MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
3-22

# Face Recognition using Local Features based on Two-layer Block Model

Wonjun Hwang[1] Ji-Yeun Kim Seokcheol Kee
Computing Lab.,
Samsung Advanced Institute of Technology

## Abstract

*This paper proposes the novel face recognition algorithm, local features based on two-layer block model, in order to improve the generalization ability of the face recognition. The traditional LDA scheme is often unstable even though it is the popular extraction technique for face recognition. In this paper, we focus on the performance stability when we have tests whose property is different from trained variations. Local Feature Analysis is adopted to transfer a face image into several local block representations by different block models, and Linear Discriminant Analysis is used to increase the discriminant power of separated block representations. The method was tested on three different face database and the system was shown to perform very well when compared to traditional approach.*

## 1 Introduction

Face recognition has been widely studied because it is the essential technology in biometrics, video surveillance, multimedia retrieval system and etc. A lot of approaches [1][2][3][4] have been proposed to improve the performance of face recognition recently. Linear Discriminant Analysis (LDA) [3] is one of the popular approaches in the field of face recognition. It is basically designed to increase the discriminatory power made by a linear transformation which maximizes the between-class scatter while minimizing the within-class scatter. The dimension of a face image is still high and for solving the small sample size problem it could be reduced by Principal Component Analysis (PCA) [2]. Other [5] has tried to solve this problem with direct LDA which discard the null space of between-class scatter, but on the other hand it keeps the null space of within-class scatter that has important information for classification. Though many algorithms related to LDA have been proposed, one of unsolved problems in LDA is easily biased to the variation of training set - overfitting problem.

LFA [4] is also popular representation algorithm when it achieved the good result in FERET test [6]. LFA could derive local topographic representations for a face image and give a description of image in terms of statistically local features and their positions. To apply this local feature to face recognition, recently LFA and LDA are combined by Yang and etc [7]. The sparsification of LFA helps the reduction of dimension of image in LDA scheme and local topological property is more useful than holistic property of PCA in recognition, but there is still structural problem because the method to select the features is designed for minimization of reconstruction error, not for increasing discriminability in face model.

In this paper, we proposed the novel recognition algorithm to merge LFA and LDA method. We do not use the existing sparsification method for selecting features but adopt the two-layer block model to make several groups with topographic local features in similar position. Each local block, flocked local features, can represent its own local property and at the same time holistic face information. Flocks of local features can easily solve the small sample size problem in LDA without discarding unselected local features, and LDA scheme can extract the important information for recognition not in focus of representation. Moreover, we can extract lots of vectors on separated viewpoint from different layer model in one face image and they have the property robust to environmental changes and overfitting problem as compared with limited number of features vectors.

The rest of this paper is organized as follows: the brief description on LFA and LDA is explained in Section 2.1 and Section 2.2, respectively and proposed algorithm - local feature based on two-layer block model is given in Section 2.3. The experimental results are given in Section 3. Conclusion is summarized in Section 4.

## 2 LFA and LDA Method based on Two-Layer Block Model

### 2.1 Theory of local feature analysis

A topographic representation based on second-order image dependencies called local features analysis (LFA) was developed by Penev and Atick [4]. Local feature analysis can makes a set of topographic and local kernels that are optimally matched to the second-order statistics of the input ensemble. Local features are basically derived from principal component eigenvectors, and consist of sphering principal component eigenvalues to equalize their variance.

Suppose that we are given a set of $M$ training images, $\varphi_i$ , $i$ =1,…, $M$ , each represented by an $N$ - dimensional vector obtained by a raster scan. The mean

[1] Address: Mt. 14-1, Nongseo-Ri, Gihueng-Eup, Yongin-Si, Gyeonggi-Do, Korea. E-mail: wj.hwang@samsung.com

vector of the image set is defined by $m = (1/M)\cdot\sum_{i=1}^{M}\varphi_i$. After subtracting the mean vector form all images, $x_i = \varphi_i - m$, we can construct an zero-mean matrix, $\mathbf{X} = [x_1,...,x_M]$ and the covariance matrix $\mathbf{S} = \mathbf{X}\cdot\mathbf{X}^T$. Eigenvector, $\mathbf{P}$, and eigenvalue, $\mathbf{D}$, are calculated by eigen analysis, $\mathbf{S} = \mathbf{P}\cdot\mathbf{D}\cdot\mathbf{P}^T$. Thus we can define a set of kernels, $\mathbf{K}$ as

$$\mathbf{K} = \mathbf{P}\cdot\mathbf{V}\cdot\mathbf{P}^T, \tag{1}$$

$$\mathbf{V} = diag(F_i/\sqrt{\lambda_i}),\ i = 1,\ldots,N, \tag{2}$$

where $\lambda_i$ are the $i$ th eigenvalues of covariance, $\mathbf{S}$. Low-pass noise filtering is performed with $F_i = \lambda_i/(\lambda_i + n^2)$ and $n$ is 0.25 in this paper. The output kernel matrix can be represented by

$$\mathbf{K} = [k_1,...,k_N], \tag{3}$$

where the columns of $\mathbf{K}$ contain the spatially local properties, and are topographic in the sense that they are indexed by spatial location as shown in Figure. 1.

Figure 1. Sample images of local features, eyebrows, nose, part around eye, cheek, and jaw.

The sparsification of LFA has been tried by some papers [4][7] because of residual correlations in the output. They reduced the dimensionality of representation by choosing a single set of kernels to minimize the difference between origin image and reconstructed image, but it is the method to address image representation. It is not assured that the selected kernels that play an important rule in reconstruction scheme are not always good in recognition scheme.

### 2.2 Theory of linear discriminant analysis

LDA [3] is a supervised learning method that uses second-order statistics to find a projection into a subspace that maximizes the between-class scatter while minimizing the within-class scatter of the projected data. A typical LDA training is carried out via two scatter matrix analyses - the between-class scatter matrix $\mathbf{S}_B$ and the within-class scatters matrices $\mathbf{S}_W$:

$$\mathbf{S}_B = \sum_{i=1}^{c} M_i (m_i - m)(m_i - m)^T, \tag{4}$$

$$\mathbf{S}_W = \sum_{i=1}^{c}\sum_{\varphi_k \in c_i} (\varphi_k - m_i)(\varphi_k - m_i)^T, \tag{5}$$

where $m_i$ is the mean image of $i$ th class $c_i$ with $M_i$ samples and $c$ is the total number of classes. The projection vector, $W$, to satisfy the basic concept of LDA is made by following equations:

$$W = \arg\max_{W} \frac{|W^T \mathbf{S}_B W|}{|W^T \mathbf{S}_W W|}. \tag{6}$$

Generally, PCA first reduces the vector dimension before applying LDA to overcome the singularity of within-class scatter matrix. The performance of LDA directly depends on the number of used PCA [8] and the best number of PCA is smaller than the dimension of input image. That is, the discriminability of the input image should be represented by some limited PCA bands and it often causes the overfitting problem in LDA scheme when the size of training set is especially small because (1) the discriminant property in whole face images can not be spanned by a handful of input vectors for LDA scheme and (2) when training the small size of training set in supervised learning scheme, the result should be easily tuned to only property of training samples. Thus it can not support different property with limited features in different test sets.

### 2.3 Proposed Algorithm – Local feature based on two-layer block model

In this paper, we apply two-layer block model for grouping the local kernels while others [9][10] applying to divide the input images into several blocks, and apply the LDA scheme to each group in order to decide which region is important for recognition instead of the sparsification. Each kernel group could emphasize its own specific local block of an image. Moreover, local feature and holistic feature could be represented at the same time. In case of another local analysis, component scheme [10], it is possible that local minimum problem occur because we have only local property. For example, open mouth component might be not equal to close one in same person but possibly equal to open one in different person, but in this paper we can overcome this local minimum problem because proposed method have the holistic characteristic, needless to say, it is feebler than the local property.

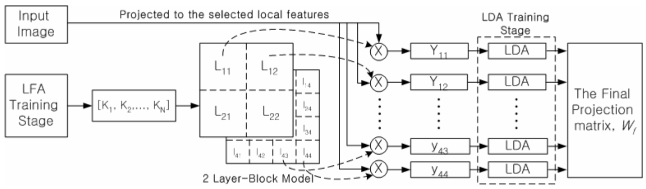


Figure 2. Concept of two-layer block model

The proposed model basically consists of two layers as shown in Figure 2. 1st-layer block model consists of 4 blocks $L_{11}$, $L_{12}$, $L_{21}$, and $L_{22}$ which has *N/4* local kernels, respectively, and 2nd-layer block model has 16 blocks, $l_{11}$, $l_{12}$,…., and $l_{44}$. Each block contains *N/16* kernels. The spatial notation of LFA kernel is given from Equation (3)

$$K(u,v) = K_{u+v\times w}, \tag{7}$$

where $w$ and $h$ are width and height of an image, respectively and $(u, v)$ is the spatial position in an image. Thus 1st-layer block model, for example, could be represented as follows;

$$L_{11} = \{K(u,v) \mid 1 \le u \le w/2,\ 1 \le v \le h/2\}$$
$$\vdots \qquad (8)$$
$$L_{22} = \{K(u,v) \mid (w/2+1) \le u \le w,\ (h/2+1) \le v \le h\}$$

In $1^{st}$-layer block model, a flocked kernel matrix $L_{xy}$ transforms zero-mean matrix $X$ to the LFA output;

$$Y_{xy} = L_{xy}^T X\ . \qquad (9)$$

The output matrix $Y_{xy}$ can be thereafter used as the input of LDA to increase the discriminability power and to reduce the feature vectors. The between-class scatter matrix and the within-class scatter matrix of chose *(x,y)* block are defined as

$$\mathbf{S}_{xy}^B = \sum_{i=1}^{c} M_i (m_{xy}^i - m_{xy})(m_{xy}^i - m_{xy})^T \quad \text{and}$$

$$\mathbf{S}_{xy}^W = \sum_{i=1}^{c} \sum_{Y_{xy}^k \in c_i} (Y_{xy}^k - m_{xy}^i)(Y_{xy}^k - m_{xy}^i)^T\ , \qquad (10)$$

where $m_{xy}^i$ is the mean vector of $Y_{xy}^i$ in $i$ th class $c_i$, $m_{xy}$ is the total mean vector of output matrix $Y_{xy}$ and $c$ is the total number of classes. The discriminant transformation matrix in this stage is $W_{xy}$, $R^{N/4} \rightarrow R^{c-1}$, and projection function of $1^{st}$-layer is derived by

$$W_{xy} = \arg\max_{W_{xy}} \frac{|W_{xy}^T \mathbf{S}_{xy}^B W_{xy}|}{|W_{xy}^T \mathbf{S}_{xy}^W W_{xy}|}\ . \qquad (11)$$

To present the discriminant facial component with compact bit-rate, we can discard the redundant features $R^{c-1} \rightarrow R^{k_1}$. Feature vector in $1^{st}$-layer block model is represent by

$$f_{xy}^1 = W_{xy}^{\ T} (L_{xy}^T(\varphi - m)) = (L_{xy}W_{xy})^T(\varphi - m)$$
$$= V_{xy}^T(\varphi - m)\ . \qquad (12)$$

The transformation set $V$ in $1^{st}$-layer block model is composed of

$$V = [(L_{11}W_{11}),(L_{12}W_{12}),\ldots,(L_{22}W_{22})]$$
$$= [V_{11},V_{12},\ldots,V_{22}]\ . \qquad (13)$$

In $2^{nd}$-layer block model, we can get the transformation set by the same way. The LFA output can be obtained from $y_{xy} = l_{xy}^T X$, and the discriminant transformation matrix, $w_{xy}$, is also obtained. The dimension of feature vector, $f_{xy}^2$, in $2^{nd}$-layer block model is reduced like $R^{N/16} \rightarrow R^{k_2}$. The optimal transformation set is defined

$$v = [(l_{11}w_{11}),(l_{12}w_{12}),\ldots,(l_{44}w_{44})]\ . \qquad (14)$$

The final projection function of local feature based on two-layer block model is represented by

$$f_i = W_f^T(\varphi_i - m) \qquad (15)$$

where $W_f = [V,v]$ is the final projection set, the number of feature vector is $(2 \times 2)\cdot k_1 + (4 \times 4)\cdot k_2$ and it is always bigger than the number of feature vector in PCA+LDA. The example basis images of final projection function are shown at Figure 3.

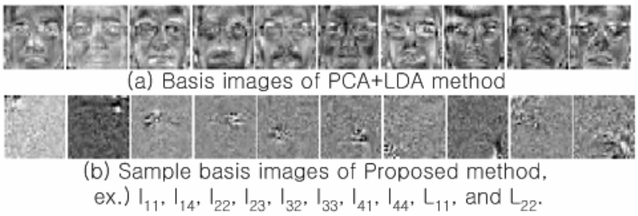


Figure 3. Sample basis images of PCA+LDA and the proposed method

We can compute the similarity between probe image $i$ and gallery image $j$ by normalized correlation and its equation is as follows

$$s(i,j) = (f_i \cdot f_j) / (\| f_i \| \cdot \| f_j \|)\ . \qquad (16)$$

## 3 Experimental results

We evaluated our algorithm on four different subsets – “Light subset” and “Pose subset” from CMU PIE database [11], moreover XM2VTS database [12] and SAIT database. In detail, “Light subset” has only 1,496 frontal face images with neutral illumination on. “Pose subset” has 1,020 images with a neutral expression under neutral illumination and its pose variation is restricted within ±22.5° in this experiment. XM2VTS has 2,360 frontal images with 4 different sessions. We built the SAIT database for one year consists of 500 individuals×5 images There are variations in illumination changes, expression changes and time elapse. All images were normalized by manual eye positions, resized to 32×32 pixels, and cropped to exclude the background as shown at Figure. 4.

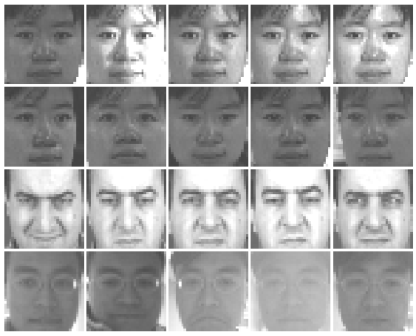

Figure 4. This is a figure. Light subset, pose subset, Illumination subset, XM2VTS, and SAIT database

We randomly select 34 individuals from respectively “Light subset” and “Pose subset,” as a training set to obtain the proper subspaces. The rest 34 subjects from respectively “Light subset” and “Pose subset” are used as a test set, and additionally “XM2VTS,” and “SAIT” database are used as only test set. In this paper we use rank order statistics displayed graphically as a “Cumulative Match Characteristic” (CMC) curve [13] as a measure of performance of face recognition.

Table 1 shows overall identification ratio of each method. Feature number of PCA+LDA-I as well as PCA+LDA-II is 33, but proposed method uses 660 features [$33\times4+33\times16$]. PCA+LDA-I and -II are the traditional PCA+LDA with different parameters. PCA+LDA-II is overtuned to training variation in PIE DB. There are big performance differences between PCA+LDA-I and -II in PIE Light test sets, but such increase is not achieved in XM2VTS test set. In other words, it is possible that the traditional PCA+LDA is easily overfitted to training variation. On the other hand, proposed method always shows better result in all test sets and the performance increases in XM2VTS and SAIT test sets are noticeable.

**Table 1**. Comparison of system performance of four different test databases. The rank 1 identification rate is written.

| | Trained variation | | Out of trained variation | |
|---|---|---|---|---|
| | PIE-Light | PIE-Pose | XM2VTS | SAIT |
| PCA+ LDA-I | 36.61 | 17.47 | 47.90 | 31.96 |
| PCA+ LDA-II | 98.54 | 24.97 | 48.92 | 49.32 |
| Proposed Method | 99.86 | 29.73 | 59.00 | 59.26 |

# 4 Conclusion

In this paper, local features based on two-layer block model were proposed for the representation of face images. By the proposed algorithm, we can have following advantages in face recognition. (1) As the number of the usable feature vectors increases in small numbers of training sets, we can represent the face model with sufficient dimension in comparison to the PCA+LDA method. Therefore our proposed system can cope with the overfitting problem. That is, the performance degradation in out of trained variation is relieved. (2) We can have a chance to analyze the local information as well as the holistic information in face model with the accentuated local block feature. (3) We can construct two different feature spaces in one person which are extracted in different scopes. For example, 1$^{st}$-layer block model is considered as a low-frequency analysis, and 2$^{nd}$-layer block model is regarded as a high-frequency analysis.

The experimental results show that the proposed description is better in face recognition as compared with the traditional PCA+LDA methods. Moreover, the proposed method can be operated well under out of trained variation.

# References

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face Recognition: A Literature Survey,” ACM Computing Survey, Vol. 36, No. 4, pp. 339-458, Dec., 2003.

[2] M. A. Turk and A. P. Pentland, “Eigenfaces for recognition,” Journal of Cognitive Neuroscience, Vol. 3, No. 1, pp. 71-86, 1991.

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenface vs. Fisherfaces: Recognition using class specific linear projection,” IEEE Trans. PAMI, vol. 19, no. 7, pp. 711-720, Jul., 1997.

[4] P. S. Penev and J. J. Atick, “Local Feature Analysis: A general statistical theory for object representation,” Network: Computation in Neural Systems, Vol. 7, No. 3, pp. 477-500, 1996.

[5] H. Yu and J. Yang, “A direct LDA algorithm for high-dimensional data – with application to face recognition,” Pattern Recognition, Vol. 34, No. 10, pp. 2067-2070, 2001.

[6] P.J. Phillips, H. Wechsler, J. Juang, and P. Rauss, “The feret database and evaluation procedure for face-recognition algorithms,” Image and Vision Computing Journal, Vol. 16, No. 5, pp. 295-306, 1998.

[7] Q. Yang, X. Ding and Z. Chen, “Discriminant Local Feature Analysis of Facial Images,” IEEE Proc., ICIP, Spain, Sep., 2003.

[8] X. Wang and X. Tang, “Random Sampling LDA for Face Recognition,” IEEE Proc., CVPR, Washington, D.C., USA, Jun. 2004.

[9] W. Hwang and H. Ko, “Real-time vehicle recognition using local feature extraction,” Electronics Letters, Vol. 37, No. 7, pp. 424-425, Mar, 2001.

[10] T. Kim, H. Kim, W. Hwang, S. Kee, and J. Kittler, "Independent Component Analysis in a Facial Local Residue Space," IEEE Proc., CVPR, Madison, USA, July, 2003.

[11] T. Sim, S. Baker and M. Bsat, “The CMU Pose, Illumination, and Expression (PIE) Database,” International Conference on Automatic Face and Gesture Recognition, pp. 53-58, May, 2002.

[12] K. Messer, J. Matas, J. Kittler and K. Jonsson, "XM2VTSDB: The Extended M2VTS Database," Audio and Video-based Biometric Person Authentication, pp. 72-77, Washington, D.C., Mar, 1999.

[13] J.M Bone, J.L. Wayman, and D.M. Blackburn, “Evaluating Facial Recognition Technology for Drug Control Applications,” ONDCP International Technology Symposium, pp. 26-28, Jun., 2001.