of face skin will be defined within the ranges as shown in (2). These thresholds have been revised after many times tests which result in satisfactory skin color segmentation as shown in Fig. 2.

$$\begin{cases} 133 \le Cr \le 173 \\ 77 \le Cb \le 127 \\ R > G > B \\ (R-G) \ge 45 \end{cases} \quad (2)$$

(a) the original image



(b) the image after skin color segmentation

Fig. 2 The processing results of skin color segmentation

B1	B2	B3
B4	B	B5
B6	B7	B8

Fig. 3 The mask of opening operation

B. Noise reduction

The procedure of noise reduction aims to remove those unnecessary pixels such as white noises. One can adopt a lowpass filter (LPF) to eliminate high frequency noises. The practical way is to perform convolution of the image and the mask as shown in (3).

$$L = \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{3}$$

Besides, since there will be a lot of noises with similar skin color existed in the image after binarization. One can remove these noises through opening operation. In this paper, a 3*3 mask as shown in Fig. 3 is used for opening (*i.e.* erosion then dilation) operation. Suppose the set A represents the binary region that needs to be processed and the set B is the mask, opening is the dilation of the erosion of the binary image A by the structuring element B: $A \circ B = (A \ominus B) \oplus B$ where $\ominus$ and $\oplus$ denote erosion and dilation, respectively. Basically, take comparison with closing operation, opening removes small objects while closing removes small holes. These two techniques can also be used to find specific shapes in an image. Opening can also be used to find things into which a specific structuring element can fit (edges, corners, *etc*.)

C. Face localization

After noise reduction, one needs to find out where the desired objects are in the processed image. Using the procedures of connected component labeling [8], the connected pixels will be regarded that belongs to the same objects. One can label these objects and determine their area, height, width and relative data. Next, since the statistics shows that the ratio of height to width of a normal human face is about 1.2. Moreover, the sharp of a human face is similar to an ellipse. That is, one can cut off those areas are possible of the neck or others based on the above two appearances.

In order to detect edge, a popular and useful algorithm is the Sobel operator. Technically, it is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function. Since the edge can be determined by Sobel operation easily, one can sketch the outline of the objective area (the region with skin color pixels) after Sobel operation.

Besides, in order to detect elliptic area in the objective, one can use the elliptic mask as shown in (4).

$$\begin{cases} x_\theta = x_0 + r\cos\theta \\ y_\theta = y_0 + 1.2r\sin\theta \end{cases}, \theta = 0^\circ \sim 360^\circ \tag{4}$$

The elliptic edge detection algorithm will convolute the edge image with elliptical edge image firstly, then find 5~10 pixels which have the maximum values of the average boundary value as (5). Next, to form the possible blocks centered at the points found in the former step; then to computer the mean square error between the possible blocks and the average face. It results in the possibility is in proportion to the mean square error.

$$\phi_g(s) = \frac{1}{N_\sigma}\sum_{i=1}^{N_\sigma}\left|g_s(i)\right| \tag{5}$$

where N_σ is the number of pixels on the elliptic circumference, $g_s(i)$ denotes the boundary value of the ith pixel on the elliptic circumference. The average boundary value, $\phi_g(s)$, can be regarded as the similarity between the detected block and an ellipse. The larger value indicates the block is nearer an ellipse.

The result of the elliptic detection after the procedures of connected component labeling and Sobel edge operation is shown in Fig. 4. From the figure, one can see that the elliptic outline indicates the detected facial region; it means the face localization have obtained an exact result.

III. REAL-TIME FACE RECOGNITION SYSTEM

Fig. 5 shows the proposed configuration of the real-time face recognition system. In the system, there consist of three procedures. The first one is to normalize every processing image in uniform size. These images will be sent to the next procedure for Principal component analysis (PCA). PCA aims to determine the eigenvector of each face. After constructing the database of the weight vector for all sample faces, one can use the magnitude of the Euclidean distance between the objective and every sample to determine which face sample is the closest and the resultant one. The face recognition will real-time identify the subject.

Proceedings of 2008 CACS International Automatic Control Conference
National Cheng Kung University, Tainan, Taiwan, Nov. 21-23, 2008

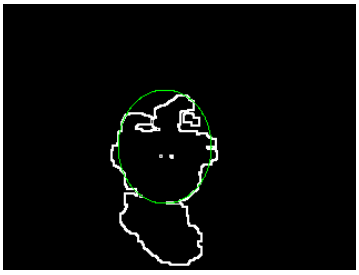


Fig. 4 The detected face after the procedure of face detection

A. Normalization of images

As mention before, robust recognition algorithms dominate the face recognition. However, if all the image data have no uniform standard, the recognition will have no rule could be followed. It possibly results in a lower recognition rate. Therefore, in this paper, all the facial images are zoomed to a uniform size as 80*100 pixels by bilinear interpolation. It will reduce the differences between subjective image and sample images. Fig. 6(a) and Fig. 6(b) illustrate the detected facial images are processed before and after normalization respectively.

Besides, all the normalized facial images in uniform size will be transformed to 8-bit gray-level images with the uniform average 128. Suppose the image has N pixels and I_i represents the gray level value of the ith pixel which can be determined from (6). Since the average I_{avg} (as shown in (7)) would like to be set to 128, the new gray level value of the ith pixel can be modified as I_{mod} as shown in (8). In this way, all the facial images can be compensated their illumination. Fig. 6(b) verifies such effects.

$$I_{gray} = 0.299R + 0.587G + 0.114B \tag{6}$$

$$I_{avg} = \frac{1}{N}\sum_{i=1}^{N} I_i \tag{7}$$

$$I_{\text{modify}} = \left(128 - I_{avg}\right) + I_i \tag{8}$$

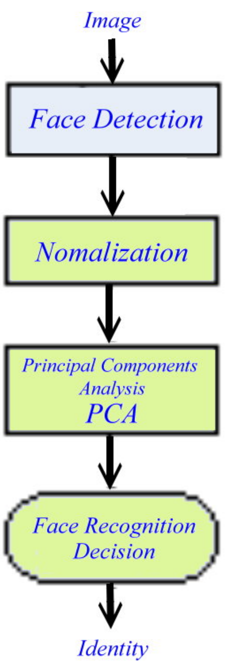
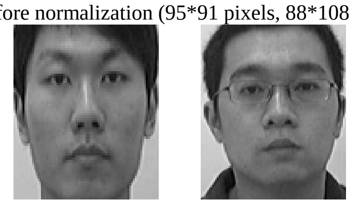


Fig. 5 The configuration of the real-time face recognition system

(a) before normalization (95*91 pixels, 88*108 pixels)



(b) after normalization (80*96 pixels, 80*96 pixels)

Fig. 6 The detected facial images before/after normalization

B. Principal component analysis (PCA)

Principal component analysis (PCA) [9, 10] is a statistical technique used for data reduction. Basically, PCA is a vector space transform and that is the simplest of the true eigenvector-based multivariate analyses. The leading eigenvectors from the eigen decomposition of the correlation or covariance matrix of the variables describe a series of uncorrelated linear combinations of the variables that contain most of the variance. In addition to data reduction, the eigenvectors from a PCA are often inspected to learn more about the underlying structure of the data. PCA is closely related to factor analysis; indeed, some statistical packages deliberately conflate the two techniques. True factor analysis makes different assumptions about the underlying structure and solves eigenvectors of a slightly different matrix.

Although the facial image has been processed in several procedures from skin color segmentation to normalization, however, the resultant image still has a large amount of data necessary to be determined. Since PCA is useful for data reduction, one can define several appropriate eigen variables to represent those data with many variance terms. That is, each set of image data can be replaced by several independent linear combinations of those influential features.

Suppose there are M samples of normalized facial images, where each one is denoted by an N*N array. To rearrange each array into a N^2*1 vector is as shown in Fig. 7. M samples have M vectors, denoted as $\Gamma_1, \Gamma_2, \ldots, \Gamma_M$. To average them, we get the average vector as

$$\Psi = \frac{1}{M}\sum_{i=1}^{M} \Gamma_i \,. \tag{9}$$

It can be regarded as the vector of the mean face that represents the common parts in all samples. While remove the common parts from each sample, one can determine each difference image as

$$\Phi_i = \Gamma_i - \Psi, \quad i = 1 \sim M \,. \tag{10}$$

Assume $A = [\Phi_1, \Phi_2, \ldots, \Phi_M]$, the covariance matrix of all facial samples, C, will be

$$C = \frac{1}{M}\sum_{i=1}^{M} \Phi_i \Phi_i^T = AA^T \,. \tag{11}$$

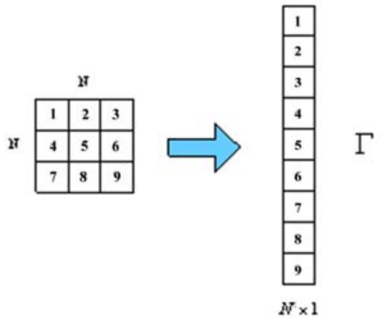


Fig. 7 The transformation from an N*N image array to a N^2*1vector

On can determine the eigenvalues and eigenvectors of C by

$$Cu_i = \lambda_i u_i \,. \tag{12}$$

From (11), one can see that the matrix A of N^2*M indicates the matrix C of N^2*N^2. It is noted that to determine the eigenvalues and the eigenvectors of such large matrix is consuming. Thus, one can compute the eigenvalues and the eigenvectors of the matrix A^TA, as shown in (13), firstly because its dimension is only M*M.

$$A^T A v_i = \varsigma_i v_i \,. \tag{13}$$

Since $C = AA^T$ from (11), to multiply both sides of (13) by the matrix A, we have

$$AA^T A v_i = \varsigma_i A v_i \,. \tag{14}$$

In comparison with (12) and (14), while

$$\begin{cases} u_i = A v_i \\ \lambda_i = \varsigma_i \end{cases}, \tag{15}$$

one can determine the eigenvalues and the eigenvectors of C from (13) and (15). The eigenvector u_i denotes the eigenface of ith sample face as shown in (16).

$$u_i = \sum_{j=1}^{M} \Phi_j v_{ij} \,. \tag{16}$$

If we sort the eigenvalues of all samples and then arrange the relative eigenvectors into an eigenspace. Determine the weighting matrix W of the eigenspace as

$$\begin{cases} \Omega_i = u_i^T(\Gamma_i - \Psi) = u_i^T \Phi_i \\ W = [\Omega_1, \Omega_2, ..., \Omega_M] \end{cases}. \tag{17}$$

The vectors in the matrix W can be regarded as trained image data of a sample. Each sample image has a relative matrix W. While finish the computation of W for all samples, it means the training of the face image database has been accomplished.

C. Decision-making principles of face recognition

In mathematics, the Euclidean distance is the ordinary distance between two points that is used to measure with a ruler. In recognition applications, the Euclidean distance can be regarded as the differences between two vectors which measures what differences between two compared data (images). The Euclidean distance between two vectors $v_1 = (a_1, a_2, ...a_k)$ and $v_2 = (b_1, b_2, ...b_k)$ will be

$$d = \sqrt{\sum_{i=1}^{k} (a_i - b_i)^2} \,. \tag{18}$$

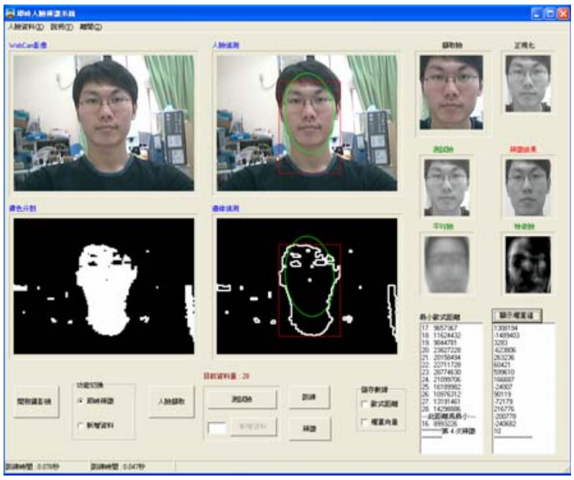


Fig. 8 The screen of the proposed system interface

In practical image recognition, one can determine the weighting vector W_a of the tested image Γ_a firstly, then compare it with the weighting vector of every trained sample in database respectively to get the relative Euclidean distance. One can say the sample with the minimum distance from the tested image, d in (19), will be regarded as the most matching face for the test. However, if the minimum is larger than a threshold, it means the tested face does not match any sample and possibly the system never meets the face before. In this status, the face image will be recorded as a new sample to strengthen the recognition database.

$$d = \min \left\| W_a - W_i \right\|, \ i = 1 \sim k \,. \tag{19}$$

IV. EXPERIMENTAL RESULTS

In this paper, we assume the processed image is of 320*240 pixels, 24 bits and BMP file format. The system interface, as illustrated in Fig. 8, will show every processed image which consist of the ones after skin color segmentation, Sobel edge detection, elliptic detection or face recognition. Besides, the data of the Euclidean distance and the relative weighting vector are still shown in the lower right of the screen.

Fig. 9 shows the results of the face detection. In which, although there are close, far, oblique, upward and downward facial images, however, the results of location detection are satisfactory. It has verified the feasibility of the proposed face detection system.

For face recognition, we initially adopt 5 images under different postures of every person among 7 members of our laboratory as training samples. After set up these samples, every member will be tested again based on the trained database. Table 1 illustrates the Euclidean distances between every two tested faces. As shown in the first row, the first member's tested face is nearest to his sample face in database. Similar results are existed in other rows. Especially, a recognition computation only costs average 0.045 sec. Fig. 10 displays the results shown in the interface for face recognition. These all indicate that the proposed system really can real-time identify the human faces.

Table 1. The European distances of all tests on face recognition

Sample face / Tested face	Member1 Sample	Member2 Sample	Member3 Sample	Member4 Sample	Member5 Sample	Member6 Sample	Member7 Sample
Member1	76,520	239,439	177,547	190,501	80,547	219,462	154,780
Member2	156,264	68,357	186,214	210,651	193,762	175,603	223,184
Member3	109,623	180,777	85,709	109,345	131,448	219,355	127,122
Member4	195,472	191,167	127,735	54,110	218,879	248,360	120,775
Member5	96,459	201,020	131,931	170,765	83,228	189,868	115,712
Member6	223,049	164,075	143,503	337,248	244,024	103,726	220,077
Member7	179,526	198,791	184,219	178,683	172,749	158,166	94,521
Average recognition time: 0.045 秒							

Fig. 9 The resultant images after face detection

Fig. 10 The results of face recognition shown in the interface

V. Conclusions

The paper has successfully accomplished a real-time face detection and recognition system. It could be applied to robotic system, access control system, security system, the human-machine interaction system and so on. From the experimental results, one can see that the proposed system can fast but exactly detect and recognize the human face even in complex background. Besides, since the decision-making algorithms proposed in this paper all are simple and useful. It indicates the system is able to be easily extended to any application system.

Acknowledgment

This work was supported by the National Science Council, Taiwan, under grant number NSC96-2221-E-218-039-MY3.

References

[1] R. L. Hsu, A. M. Mohamed, A. K. Jain, “Face Detection in Color Image,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.24, No.5, pp. 696-706, 2002.

[2] H. Hongo, A. Murata, K. Yamamoto, “Consumer Products User Interface Using Face and Eye Orientation,” *Proceedings of IEEE International Symposium on Consumer Electronics*, pp.87-90, 1997.

[3] J. Miao, B. Yin, K. Wang, L. Shen, X. Chen, “A hierarchical Multiscale and Multiangle System for Human Face Detection in a Complex Background Using Gravity-Center Template,” *Pattern Recognition*, Vol.32, pp.1237-1248, 1999.

[4] C. Anderson, B. Peter, G. Wal, “Change detection and tracking using pyramid transformation techniques,” *Proceedings of SPIE - Intelligent Robots and Computer Vision*, Vol. 5, pp.72-78, 1985.

[5] H. A. Rowley, S. Baluja, T. Kanade, “Neural Network-Based Face Detection,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.20, No. 1, pp. 23-38, 1998.

[6] W. S. Lee, H. J. Lee, J. H. Chung, “Wavelet-Based FLD for Face Recognition,” *Proceedings of the 43rd IEEE Midwest Symposium on Circuits and System*, Vol.2, pp.734-737, 2000.

[7] H. P. Graf, E. Chen, E. Cosatto, “Locating Faces and Facial Parts,” *Proceedings of First International Workshop Automatic Face and Gesture Recognition*, pp.41-46, 1995.

[8] L. S. Shapiro, G. C. Stockman, Computer Vision, Prentice-Hall, NJ., 2001.

[9] M. Kirby, L. Sirovich, “Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces,” *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol.12, No.1, 1990.

[10] M. Turk, A. Pentland, “Eigenfaces for Recognition,” *Journal of Cognitive Neuroscience*, Vol.3, No.1, pp.71-86, 1991.