Component-based LDA Face Descriptor for Image Retrieval

Tae-Kyun Kim, Hyunwoo Kim, Wonjun Hwang, Seok Cheol Kee, Jong Ha Lee
Human Computer Interaction Lab.
Samsung AIT
Republic of Korea
{taekyun,hwkim}@sait.samsung.co.kr

Abstract

We present a component-based face descriptor with LDA (Linear Discriminant Analysis) and a simple pose classification. Our algorithm has been developed to deal with face image retrieval in huge database such as those in internet environments. Such retrieval requires a compact face descriptor and an efficient recognition algorithm that is robust to variations in lighting and facial poses. Partitioning of a face image into components facilitates the development of an efficient and robust algorithm as follows. First, compensation for light and pose variations is much more easily done on individual components than on the whole image. Second, pose variation is compensated by classifying facial pose and aligning facial components. Finally, LDA is more effective at the component level which has simplified statistics than the whole image. Experimental results on MPEG-7 database show an impressive accuracy of our algorithm compared with conventional LDA methods.

1. Introduction

In video processing and analysis, the human face is a key component for visual discrimination and identification. Since the early 1990s, many methods for face recognition and facial expression analysis have been extensively developed. Recently, face descriptors for MPEG-7 have been proposed for face retrieval in video streams [1-7]. The face descriptor should meet the following requirements. The descriptor associated with each face image should be extracted without the prior knowledge about which group of images belongs to the same person. Each image in the data set should be used as a query image in order to retrieve the other images of the same person from the data set. The ground truth of each image is the set of the other images of the person in the query image [1]. A challenging problem is to retrieve face images with large variations in lighting and pose. The descriptor should be compact and learn statistics only from the training images of other persons. Several descriptions have been presented [2-7].

To compensate image variation due to illumination change, Wang and Tan proposed the 2^{nd}-order Eigenface method [2] and Kamei and Yamada extended their work to use

a confidence factor describing face symmetry and intensity variation due to illumination change [5]. Kim et al. developed the 2nd-order PCA Mixture Model (PMM) method [4]. The 2nd-order approaches attempt to remove effects due to the change in illumination by removing the component of the image lying in the subspaces spanned by the first few eigenvectors. However, the approaches seem to be weak under pose variation because they are describing a holistic pixel distribution, which is vulnerable to pose change.

To compensate image variation due to pose change as well as illumination change, Nefian and Davies used the DCT-based embedded Hidden Markov Model (eHMM) for face description [3], while Kim et. al. proposed eHMM method with the 2nd-order Block-specific Eigenvectors [7]. The eHMM algorithm deals with pose variation using embedded states corresponding to facial regions implicitly and segmenting an observation image into overlapping blocks, but it may resort to local minima if the initial solution is not close to the global minima. Wiskott et al. [8] developed Gabor wavelet based algorithm called elastic bunch graph matching. These algorithms are, however, computationally expensive. In the face retrieval where the descriptor should be extracted for each face image without any prior knowledge of the same person, eHMM-based methods have been found to have poor performance.

In this paper, we propose a new approach dealing with pose and illumination variation with a very efficient face description in terms of both accuracy and size. We introduce a component-based LDA face representation. Most closely related work is Heisele et. al.'s algorithm [9]. They detect facial components, and their grayscale values are concatenated into a single feature vector. Then, SVM (Support Vector Machine) algorithm is applied to the collection of feature vectors and the extracted support vectors are used for classification. Although they show that their component-based algorithm can simplify SVM classifier giving better accuracy, SVM is very time consuming for huge databases and facial component detection is very difficult in natural environments.

Our algorithm, however, combines the component-based representation with LDA and simple pose classification, resulting in a compact face descriptor with a high accuracy, called the 'component-based LDA face descriptor.' First, to simplify image statistics, we adopt the component-based scheme in which a face image is separated into several facial components. To compensate the effect of pose variation, the components are then aligned by calculating translation offsets between the corresponding components. In addition, to compensate the effect of illumination variation, the components are encoded by LDA. The combination of component based representation and LDA effectively solves the problems of face retrieval and person identification.

Section 2 describes the component-based approach and Section 3 reviews LDA. The component-based LDA approach is presented in the next Section. Experimental results and conclusions are presented in Section 5 and 6, respectively.

2. Component-based Representation

A face image of our descriptor is represented component by component. We separate a face image into several facial components corresponding to forehead, eyes, nose and mouth. Compared with the holistic image representation, it is more robust to illumination and/or pose variation in face encoding, and it has flexibility in similarity matching and in alignment adjustment.

First, image variation due to pose and/or illumination change within each component patch is smaller than that in a whole image space, simplifing the pre-processing [6]. Generally, holistic approaches based on PCA/ICA/LDA (Principal

Component Analysis/ Independent Component Analysis/ Linear Discriminant Analysis) encode the greyscale correlation among every pixel position statistically and image variation due to lighting and camera geometry results in severe change of face representations. Since our component-based scheme encodes the facial components separately, image variations are limited to each component region. Most of all, the pre-processing within small patches is easier than that in the whole image region. Because a facial component has less statistical complexity than the whole face image, the linear encoding like PCA/ICA/LDA in a component region becomes more suitable than that for the whole face region. In addition, separated facial components have partial overlaps with neighboring components, preserving the component adjacency relationships important for personal identification. Experimental result shows that the component encoding followed by even simple sum of matching scores of components outperforms holistic encoding methods in person identification.

Second, a facial component with large variation is weighted less in the matching stage. In matching stage, since each facial component can be considered as a separate classifier, the outputs can be weighted by its discriminability and a priori knowledge. Similarity, in the component scheme, face occlusion such as wearing sunglasses or masks can be more easily identified and dealt with for person identification. For example, sunglass or mask patterns are trained and it is compared with facial components corresponding to eyes or mouth for discrimination. Furthermore, when the component positions are well aligned by facial component detection or dense matching methods, the pose variation can be compensated, resulting in a further accuracy improvement. In [9], the recognition accuracy was improved after component alignment.

Third, for pose compensation, dense optical flow or global projective/affine transformation is needed in the whole image representation, while translational offset can be enough in the component-based representation. For detail, refer to Section 2.1. Figure 1 shows an example of facial component separation, where separation is fixed relatively to the eye positions.

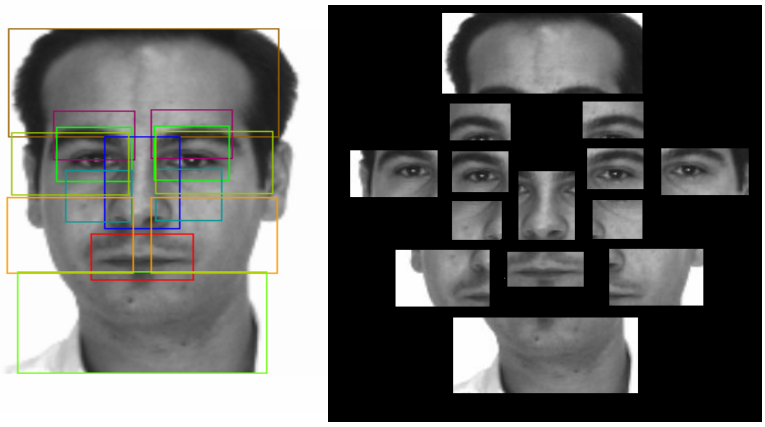

Figure 1. An example of facial component separation.

2.1 Pose Classification and Component Alignment

As already mentioned, when the component positions are well aligned by component detection or dense matching methods, the pose variation can be compensated and it may result in further accuracy improvement. However, the dense matching or perfect

component detection is very difficult and need heavy computation. In this work, we efficiently compensate pose variation by combining a pose classification technique and 2D translation estimation of components.

First, let us consider pose classification stage. During training the eye positions are given for each face image. The face images are then five pose sets – frontal, left, right, up, and down – by manual clustering. From the set of each pose class, eigenfaces are extracted by PCA. During the pose classification stage a test image is projected into the five different Eigen-subspaces corresponding to the first a few eigenfaces of each pose class. The image is classified into the class with the smallest projection error [13].

Next, a pose compensation stage follows. Generally, according to assigned pose class, images corresponding to front faces are selected as references and those of other poses are warped to the references using affine/projective transformations. The warping needs heavy computation and mis-classification results in accuracy degradation. However, when we considered facial component patches, which are smaller than a whole face region, the translation offsets can approximate affine/projective transformation. The translation offsets can be computed from the warping of the average positions of fiducial points on faces in the training data set. Therefore, without dense matching or warping, the components are aligned and image variation due to pose change is largely removed. Note that there is a trade-off between the size/number of facial component patches and the size of descriptor.

After the removal of pose variation through component alignment, the corresponding components is encoded for description and used for similarity computation, resulting in a higher face-recognition accuracy.

3. Linear Discriminant Analysis (LDA)

When the training data set are labeled for each identity, supervised learning techniques like LDA are more profitable for face feature extraction compared with methods of unsupervised learning. When we apply the supervised learning LDA, we can remove the illumination variation and pose variation as well in encoding. LDA still keeps identity information.

LDA or Fisher's Linear Discriminant (FLD) is a class specific method in the sense that it represents data to make it useful for classification [10,11]. Given a set of N images $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ with each image belongs to one of c classes $\{X_1, X_2, \ldots, X_c\}$, LDA selects a linear transformation matrix $\mathbf{W}$ in such a way that the ratio of the between-class scatter and the within-class scatter is maximized.

Mathematically, the between-class scatter matrix and the within-class scatter matrix are defined by

$$\mathbf{S}_B = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T \qquad (1)$$

and

$$\mathbf{S}_W = \sum_{i=1}^{c} \sum_{\mathbf{x}_k \in X_i} (\mathbf{x}_k - \mu_i)(\mathbf{x}_k - \mu_i)^T , \qquad (2)$$

respectively, where μ_i denotes the mean image of class X_i, μ denotes the mean image of entire data set, and N_i denotes the number of images in class X_i. If the

within-class scatter matrix $\mathbf{S}_W$ is not singular, LDA finds an orthonormal matrix $\mathbf{W}_{opt}$ maximizing the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix. That is, the LDA projection matrix is represented by

$$\mathbf{W}_{opt} = \arg\max_{\mathbf{W}} \frac{\left|\mathbf{W}^T \mathbf{S}_B \mathbf{W}\right|}{\left|\mathbf{W}^T \mathbf{S}_W \mathbf{W}\right|} = \left[\mathbf{w}_1 \quad \mathbf{w}_2 \quad \cdots \quad \mathbf{w}_m\right]. \quad (3)$$

The set of the solution $\{\mathbf{w}_i \mid i = 1,2,\ldots,m\}$ is that of generalized eigenvectors of $\mathbf{S}_B$ and $\mathbf{S}_W$ corresponding to the m largest eigenvalues $\{\lambda_i \mid i = 1,2,\ldots,m\}$, i.e., $\mathbf{S}_B \mathbf{w}_i = \lambda_i \mathbf{S}_W \mathbf{w}_i, i = 1,2,\ldots,m$. Generally, to overcome the singularity of $\mathbf{S}_W$, PCA first reduces the vector dimension before applying LDA. The each LDA feature vector is represented by the vector projections $\mathbf{y}_k = \mathbf{W}_{opt}{}^T \mathbf{x}_k$, $k = 1,2,\ldots,N$.

4. Component-based LDA Face Descriptor

To take advantage of both the good linear property and robustness to image variation of the component-based approach, we combine LDA with the component-based representation. LDA is applied to the separated facial components separately and this improves the accuracy. In this proposal, LDA applied to the whole face is called 'the holistic LDA method' and LDA applied to the components is called 'the component-based LDA method'. We can also combine both methods and this is called 'the combined LDA method'. Mathematically, for the holistic LDA method, a face image $\mathbf{x}$ is represented by a LDA feature vector $\mathbf{y}^0 = \left(\mathbf{W}^0\right)^T \mathbf{x}$ with a LDA transformation matrix $\mathbf{W}^0 \equiv \mathbf{W}_{opt}$.

4.1 Face Description

First, for training data set, the LDA transformation matrix is extracted. Given a set of N training images $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, all the images are separated into L facial components by the facial component separation algorithm. All patches of each component are gathered together and are represented in vector form; the k-th component is denoted as $\{\mathbf{z}_1{}^k, \cdots, \mathbf{z}_N{}^k\}$. Then, for the set of each component, a LDA transformation matrix is trained. For the k-th facial component, the corresponding LDA matrix $\mathbf{W}^k$ is computed. Finally, we store the set of LDA transformation matrices, $\{\mathbf{W}^1, \cdots, \mathbf{W}^L\}$, to be used for the test stage.

In test data set, the L vectors $\{\mathbf{z}^1, \cdots, \mathbf{z}^L\}$ corresponding to facial component patches are extracted from a face image $\mathbf{x}$. A set of LDA feature vectors $\mathbf{y} = \{\mathbf{y}^1, \cdots, \mathbf{y}^L\}$ is extracted by transforming the component vectors by the

corresponding LDA transformation matrices, respectively. The feature vectors are computed by

$$\mathbf{y}^k = (\mathbf{W}^k)^T \mathbf{z}^k,\ k = 1,2,\ldots,L. \tag{4}$$

Therefore, for the component-based LDA method, a face image $\mathbf{x}$ is compactly represented by a set of LDA feature vectors $\{\mathbf{y}^1,\cdots,\mathbf{y}^L\}$, and, for the combined LDA method, a set of LDA feature vectors $\{\mathbf{y}^0;\mathbf{y}^1,\cdots,\mathbf{y}^L\}$. Note that for the holistic LDA method, a face image $\mathbf{x}$ is represented by a LDA feature vector $\mathbf{y}^0$.

4.2 Similarity Matching

Given two face images $\mathbf{x}_1, \mathbf{x}_2$ represented by LDA feature vector set $\mathbf{y}_1, \mathbf{y}_2$ the similarity $d(\mathbf{y}_1, \mathbf{y}_2)$ for the component-based LDA method is measured by weighted sum of cross-correlations between the corresponding components as

$$d(\mathbf{y}_1, \mathbf{y}_2) = \frac{1}{L}\sum_{k=1}^{L} w_k \frac{\mathbf{y}_1^k \cdot \mathbf{y}_2^k}{\left\|\mathbf{y}_1^k\right\| \left\|\mathbf{y}_2^k\right\|}, \tag{5}$$

where $\mathbf{y}_1^k, \mathbf{y}_2^k$ denote the LDA feature vectors of the k th facial component of the face image $\mathbf{x}_1, \mathbf{x}_2$, respectively, and w_k denotes weighting factors of the k th facial component. For the combined LDA method, k starts from 0 instead of 1, and for the holistic LDA method, k has only 0.

5. Experimental Results and Discussion

Database

The experimental face database consists of 3175 images of 635 faces (5 images of each face). The images in the database are manually normalized in 46x56 pixels^2 giving fixed eye positions. Some of the images are selected from well-known public databases: 133*5 images from AR DB, 15*5 from Yale DB, 40*5 from ORL DB, 30*5 from Bern DB and 122*5 from FERET. The 223*5 images in the database are taken under lighting variation (light set), and 412*5 images at different view angles (pose set).

Protocol of experiments

Three different experiments were performed to show the accuracy and generalization performance of algorithms. In Experiment 1, 200 images (5 images of 40 persons), each half from a light set or pose set, were used for training and the others were used for the test. Experiment 2 has 800 training images, which consisting of 5 images of 160 persons from both light and pose sets, and 2375 test images. In Experiment 3, approximately a half, 1685 images (5 images of 337 persons) of all database images were used for training and the other half (5 images of 298 persons) for the test.

Feature Selection Scheme

The class discriminability of basis vectors defined in (1) was calculated for the training set and the best combination of the k most discriminable basis vectors were chosen.

$$r = \frac{\sigma_B}{\sigma_W} \tag{6}$$

where $\sigma_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)^2$, $\sigma_W = \sum_{i=1}^{c} \sum_{\mathbf{x}_k \in X_i} (\mathbf{x}_k - \mu_i)^2$.

The order of the class discriminability of vectors was approximately identical to the sequence of eigenvalues of the vectors on LDA. In the proposed component scheme, the same number of bases for each component was chosen. Variation of the ratio of component basis numbers did not largely affect the performance.

Component Weighting Scheme

For combining components, the weighted sum rule of cross-correlation between the corresponding components was adopted. The weight of each component was determined proportionally to square of a reciprocal of FIR (False Identification Rate) on the training set. In the combined LDA, the holistic LDA and the result of total component LDA have a similar weight. In these experiments, the components around forehead were dominant in recognition and this may be because the data set does not include large variations over time relative to illumination and pose changes. Fixed hair-styles of people provided consistency of face images over time.

Generalization test: holistic vs. component LDA

The cumulative FIR (False Identification Rate) graphs of the holistic LDA and component LDA are shown in the Figure 2 (a), (b), and (c). Note that the component LDA highly outperforms the holistic LDA in the case of small training data like Experiment 1 and 2. It has similar performance in Experiment 3, where a half of data set for training and the other half for the test. The holistic LDA over-learned from the training set in all the cases giving a poor generalization. As shown in Figure 3, most of the important facial information appears at a certain region of faces. Intensity variations of the holistic LDA basis images are around forehead, eyes of faces. If the test faces had more discriminative information in other parts, the learned basis vectors would not represent faces effectively. Compared to the holistic approach, the component LDA learns evenly from the whole region of a face by selecting the same number of basis vectors of all separated components.

Although the component scheme encodes a face image with the benefit of good linear property and robustness to image variations in the components, it lacks information between the components. To overcome this problem, the combined LDA has been proposed. Figure 2 (d) shows that the proposed combine method of the holistic and component LDA improves the performance in the first rank FIR dramatically. The first rank FIRs of the holistic and combined method are 0.0355 and 0.0208, respectively.

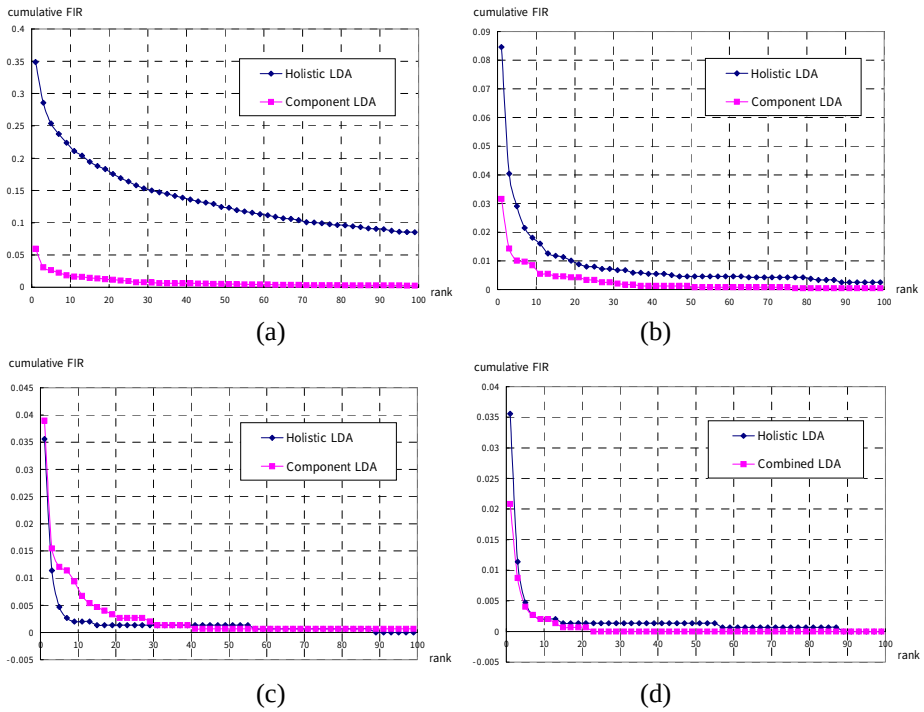

Figure 2. Cumulative FIR plots.
(a) Experiment 1. (b) Experiment 2. (c) and (d) Experiment 3.

Figure 3. The first 10 basis images of the holistic LDA.

Computational Complexity and Size of Descriptor

Table 1 compares our approach with the holistic LDA method in terms of computational complexity and size of descriptor. In feature extraction, the component LDA is approximately half times simpler than the holistic LDA while the combined LDA is comparable with the holistic LDA. In matching complexity and size of descriptors, the component LDA and the combined LDA use 2.5 and 3.5 times more computations, compared with the holistic approach, but give better generalization performances.

Table 1. Computational complexity and descriptor size.

		Holistic LDA	Component LDA	Combined LDA
Feature	Additions	N_0 *(N-1)=103000	5* N_1 *(N_{avg} - 1)=38100	N_0 *(N-1)+ 5* N_1 *(N_{avg} –1)=141100

Extraction Complexity	Multiplications	N_0*N=103040	5* N_1*N_{avg}=38200	N_0*N+5* N_1*N_{avg}=141240
Matching Complexity	Additions	3*(N_0-1) = 117	5*3*(N_1-1)+4=289	5*3*(N_1-1)+5+3*(N_0-1)=407
	Multiplications	3*N_0 = 120	5*(3*N_1+1)=305	5*(3*N_1+1)+ 3*N_0 +1=426
Size of Descriptor in Bits		40*4	100*4	140*4

N_0 : the number of elements of a holistic feature vector(=40), N_1 : the number of elements of one component feature vector(=20) , N : holistic input image size(=46*56), N_{avg} : average size of component input image (=382).

6. Concluding Remarks

In this paper, we proposed the component-based LDA face descriptor with simple pose compensation. It showed higher retrieval accuracy compared with the conventional LDA method with small size of descriptor and low computational complexity. Experimental results gave better generalization performance than the conventional LDA approach, and it shows the feasibility in very huge database. Our future work will be to develop a more sophisticated classifier for our LDA projections and to extend the size of current data set to find and solve problems in huge database.

References

[1] M. Abdel-Mottaleb, J. H. Connell, R. M. Bolle, and R. Chellappa, “Face descriptor syntex,” Merging proposals P181, P551, and P650, ISO/MPEG m5207, Melbourne, 1999.

[2] L. Wang, and T. K. Tan, “Experimental Results of Face Description Based on the 2nd-order Eigenface Method,” ISO/IEC JTC1/SC21/WG11/M6001, Geneva, May 2000.

[3] A. Nefian, and B. Davies, “Standard Support for Automatic Face Recognition,” ISO/IEC JTC1/SC21/WG11/M7251, Sydney, July 2001. Also appeared in A. Nefian and M. Hayes, “An embedded hmm-based approach for face detection and recognition,” In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 6, pp. 3553–3556, 1999.

[4] H. C. Kim, D. Kim, S. Y. Bang, Y. S. Choi, and J. Kim, “Proposal for Face Description Using 2nd-order PCA Mixture Model,” ISO/IEC JTC1/SC21/WG11 M7286, Sydney, July 2001.

[5] T. Kamei, and A.Yamada, “Extension of the Face Recognition Descriptor Using a Confidence Factor,” ISO/IEC JTC1/SC21/WG11 M7689, Pattaya, December 2001.

[6] Toshio Kamei, Akio Yamada, “Proposal of the Face Recognition Descriptor based on Fourier spectral Principal Component Analysis,” ISO/IEC JTC1/SC21/WG11 M7953, Juju Island, March 2002.

[7] Daijin Kim, Min-Sub Kim, Sung Yang Bang, Sang Youn Lee, Young Sik Choi, “Face Recognition Descriptor Using the Embedded HMM with the 2nd-order Block-specific Eigenvectors,” ISO/IEC JTC1/SC21/WG11 M7997, Juju Island, March 2002.

[8] L. Wiskott, J.-M. Fellous, N. Kr¨uger, and C. von der Malsburg. “Face recognition by elastic bunch graph matching,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 775 –779, 1997.
[9] B. Heisele, P. Ho, and T. Poggio, “Face Recognition with Support Vector Machines: Global versus Component-based Approach,” In Proc. IEEE International Conference on Computer Vision, 2001.
[10] M. S. Bartlett, “Face Image Analysis by Unsupervised Learning,” Kluwer Academic Publishers, 2001.
[11] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” IEEE Trans. On Pattern Recognition and Machine Intelligence, Vol. 19, No. 7, pp. 711- 720, July 1997.
[12] A.M. Martinez, A.C. Kak, “PCA versus LDA,” IEEE Trans. On Pattern Recognition and Machine Intelligence, Vol. 23, No. 3, pp. 228-233, July 1997.
[13] B. Moghaddam, A. Pentland, “Face Recognition using View-based and Modular Eigenspaces,” SPIE Automatic Systems for the Identification and Inspection of Humans, Vol. 2277, July 1994.