Line-based PCA and LDA approaches for Face Recognition

Vo Dinh Minh Nhat, and Sungyoung Lee

Kyung Hee University – South of Korea
{vdmnhat, sylee}@oslab.khu.ac.kr

Abstract. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) techniques are important and well-developed area of image recognition and to date many linear discrimination methods have been put forward. Despite these efforts, there persist in the traditional PCA and LDA some weaknesses. In this paper, we propose a new Line-based methodes called Line-based PCA and Line-based LDA that can outperform the traditional PCA and LDA methods. As opposed to conventional PCA and LDA, those new approaches are based on 2D matrices rather than 1D vectors. That is, we firstly divide the original image into blocks. Then, we transform the image into a vector of blocks. By using row vector to represent each block, we can get the new matrix which is the representation of the image. Finally PCA and LDA can be applied directly on these matrices. In contrast to the covariance matrices of traditional PCA and LDA approaches, the size of the image covariance matrices using new approaches are much smaller. As a result, those new approaches have three important advantages over traditional ones. First, it is easier to evaluate the covariance matrix accurately. Second, less time is required to determine the corresponding eigenvectors. And finally, block size could be changed to get the best results. Experiment results show our method achieves better performance in comparison with the other methods.(1)

Index Terms – Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Face Recognition.

1. Introduction

Face recognition research has been started in the late 70s and is one of the active and exciting researches in computer science and information technology areas since 1990 [1]. Generally, there are three phases for face recognition, mainly face representation, face detection, and face identification. Face representation is the first task, that is, how to model a face. The way to represent a face determines the successive algorithms of detection and identification. There are a variety of

(1) This research was supported by the MIC (Ministry of Information and Communication), Korea, under the ITRC(Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Assessment)
Corresponding Authors : Vo Dinh Minh Nhat (vo_dinhminhnhat@yahoo.com), and SungYoung Lee (sylee@oslab.khu.ac.kr)

approaches for face representation, which can be roughly classified into three categories: template-based, feature-based, and appearance-based. The simplest template-matching approaches represent a whole face using a single template, i.e., a 2-D array of intensity, which is usually an edge map of the original face image. In a more complex way of template-matching, multiple templates may be used for each face to account for recognition from different viewpoints. Another important variation is to employ a set of smaller facial feature templates that correspond to eyes, nose, and mouth, for a single viewpoint. The most attractive advantage of template-matching is the simplicity, however, it suffers from large memory requirement and inefficient matching. In feature-based approaches, geometric features, such as position and width of eyes, nose, and mouth, eyebrow's thickness and arches, face breadth, or invariant moments, are extracted to represent a face. Feature-based approaches have smaller memory requirement and a higher recognition speed than template-based ones do. They are particularly useful for face scale normalization and 3D head model-based pose estimation. However, perfect extraction of features is shown to be difficult in implementation. Eigenfaces approach is one of the earliest appearance-based face recognition methods, which was developed by M. Turk and A. Pentland [2] in 1991. This method utilizes the idea of the PCA and decomposes face images into a small set of characteristic feature images called eigenfaces. Recognition is performed by projecting a new face onto a low dimensional linear “face space” defined by the eigenfaces, followed by computing the distance between the resultant position in the face space and those of known face classes.

The Fisherface method [4] combines PCA and the Fisher criterion [9] to extract the information that discriminates between the classes of a sample set. It is a most representative method of LDA. Nevertheless, Martinez *et al*. demonstrated that when the training data set is small, the Eigenface method outperforms the Fisherface method [7]. Should the latter be outperformed by the former? This provoked a variety of explanations. Liu *et al.* thought that it might have been because the Fisherface method uses all the principal components, but the components with the small eigenvalues correspond to high-frequency components and usually encode noise [11], leading to recognition results that are less than ideal. In line with this theory, they presented two enhanced Fisher linear discrimination (FLD) models (EFMs) [11] and an enhanced Fisher classifier [12] for face recognition. Their experiential explanation lacks sufficient theoretical demonstration, however, and EFM does not provide an automatic strategy for selecting the components. Chen *et al.* proved that the null space of the within-class scatter matrix contains the most discriminative information when a small sample size problem takes place [13]. Their method is also inadequate, however, as it does not use any of the information outside the null space. In [5], Yu *et al.* propose a direct LDA (DLDA) approach to solve this problem. It removes the null space of the between-class scatter matrix firstly by doing eigen-analysis. Then a simultaneous diagonalization procedure is used to seek the optimal discriminant vectors in the subspace of the between-class scatter matrix. However, in this method, removing the null space of the between-class scatter matrix by dimensionality reduction would indirectly lead to the losing of the null space of the within-class scatter matrix which contains considerable discriminative information. Rui Huang [10] proposed the method in which the null space of total scatter matrix which has been proved to be the common null space of both between-class and within-class

scatter matrix, and useless for discrimination, is firstly removed. Then in the lower-dimensional projected space, the null space of the resulting within-class scatter matrix is calculated. This lower-dimensional null space, combined with the previous projection, represents a subspace of the whole null space of within-class scatter matrix, and is really useful for discrimination. The optimal discriminant vectors of LDA are derived from it.

However, in the previous PCA and LDA-based face recognition techniques, the 2D face image matrices must be previously transformed into 1D image vectors. The resulting image vectors of faces usually lead to a high dimensional image vector space, where it is difficult to evaluate the covariance matrices accurately due to its large size and the relatively small number of training samples. Fortunately, the eigenvectors can be calculated efficiently using the SVD techniques and the process of generating these covariance matrices is actually avoided. However, this does not imply that the eigenvectors can be evaluated accurately in this way since the eigenvectors are statistically determined by the between-class and within-class covariance matrices, no matter what method is adopted for obtaining them. In this paper, new approaches called Line-based PCA and Line-based LDA are developed for image feature extraction. As opposed to conventional PCA and LDA, Line-based PCA and Line-based LDA is based on 2D matrices rather than 1D vectors. That is, we firstly divide the original image into blocks. Then, we transform the image into a vector of blocks. By using row vector to represent each block, we can get the new matrix which is the representation of the image. Finally PCA and LDA can be applied directly on these matrices. In contrast to the covariance matrices of traditional PCA and LDA approaches, the size of the image covariance matrices using new approaches are much smaller. As a result, those new approaches have three important advantages over traditional ones. First, it is easier to evaluate the covariance matrix accurately. Second, less time is required to determine the corresponding eigenvectors. And finally, block size could be changed to get the best results. The remainder of this paper is organized as follows: In Section 2, the traditional PCA and LDA methods are reviewed. The idea of the proposed methods and their algorithms are described in Section 3. In Section 4, experimental results are presented on the ORL face image database to demonstrate the effectiveness of our methods. Finally, conclusions are presented in Section 5.

2. PCA and LDA

Let us consider a set of N sample images $\{x_1, x_2, ..., x_N\}$ taking values in an n-dimensional image space, and assume that each image belongs to one of c classes $\{C_1, C_2, ..., C_c\}$. Let N_i be the number of the samples in class $C_i (i = 1, 2, ..., c)$, $\mu_i = \frac{1}{N_i} \sum_{x \in C_i} x$ be the mean of the samples in class C_i, $\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$ be the mean of all samples. A linear transformation maps the original n-dimensional image

space into an *m*-dimensional feature space, where *m* < *n*. The new feature vectors $y_k \in \mathbb{R}^m$ are defined by the following linear transformation :

$$y_k = W^T x_k \tag{1}$$

where $k = 1, 2, ..., N$ and $W \in \mathbb{R}^{nxm}$ is a matrix with orthonormal columns.

If the total scatter matrix is defined as

$$S_T = \frac{1}{N} AA^T = \frac{1}{N} \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T \tag{2}$$

Then after applying the linear transformation W^T, the scatter of the transformed feature vectors $\{y_1, y_2, ..., y_N\}$ is $W^T S_T W$. In PCA, the projection W_{opt} is chosen to maximize the determinant of the total scatter matrix of the projected samples, i.e.,

$$W_{opt} = \arg\max_W \left| W^T S_T W \right| = [w_1 w_2 ... w_m] \tag{3}$$

where $\{w_i \,|\, i = 1, 2, ..., m\}$ is the set of *n*-dimensional eigenvectors of S_T corresponding to the *m* largest eigenvalues.

In LDA, the projection W_{opt} is chosen to maximize the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples, i.e.,

$$W_{opt} = \arg\max_W \frac{\left| W^T S_b W \right|}{\left| W^T S_w W \right|} = [w_1 w_2 ... w_m] \tag{4}$$

where $\{w_i \,|\, i = 1, 2, ..., m\}$ is the set of generalized eigenvectors of S_b and S_w corresponding to the m largest generalized eigenvalues $\{\lambda_i \,|\, i = 1, 2, ..., m\}$, i.e.,

$$S_b w_i = \lambda_i S_w w_i \quad i = 1, 2, ..., m \tag{5}$$

with the between-class scatter matrix S_b is defined as

$$S_b = \frac{1}{N} \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T = \frac{1}{N} \Phi_b \Phi_b^T \tag{6}$$

and the within-class scatter matrix S_w is defined as

$$S_w = \frac{1}{N}\sum_{i=1}^{c}\sum_{x_k \in C_i}(x_k - \mu_i)(x_k - \mu_i)^T = \frac{1}{N}\Phi_w \Phi_w^T \quad \textbf{(7)}$$

3. Line-based PCA and LDA approaches

In the traditional PCA and LDA-based face recognition techniques, the 2D face image matrices must be previously transformed directly into 1D image vectors. The resulting image vectors of faces usually lead to a high dimensional image vector space. However, in our proposed approaches, we firstly divides the original image into $s = hxw$ size blocks with h, w are the height and width of the block. Then, we transform the image into a vector of blocks. By using row vector r with $r^T \in \mathbb{R}^s$ to represent each block (actually, each block is a line of the raw image, so we call these approaches line-based ones), we can get the matrix $X \in \mathbb{R}^{kxs}$ which is the representation of the image , with k is the number of blocks. See fig. 1 for the process.

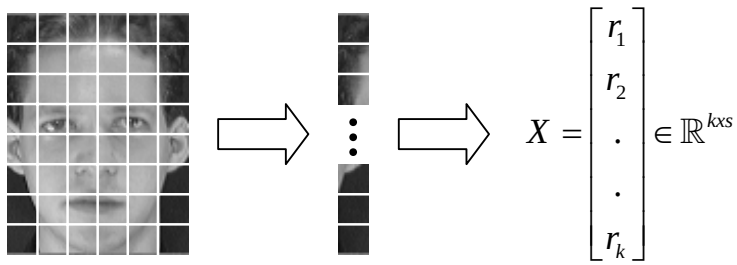

Fig. 1. The process of getting representation of each image

Now, set of N sample images are represented as $\{X_1, X_2, ..., X_N\}$ with $X_i \in \mathbb{R}^{kxs}$. Then the between-class scatter matrix S_b is re-defined as

$$S_b = \frac{1}{N}\sum_{i=1}^{c} N_i(\mu_{C_i} - \mu_X)(\mu_{C_i} - \mu_X)^T \quad \textbf{(8)}$$

and the within-class scatter matrix S_w is re-defined as

$$S_w = \frac{1}{N}\sum_{i=1}^{c}\sum_{X_k \in C_i}(X_k - \mu_{C_i})(x_k - \mu_{C_i})^T \quad \textbf{(9)}$$

The total scatter matrix is re-defined as

$$S_T = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_X)(X_i - \mu_X)^T \quad \textbf{(10)}$$

with $\mu_X = \sum_{i=1}^{N} X_i \in \mathbb{R}^{kxs}$ is the mean image of all samples and $\mu_{C_i} = \frac{1}{N_i}\sum_{X \in C_i} X$ be the mean of the samples in class C_i.

Similarly, a linear transformation mapping the original kxs image space into an mxs feature space, where $m < k$. The new feature matrices $Y_i \in \mathbb{R}^{mxs}$ are defined by the following linear transformation :

$$Y_i = W^T X_i \in \mathbb{R}^{mxs} \quad \textbf{(11)}$$

where $i = 1, 2, ..., N$ and $W \in \mathbb{R}^{kxm}$ is a matrix with orthonormal columns.

In Line-based PCA approach, the projection W_{opt} is chosen with the criterion same as that in (3), and criterion (4) is used in case of Line-based LDA approach.

After a transformation by Line-based PCA or Line-based LDA, a feature matrix is obtained for each image. Then, a nearest neighbor classifier is used for classification. Here, the distance between two arbitrary feature matrices Y_i and Y_j is defined by using Euclidean distance as follows :

$$d(Y_i, Y_j) = \sqrt{\sum_{u=1}^{k}\sum_{v=1}^{s}(Y_i(u,v) - Y_j(u,v))^2} \quad \textbf{(12)}$$

Given a test sample Y_t, if $d(Y_t, Y_c) = \min_j d(Y_t, Y_j)$, then the resulting decision is Y_t belongs to the same class as Y_c.

4. Experimental results

This section evaluates the performance of our propoped algorithms Line-based PCA and Line-based LDA compared with that of the original PCA and LDA algorithms based on using ORL. In the ORL database, there are ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement).

In our experiments, firstly we tested the recognition rates with different number of training samples. $k(k = 2,3,4,5)$ images of each subject are randomly selected from the database for training and the remaining images of each subject for testing. For each value of k, 5 runs are performed with different random partition between training set and testing set. The block size 3 by 3 is used in this first experiment with Line-based PCA and Line-based LDA. Table 1. shows the recognition results of the best recognition accuracy among all the dimension of feature vectors. It means we test on all dimension of feature vectors and choose the best recognition accuracy.

Table 1. The recognition rates on ORL database with different training samples of four methods (PCA, LDA, Line-based PCA – 3x3 block size, Line-based LDA – 3x3 block size)

Training samples	2	3	4	5
PCA (Eigenfaces)	83.4	87.07	89.3	91.5
LDA (Fisherfaces)	78.83	86.9	91.03	93.6
Line-based PCA	85.55	89.05	92.72	95.03
Line-based LDA	86.52	89.12	94.23	95.94

Next, we try to test our approaches when the block size is changed. And several results can be shown in the Table 2 & 3. The same protocol as previous experiments, we choose the recognition result of the dimension feature vectors which give the best accuracy.

Table 2. The recognition rates with different block sizes of Line-based PCA .

	Training samples			
Size of block	2	3	4	5
[2x2]	85.97	89.65	93.60	95.68
[3x3]	85.55	89.05	92.72	95.03
[5x5]	87.31	90.16	94.06	96.04
[10x2]	85.29	89.79	92.70	95.06
[10x10]	82.67	87.54	89.92	92.79

Table 3. The recognition rates with different block sizes of Line-based LDA .

	Training samples			
Size of block	2	3	4	5
[2x2]	86.77	90.4	94.23	96.48
[3x3]	86.52	89.12	94.23	95.94
[5x5]	87.88	90.92	94.86	96.89
[10x2]	86.17	90.41	93.6	95.98
[10x10]	83.51	88.52	90.55	93.57

From Table 2&3, it seems to be that the block size 5x5 give the best recognition results among all. However we still not yet find the relationship between the block size and the recognition result.

5. Conclusions

New methods for face recognition have been proposed in this paper. As opposed to conventional PCA and LDA, Line-based PCA and Line-based LDA is based on 2D matrices rather than 1D vectors. That is, we firstly divide the original image into blocks. Then, we transform the image into a vector of blocks. By using row vector to represent each block, we can get the new matrix which is the representation of the image. Finally PCA and LDA can be applied directly on these matrices. In contrast to the covariance matrices of traditional PCA and LDA approaches, the size of the image covariance matrices using new approaches are much smaller. As a result, those new approaches have three important advantages over traditional ones. First, it is easier to evaluate the covariance matrix accurately. Second, less time is required to determine the corresponding eigenvectors. And finally, block size could be changed to get the best results.

Reference

[1] W. Zhao, R. Chellappa, A. Rosenfeld, J. Phillips: Face Recognition: A Literature Survey. Technical Report, CFAR-TR00-948, Univ. of Maryland, 2000. (Revised 2002)

[2] M. Turk, A Pentland: Eigenfaces for recognition. Journal of Cognitive Neuroscience, Vol. 3 (1991) 71-86.

[3] W. Zhao, R. Chellappa, P.J. Phillips: Subspace Linear Discriminant Analysis for Face Recognition. Technical Report CAR-TR-914, 1999.

[4] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman: Eigenfaces vs. fisherface: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Machine Intell., vol. 19 (1997) 711–720.

[5] H. Yu, J. Yang: A direct LDA algorithm for high-dimensional data with application to face recognition. Pattern Recognit., vol. 34 (2001) 2067–2070.

[6] M. Loog, R. P. W. Duin, R. Haeb-Umbach: Multiclass linear dimension reduction by weighted pairwise fisher criteria. IEEE Trans. Pattern Anal. Machine Intell., vol. 23 (2001) 762–766.

[7] A. M. Martinez, A. C. Kak: PCA versus LDA. IEEE Trans. Pattern Anal. Machine Intell., vol. 23 (2001) 228–233.

[8] D. H. Foley, J. W. Sammon: An optimal set of discrimination vectors. IEEE Trans. Comput., vol. C-24 (1975) 281–289.

[9] R. A. Fisher: The use of multiple measurements in taxonomic problems. Ann. Eugenics, vol. 7 (1936) 178–188.

[10] Rui Huang, Qingshan Liu, Hanqing Lu, Songde Ma: Solving the small sample size problem of LDA. Pattern Recognition, 2002. Proceedings. 16th International Conference on , Vol 3 (2002)

[11] C. Liu, H. Wechsler: Robust coding scheme for indexing and retrieval from large face databases. IEEE Trans. Image Processing, vol. 9 (2000) 132–137.

[12] Chengjun Liu, Wechsler H.: A shape- and texture-based enhanced Fisher classifier for face recognition. IEEE Trans. Image Processing, vol. 10(2001) 598–608.

[13] L. Chen, H. M. Liao, M. Ko, J. Lin, G. Yu: A new LDA-based face recognition system which can solve the small sample size problem. Pattern Recognit., vol. 33 (2000) 1713–1726.