New feature extraction approaches for face recognition

Vo Dinh Minh Nhat and Sungyoung Lee

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

Abstract. All the traditional PCA-based and LDA-based methods are based on the analysis of vectors. So, it is difficult to evaluate the covariance matrices in such a high-dimensional vector space. Recently, two-dimensional PCA (2DPCA) and two-dimensional LDA (2DLDA) have been proposed in which image covariance matrices can be constructed directly using original image matrices. In contrast to the covariance matrices of traditional 1D approaches (PCA and LDA), the size of the image covariance matrices using 2D approaches (2DPCA and 2DLDA) are much smaller. As a result, it is easier to evaluate the covariance matrices accurately and computation cost is reduced. However, a drawback of 2D approaches is that it needs more coefficients than traditional approaches for image representation. Thus, 2D approach needs more memory to store its features and costs more time to calculate distance (similarity) in classification phase. In this paper, we develop a new image feature extraction methods called *two-stage 2D* subspace approaches to overcome the disadvantage of 2DPCA and 2DLDA. The initial idea of *two-stage 2D* subspace approaches which consist of *two-stage* 2DPCA and *two-stage* 2DLDA is to perform 2DPCA or 2DLDA twice: the first one is in horizontal direction and the second is in vertical direction. After the two sequential 2D transforms, the discriminant information is compacted into the up-left corner of the image. Experiment results show our methods achieve better performance in comparison with the other approaches with the lower computation cost.

Index Terms – Principle component analysis (PCA), Linear Discriminant Analysis (LDA), Face Recognition.

1. Introduction

PRINCIPAL component analysis (PCA), also known as Karhunen-Loeve expansion, is a classical feature extraction and data representation technique widely used in the

(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)

areas of pattern recognition and computer vision. Within this context, Turk and Pentland [1] presented the well-known Eigenfaces method for face recognition in 1991. Since then, PCA has been widely investigated and has become one of the most successful approaches in face recognition. However, PCA could not capture even the simplest invariance unless this information is explicitly provided in the training data. It also cannot make full use of pattern separability information like the Fisher criterion, and its recognition effect is not ideal when the size of the sample set is large.

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. In [14], a common vector for each individual class is obtained by removing all the features that are in the direction of the eigenvectors corresponding to the nonzero eigenvalues of the scatter matrix of its own class. The common vectors are then used for recognition. In their case, instead of using a given class's own scatter matrix, they use the within-class scatter matrix of all classes to obtain the common vectors.

However, all the previous traditional PCA-based and LDA-based methods are based on the analysis of vectors. So, it is difficult to evaluate the covariance matrices in such a high-dimensional vector space. Recently, two-dimensional PCA (2DPCA)

[15] and two-dimensional LDA (2DLDA) [16] have been proposed in which image covariance matrices can be constructed directly using original image matrices. In contrast to the covariance matrices of traditional 1D approaches (PCA and LDA), the size of the image covariance matrices using 2D approaches (2DPCA and 2DLDA) are much smaller. As a result, it is easier to evaluate the covariance matrices accurately and computation cost is reduced. However, a drawback of 2D approaches is that it needs more coefficients than traditional approaches for image representation. Thus, 2D approach needs more memory to store its features and costs more time to calculate distance (similarity) in classification phase. In this paper, we develop a new image feature extraction methods called *two-stage 2D* subspace approaches to overcome the disadvantage of 2DPCA and 2DLDA. The initial idea of *two-stage 2D* subspace approaches which consist of *two-stage* 2DPCA and *two-stage* 2DLDA is to perform 2DPCA or 2DLDA twice: the first one is in horizontal direction and the second is in vertical direction. After the two sequential 2D transforms, the discriminant information is compacted into the up-left corner of the image. The remainder of this paper is organized as follows: In Section 2, the traditional PCA and LDA methods are reviewed. 2DPCA, 2DLDA and the proposed methods are described in Section 3. In Section 4, experimental results are presented for the ORL face image databases to demonstrate the effectiveness of our methods. Finally, conclusions are presented in Section 5.

2. PCA and LDA

One approach to coping with the problem of excessive dimensionality of the image space is to reduce the dimensionality by combining features. Linear combinations are particular, attractive because they are simple to compute and analytically tractable. In effect, linear methods project the high-dimensional data onto a lower dimensional subspace.

Suppose that we have *N* sample images $\{x_1, x_2, ..., x_N\}$ taking values in an *n*-dimensional image space. Let us also consider a linear transformation mapping 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.

Different objective functions will yield different algorithms with different properties. PCA aims to extract a subspace in which the variance is maximized. Its objective function is as follows:

$$W_{opt} = [w_1 w_2 ... w_m] = \arg\max_W \left| W^T S_T W \right| \tag{2}$$

with the total scatter matrix is defined as

$$S_T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T \qquad (3)$$

and $\mu \in \mathbb{R}^n$ is the mean image of all samples.

The optimal projection $W_{opt} = [w_1 w_2 ... w_m]$ is the set of *n*-dimensional eigenvectors of S_T corresponding to the *m* largest eigenvalues.

While PCA seeks directions that are efficient for representation, Linear Discriminant Analysis seeks directions that are efficient for discrimination. 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 X_i, $\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$ be the mean of all samples. Then the between-class scatter matrix S_b is defined as

$$S_b = \frac{1}{N} \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T = \frac{1}{N} \Phi_b \Phi_b^T \qquad (4)$$

and the within-class scatter matrix S_w is defined as

$$S_w = \frac{1}{N} \sum_{i=1}^{c} \sum_{x_k \in C_i} (x_k - \mu_i)(x_k - \mu_i)^T = \frac{1}{N} \Phi_w \Phi_w^T \qquad (5)$$

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.,

$$W_{opt} = \arg\max_W \frac{|W^T S_b W|}{|W^T S_w W|} = [w_1 w_2 ... w_m] \qquad (6)$$

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.,

$$S_b w_i = \lambda_i S_w w_i \quad i = 1, 2, ..., m \qquad (7)$$

3. Two-dimensional PCA, Two-dimensional LDA and our proposed approaches

In 2D approach, the image matrix does not need to be previously transformed into a vector, so a set of *N* sample images is represented as $\{X_1, X_2, ..., X_N\}$ with $X_i \in \mathbb{R}^{kxs}$. The total scatter matrix is defined as

$$G_T = \sum_{i=1}^{N} (X_i - \mu_X)(X_i - \mu_X)^T \quad \textbf{(8)}$$

with $\mu_X = \frac{1}{N}\sum_{i=1}^{N} X_i \in \mathbb{R}^{kxs}$ is the mean image of all samples. $G_T \in \mathbb{R}^{kxk}$ is also called image covariance (scatter) matrix.

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 :

$$Y_i = W^T (X_i - \mu_X) \in \mathbb{R}^{mxs} \quad \textbf{(9)}$$

where $i = 1, 2, ..., N$ and $W \in \mathbb{R}^{kxm}$ is a matrix with orthonormal columns. In 2DPCA, the projection W_{opt} is chosen to maximize $tr(W^T G_T W)$. The optimal projection $W_{opt} = [w_1 w_2 ... w_m]$ with $\{w_i \,|\, i = 1, 2, ..., m\}$ is the set of *n*-dimensional eigenvectors of G_T corresponding to the *m* largest eigenvalues.

In 2DLDA, the between-class scatter matrix S_b is re-defined as

$$S_b = \frac{1}{N}\sum_{i=1}^{c} N_i (\mu_{C_i} - \mu_X)(\mu_{C_i} - \mu_X)^T \quad \textbf{(10)}$$

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

$$S_w = \frac{1}{N}\sum_{i=1}^{c} \sum_{X_k \in C_i} (X_k - \mu_{C_i})(X_k - \mu_{C_i})^T \quad \textbf{(11)}$$

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 :

$$Y_i = W^T (X_i - \mu_X) \in \mathbb{R}^{mxs} \quad \textbf{(12)}$$

where $i = 1, 2, ..., N$ and $W \in \mathbb{R}^{kxm}$ is a matrix with orthonormal columns. And the projection W_{opt} is chosen with the criterion same as that in (6).

After a transformation by 2DPCA or 2DLDA, 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 :

$$d(Y_i, Y_j) = \sqrt{\sum_{u=1}^{k} \sum_{v=1}^{s} (Y_i(u,v) - Y_j(u,v))^2} \quad \textbf{(13)}$$

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 .

The 2D approach can eliminate the correlations between image columns and compress the discriminant information optimally into a few of columns in horizontal direction. However, it disregards the correlations between image rows and the data compression in vertical direction. So, its compression rate is far lower than 1D approach and more coefficients are needed for the representation of images. This must lead to a slow classification speed and large storage requirements for large-scaled databases. In this section, we will suggest a way to overcome the weakness of 2DPCA and 2DLDA. Our idea is simple, just to perform 2DPCA or 2DLDA twice: the first one is in horizontal direction and the second is in vertical direction (note that any operation in vertical direction can be equivalently implemented by an operation in horizontal direction by virtue of the transpose operation of matrix). Specifically, given image X_i , we obtain its feature matrix Y_i after the first 2DPCA or 2DLDA transform. Then, we transpose Y_i and input Y_i^T into 2DPCA or 2DLDA, and the resulting feature matrix Z_i could be obtained. This process is illustrated in Fig. 1.

Detailed implementation of *two-stage* 2DPCA or 2DLDA can be summarized as follow

Step 1 : Horizontal 2DPCA or 2DLDA

- Given training images X_i with $i = 1..N$, calculate the total scatter matrix, or between-class scatter matrix and within-class scatter matrix by (8), (10), (11).

- After first transformation by 2DPCA or 2DLDA, a feature matrix is obtained for each image $Y_i = W_{1opt}^T X_i$ with $W_{1opt} = [w_1 w_2 ... w_m]$ is the optimal projection.

Step 2 : Vertical 2DPCA or 2DLDA

- Similarly, input Y_i^T into 2DPCA or 2DLDA, and the resulting feature matrix $Z_i = W_{2opt}^T Y_i^T$ could be obtained.

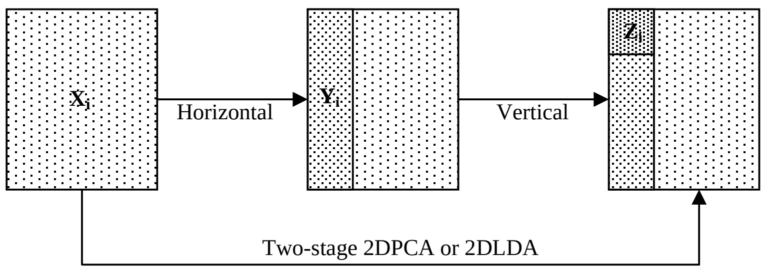

Fig. 1 : The illustration of *two-stage* 2DPCA or 2DLDA

4. Experimental results

This section evaluates the performance of our propoped algorithms *two-stage* 2DPCA and *two-stage* 2DPCA with the following algorithms : Eigenfaces (PCA) , Fisherfaces (LDA), Direct LDA [5], 2DPCA and 2DLDA based on using ORL face database. 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, 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. Since the number of projection vectors, has a considerable impact on the results of the different algorithms, we choose the value that corresponds to the best classification result on the image set consisting of the first k images of each subject as its optimal value. In all of the experiments, the nearest neighbor algorithm under the Euclidean distance is employed to classify the test images. For each value of k , 30 runs are performed with different random partition between training set and testing set. *Table 1* shows the average recognition rates (%) of different approaches with ORL database, while *Table 2* show us the performances of those approaches in terms of computation time.

Table 1. Comparison of the average error rates (%) of different approaches on the ORL database.

k	2	3	4	5	6
Eigenfaces - PCA	82.46	89.27	92.73	95.24	96.11
Fisherfaces - LDA	79.47	86.84	90.37	91.89	93.93
Direct LDA	80.94	89.19	92.71	95.46	96.97
2DPCA	85.05	90.18	93.89	96.18	97.11
***Two-stage* 2DPCA**	**86.02**	**90.89**	**94.36**	**97.80**	**98.20**
2DLDA	87.41	92.34	95.54	96.12	96.70
***Two-stage* 2DLDA**	**89.14**	**93.56**	**96.87**	**97.36**	**98.15**

Table 2. The average CPU time (s) consumed for training and testing, the top recognition rates (%) and the corresponding number of samples of seven methods. (CPU : PIV 2.4 GHz, RAM 512M)

k	2	3	4	5	6
Eigenfaces - PCA	27.59	37.70	52.34	36.08	32.15
Fisherfaces - LDA	16.18	32.71	44.08	68.96	81.09
Direct LDA	15.59	15.61	15.31	17.61	17.31
2DPCA	0.50	1.10	1.25	1.39	1.52
***Two-stage* 2DPCA**	**0.35**	**0.90**	**1.16**	**1.19**	**1.31**
2DLDA	0.58	1.21	1.34	1.45	1.63
***Two-stage* 2DLDA**	**0.41**	**0.98**	**1.22**	**1.34**	**1.49**

We can see that our methods achieve the better recognition rate compared to the other approaches with a lower computation cost.

5. Conclusions

New 2DPCA-based and 2DLDA-based methods for face recognition have been proposed in this paper. The proposed methods can outperform the other methods in terms of both recognition rate and computation cost. The initial idea of *two-stage* 2DPCA or *two-stage* 2DPCA is to perform 2DPCA or 2DLDA twice: the first one is in horizontal direction and the second is in vertical direction. After the two sequential transforms, the discriminant information is compacted into the up-left corner of the image. The effectiveness of the proposed approaches can be seen through our experiments based on ORL face databases.

Reference

1. M. Turk, A. Pentland : Eigenfaces for face recognition. Int. J. Cog. Neurosci., vol. 3, (1991) 71–86.

2. W. Zhao, R. Chellappa, P.J. Phillips: Subspace Linear Discriminant Analysis for Face Recognition. Technical Report CAR-TR-914, Center for Automation Research, University of Maryland, 1999.
3. D. L. Swets, J. J. Weng: Using discrimination eigenfeatures for image retrieval. IEEE Trans. Pattern Anal. Machine Intell., vol. 18 (1996) 831–836.
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. [9] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” Ann. Eugenics, vol. 7, pp. 178–188, 1936.
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) 29 – 32.
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.
14. Cevikalp, H., Neamtu, M., Wilkes. M., Barkana, A.: Discriminative common vectors for face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on , Vol. 27 (2005) 4 – 13.
15. Jian Yang, Zhang, D., Frangi, A.F., Jing-yu Yang: Two-dimensional PCA: a new approach to appearance-based face representation and recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on , Vol. 26 (2004) 131 – 137
16. Ming Li, Baozong Yuan : 2D-LDA: A statistical linear discriminant analysis for image matrix. Pattern Recognition Letters, Vol. 26, Issue 5 (2005) 527-532.
17. “The ORL face database” http://www.uk.research.att.com/facedatabase.html.