Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition 39 (2006) 1396 – 1400
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

$(2D)^2$LDA: An efficient approach for face recognition

S. Noushath[a,*], G. Hemantha Kumar[a], P. Shivakumara[b]

[a]Department of Studies in Computer Science, University of Mysore, Manasagangothri, Mysore 570006, Karnataka, India
[b]Department of Computer Science, School of Computing, National University of Singapore (NUS), 3 Science, Drive 2, Singapore 117543

Received 31 August 2005; accepted 24 January 2006

Abstract

Although 2DLDA algorithm obtains higher recognition accuracy, a vital unresolved problem of 2DLDA is that it needs huge feature matrix for the task of face recognition. To overcome this problem, this paper presents an efficient approach for face image feature extraction, namely, $(2D)^2$LDA method. Experimental results on ORL and Yale database show that the proposed method obtains good recognition accuracy despite having less number of coefficients.

Keywords: Principal component analysis; Linear discriminant analysis; Face recognition

1. Introduction

Linear discriminant analysis (LDA) is a well-known feature extraction and data representation technique widely used in the areas of pattern recognition for feature extraction and dimension reduction. The objective of LDA is to find the optimal projection so that the ratio of the determinants of the between-class and the within-class scatter matrices of the projected samples reaches its maximum. However, concatenating 2D matrices into 1D vectors leads to very high dimensional nature of image vector, where it is difficult to evaluate the scatter matrices accurately due to its large size and the relatively small number of training samples. Furthermore, the within-class scatter matrix is always singular, making the direct implementation of LDA algorithm an intractable task.

To overcome these problems, a new technique called 2DLDA [1] was recently proposed, which directly computes eigenvectors of the so called scatter matrices without matrix-to-vector conversion. Because the size of the scatter matrices is equal to the width of the images, which is quite small compared to the size of the scatter matrices in LDA, 2DLDA evaluates the scatter matrices more accurately and computes the corresponding eigen vectors more efficiently. It was reported in Ref. [1] that the recognition accuracy on several databases was higher using 2DLDA than other PCA and LDA-based algorithms.

However, the main drawback of 2DLDA is that it needs more coefficients for image representation than conventional PCA- and LDA-based schemes. For an image size of 112×92, the commonly used image size in face recognition, the number of coefficients used by 2DLDA for classification is $112 \times d$, where d is set to no less than 5 for satisfactory accuracy.

In this paper, we first indicate that 2DLDA is essentially working in the row-direction of images, and then propose an alternative 2DLDA which works in the column direction of images. By simultaneously combining row and column directions, we develop two-directional 2DLDA, i.e. $(2D)^2$LDA, for efficient representation and recognition. Experimental results on ORL and Yale database shows that the proposed method obtains same or even better recognition accuracy than 2DLDA, while the number of coefficients needed by the former for image representation is much smaller than that of the latter.

* Corresponding author. Tel.: +91 821 2510789; fax: +91 821 2510789.
E-mail address: nawali_naushad@yahoo.co.in (S. Noushath).

0031-3203/$30.00
doi:10.1016/j.patcog.2006.01.018

2. Overview of 2DLDA approach

2DLDA is an effective feature extraction and discrimination approach [1] in face recognition. Formally, it can briefly be formulated as follows: Suppose $\{A_k\}_{k=1}^{N}$ are the training images, which contain C classes, and the ith class C_i has n_i samples ($\sum_{i=1}^{C} n_i = N$). 2DLDA attempts to seek a set of optimal discriminating vectors to form a transform $X_d = \{x_1, x_2, \ldots, x_d\}$ by maximizing the 2D Fisher criterion denoted as

$$J(X) = \frac{X^{\mathrm{T}} G_b X}{X^{\mathrm{T}} G_w X}. \tag{1}$$

In Eq. (1), T denotes matrix transpose, G_b and G_w, respectively, are between-class and within-class scatter matrices:

$$G_b = \frac{1}{N} \sum_{i=1}^{C} n_i (\overline{A}_i - \overline{A})^{\mathrm{T}} (\overline{A}_i - \overline{A}), \tag{2}$$

$$G_w = \frac{1}{N} \sum_{i=1}^{C} \sum_{j \in C_i} (A_j - \overline{A_i})^{\mathrm{T}} (A_j - \overline{A_i}), \tag{3}$$

$\overline{A}_i, \overline{A}$ denote the means of ith class and the whole training set, respectively. A_j is the jth image in the class C_i. The goal of 2DLDA scheme is to find the optimal discriminating vectors X_{opt} in order to maximize $J(X)$. Obviously, the optimal discrimination vectors X_{opt} are the eigenvector corresponding to the dominant eigenvalues of eigenstructure $G_w^{-1} G_b$. It has been proved that the optimal value for the discriminating vectors X_{opt} is composed of the orthonormal eigenvectors $x_1, x_2, \ldots, x_d$ of $G_w^{-1} G_b$ corresponding to the d largest eigenvalues. Now, given an image $A_{m \times n}$, all the projections of the image matrix in the d-directions make up md-dimensional vector, which is the 2DLDA feature vector.

2.1. Proposed alternative-2DLDA

Let $A_k = [(A_k^{(1)})^{\mathrm{T}}, (A_k^{(2)})^{\mathrm{T}}, \ldots, (A_k^{(m)})^{\mathrm{T}}]^{\mathrm{T}}$, $\overline{A}_i = [(\overline{A}_i^{(1)})^{\mathrm{T}}, (\overline{A}_i^{(2)})^{\mathrm{T}}, \ldots, (\overline{A}_i^{(m)})^{\mathrm{T}}]^{\mathrm{T}}$, $\overline{A} = [(\overline{A}^{(1)})^{\mathrm{T}}, (\overline{A}^{(2)})^{\mathrm{T}}, \ldots, (\overline{A}^{(m)})^{\mathrm{T}}]^{\mathrm{T}}$, where $A_k^{(j)}, \overline{A}_i^{(j)}, \overline{A}^{(j)}$ denote the jth row vectors of A_k, $\overline{A}_i$ and $\overline{A}$, respectively. Then Eqs. (2) and (3) can be written as:

$$G_b = \frac{1}{N} \sum_{i=1}^{C} n_i \sum_{j=1}^{m} (\overline{A}_i^{(j)} - \overline{A}^{(j)})^{\mathrm{T}} (\overline{A}_i^{(j)} - \overline{A}^{(j)}), \tag{4}$$

$$G_w = \frac{1}{N} \sum_{i=1}^{C} \sum_{k \in C_i} \sum_{j=1}^{m} (A_k^{(j)} - \overline{A}_k^{(j)})^{\mathrm{T}} (A_k^{(j)} - \overline{A}_k^{(j)}). \tag{5}$$

Eq. (5) reveals that the scatter matrix G_w can be obtained from the outer product of row vectors of images, assuming the training images have zero mean [2]. For this reason, we claim that original 2DLDA is working in the row direction of images. Apparently, a natural extension is to use the outer product between column vectors of images to construct G_b and G_w.

Let

$$A_k = [(A_k^{(1)}), (A_k^{(2)}), \ldots, (A_k^{(n)})],$$
$$\overline{A}_i = [(\overline{A}_i^{(1)}), (\overline{A}_i^{(2)}), \ldots, (\overline{A}_i^{(n)})],$$
$$\overline{A} = [(\overline{A}^{(1)}), (\overline{A}^{(2)}), \ldots, (\overline{A}^{(n)})],$$

where $A_k^{(j)}, \overline{A}_i^{(j)}, \overline{A}^{(j)}$, respectively denote the jth column vectors of A_k, $\overline{A}_i$ and$\overline{A}$.

Let Z denotes an m-dimensional unitary column vector. Projecting the image matrix $A_{m \times n}$ onto Z yields a $q \times n$ feature matrix, i.e, $B = Z^{\mathrm{T}} A$. Similar to Eq. (1), the following criterion is adopted to find the optimal projection vector Z and is given by $J(Z) = trace(S_b^z)/trace(S_w^z)$, where S_b^z and S_w^z are, respectively, given by $1/N \sum_{i=1}^{C} n_i (\overline{y}^i - \overline{y})(\overline{y}^i - \overline{y})^{\mathrm{T}}$ and $1/N \sum_{i=1}^{C} \sum_{j \in C_i} (y_j - \overline{y}^i)(y_j - \overline{y}^i)^{\mathrm{T}}$. Here $\overline{y}$ and $\overline{y}^i$, respectively, denote the global and the mean vector of ith class in the projection space.

It is easy to verify that $trace(S_b^z) = Z \cdot G \cdot Z^{\mathrm{T}}$ and $trace(S_w^z) = Z \cdot G_w \cdot Z^{\mathrm{T}}$ where G_b and G_w are now given as

$$G_b = \frac{1}{N} \sum_{i=1}^{C} n_i \sum_{j=1}^{m} (\overline{A}_i^{(j)} - \overline{A}^{(j)})(\overline{A}_i^{(j)} - \overline{A}^{(j)})^{\mathrm{T}}, \tag{6}$$

$$G_w = \frac{1}{N} \sum_{i=1}^{C} \sum_{k \in C_i} \sum_{j=1}^{m} (A_k^{(j)} - \overline{A}_k^{(j)})(A_k^{(j)} - \overline{A}_k^{(j)})^{\mathrm{T}}. \tag{7}$$

Similarly, the optimal projection matrix $Z_{\mathrm{opt}} = [z_1, z_2, \ldots, z_q]$ can be obtained by computing the orthonormal eigenvectors of $G_w^{-1} G_b$ corresponding to the q largest eigenvalues thereby maximizing $J(Z)$.

2.2. Proposed (2D)^2 LDA method: 2-directional 2-dimensional LDA

We reasoned in Section 2.1 that 2DLDA works in the rowwise direction reflecting the information between row of images to learn an optimal matrix X from a set of training images, and then project an $m \times n$ image A onto X, yielding m by d matrix, i.e. $Y_{m \times d} = A_{m \times n} \cdot X_{n \times d}$. Similarly, the alternative 2DLDA learns optimal projection matrix Z reflecting information between columns of images and then projects A onto Z, yielding a q by n matrix, i.e. $B_{q \times n} = Z_{m \times q}^{\mathrm{T}} \cdot A_{m \times n}$.

Suppose we have obtained the projection matrices X (as in Section 2) and Z (as in Section 2.1), projecting the m by n image A onto X and Z simultaneously, yielding a q by d matrix C,

$$C = Z^{\mathrm{T}} \cdot A \cdot X. \tag{8}$$

The matrix C is also called the coefficient matrix in image representation. When used for face recognition, the matrix C is also called the feature matrix. After projecting each training image $A_k = (k = 1, 2, 3, \ldots, N)$ onto X and Z, we obtain the feature matrices C_K $(k = 1, 2, 3, \ldots, N)$. Given a test face image A, first use Eq. (8) to get the feature matrix C, then a nearest neighbor classifier is used for classification. Here, the distance between C and C_k is defined by

$$d(C, C_k) = \|C - C_k\| = \sqrt{\sum_{i=1}^{q}\sum_{j=1}^{d}(C^{(i,j)} - C_k^{(i,j)})^2}.$$

3. Experimental results

In this section, we experimentally evaluate our proposed alternative-2DLDA and $(2D)^2$LDA methods with PCA [3], 2DPCA [4], alternative-2DPCA [2], $(2D)^2$PCA [2] and 2DLDA [1] methods, on two well-known face databases: ORL and Yale. While the ORL database is used to test the performance of the face recognition algorithms under the condition of minor variation of scaling and rotation, the Yale database is used to examine the performance of the algorithms under the condition of varied facial expression and lighting configuration. All of our experiments are carried out on a PC machine with P4 3 GHz CPU and 512 MB RAM memory under Matlab 7 platform.

3.1. Results on ORL database

The ORL database (http://www.uk.research.att.com/facedatabase.html) contains 112×92 sized 400 frontal faces: 10 tightly, cropped images of 40 individuals with variation in pose, illumination, facial expression (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). In our experiments, we split the whole database into two parts evenly. First five images of each class is used for training and the rest of the five images are used for testing. This experiment is repeated 25 times by varying projection vectors d (where $d = 1, 2, 3, \ldots, 20, 25, 30, 35, 40, 45$). Since d, the number of projection vectors, has a considerable impact on different algorithms, we chose the value that corresponds to the best classification result on the image set. Table 1 gives the comparisons of seven methods on top recognition accuracy, corresponding dimension of feature vector (for PCA) or feature matrices (for the other six methods) and running times. It can be found from Table 1 that the top recognition accuracy of proposed alternative 2DLDA method is comparable with other methods. Table 1 also reveals that top recognition accuracy of $(2D)^2$LDA method is significantly higher than other methods despite having reduced feature matrix. Finally, Table 1 shows that $(2D)^2$LDA and $(2D)^2$PCA methods consume least running time among the other methods.

Table 1
Comparison of seven methods on ORL database

Methods	Top recognition rate	Dimension	Running time
PCA	95.50	40	21.78
2DPCA	97.00	112×11	7.89
Alternative 2DPCA	97.50	13×92	6.73
$2D^2$ PCA	97.75	7×7	3.89
2DLDA	98.00	112×7	6.84
Alternative 2DLDA	98.00	9×92	7.05
$2D^2$LDA	98.50	8×8	4.14

We also tested the performance of proposed method under noise conditions. For this, we randomly select one image from each class and generate 10 noisy images (with salt and pepper noise) for that class by varying noise density from 0.1 to 1.0. So effectively, we created 400 noisy images corresponding to 40 different classes. We used 400 original images of ORL for training, and during testing we used the noisy images thus created. So size of both testing and training database is 400. Again, we repeat this experiment 25 times for each noise density by varying the principal components. Fig. 1 is the graphical plot of top recognition rate against noise density. From Fig. 1 it is clear that proposed $(2D)^2$LDA method outperforms other methods in terms of recognition rate despite having reduced feature matrix. Moreover, it can be found from Fig. 1 that top recognition rate achieved by alternative-2DLDA is comparable with other methods.

In order to make full use of the available data, we randomly select p images from each subject to construct the training data set, the remaining images being used as the test images. To ensure sufficient training, a value of at least 2 is used for p. As mentioned earlier, for each p, all the algorithms are repeated 25 times for varying number of projection vectors. Table 2 shows the top recognition accuracy

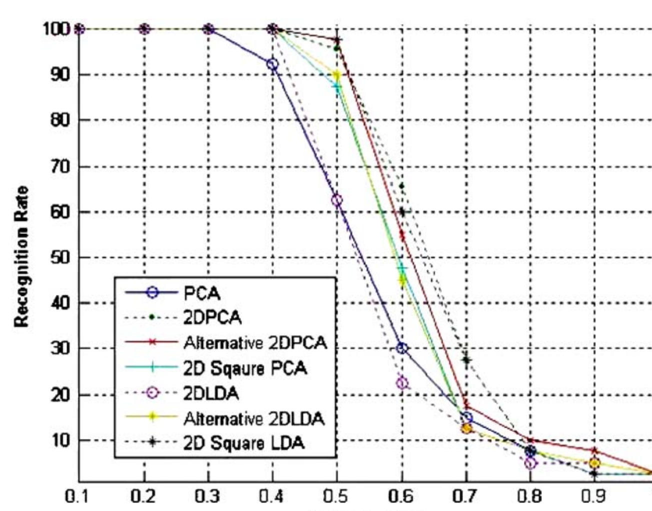

Fig. 1. Performance of different methods under noise conditions.

Table 2
Comparison of different approaches in terms of top recognition accuracy on ORL database

Method	No. of training samples per class			
	2	4	6	8
PCA	85.50 (40)	93.25 (40)	98.00 (13)	99.00 (19)
2DPCA	89.75 (112 × 3)	95.75 (112 × 3)	98.50 (112 × 6)	99.25 (112 × 5)
Alternative 2DPCA	89.50 (16 × 92)	95.25 (13 × 92)	98.50 (8 × 92)	99.50 (10 × 92)
$2D^2$ PCA	88.75 (9 × 9)	95.00 (11 × 11)	98.50 (7 × 7)	99.50 (5 × 5)
2D LDA	90.25 (112 × 4)	95.25 (112 × 3)	99.00 (112 × 5)	99.50 (112 × 7)
Alternative 2DLDA	88.75 (8 × 92)	95.25 (8 × 92)	98.25 (9 × 92)	99.75 (3 × 92)
$2D^2$ LDA	89.25 (7 × 7)	96.00 (11 × 11)	98.75 (8 × 8)	99.75 (5 × 5)

Table 3
Comparison of different approaches in terms of top recognition rate on Yale database

Methods	Number of training samples per class			
	2	4	6	8
PCA	74.67	88.67	94.00	99.33
2DPCA	85.15	93.33	94.54	99.39
Alternative 2DPCA	84.67	93.33	96.67	99.33
$(2D)^2$PCA	84.67	92.67	96.67	99.33
2DLDA	84.00	94.54	95.75	99.39
Alternative 2DLDA	85.33	92.67	96.00	99.39
$(2D)^2$LDA	86.67	93.33	94.54	99.33

achieved by all the methods for varying number of training samples. The values in the parenthesis denote the dimension of feature vector (for PCA) or feature matrices (for other six methods) to attain top recognition accuracy.

Although PCA method uses less number of coefficients for recognition purpose, one of the inextricable problem of this PCA method is that its computational time is considerably high. This is not difficult to understand because the size of the covariance matrix used in this method is very huge compared to other methods.

3.2. Results on Yale database

Yale database (http://cvc.yale.edu/projects/yalefaces/yalefaces.html) contains 165 grayscale images of 15 individuals, each of which is cropped with the size of 225 × 195. There are 11 images per subject, one per different facial expression or lighting configurations. We considered this database in order to evaluate the performance of methods under the condition when facial expression and lighting conditions are changed.

Table 3 shows the top recognition accuracy obtained by different methods for varying number of training samples. As previously mentioned, we repeated each experiment 25 times by varying the number of projection vectors. Table 3 reveals that proposed alternative 2DLDA and $(2D)^2$LDA methods are comparable to 2DPCA, alternative 2DPCA, $(2D)^2$PCA and 2DLDA methods in terms of recognition

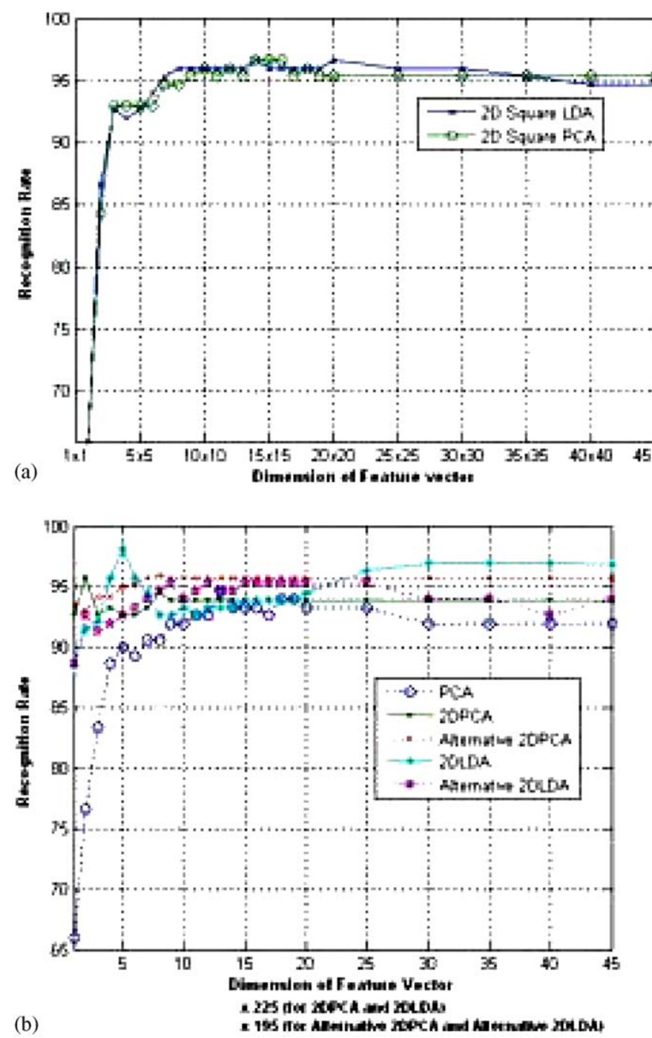

Fig. 2. Recognition performance of different approaches with varying dimension of feature vectors.

accuracy. The recognition performance of $(2D)^2$LDA and $(2D)^2$PCA, and PCA, 2DLDA, alternative 2DLDA, 2DPCA, and alternative 2DPCA with varying dimension of feature vectors for five training samples is given in Fig. 2(a) and (b), respectively. It can be easily ascertained from Fig. 2 that the $(2D)^2$LDA method, with reduced feature vector, obtains same or even good recognition accuracy when compared with other methods.

4. Conclusion

In this paper, an efficient face representation and recognition method called $(2D)^2$LDA is proposed. The main difference between $(2D)^2$LDA and existing 2DLDA is that the latter only works in the row direction of face images, while the former works simultaneously in the row and the column directions of face images. The major advantage of the proposed method is that it requires fewer number of coefficients and least computing time for face image representation and recognition unlike standard PCA, 2DPCA, and 2DLDA methods. Experimental results show the effects of the proposed method.

References

[1] L. Ming, B. Yuan, 2D-LDA: a statistical linear discriminant analysis for image matrix, Pattern Recognition Lett. 26 (5) (2005) 527–532.
[2] D. Zhang, Z.-H. Zhou, $(2D)^2$PCA: 2-directional 2-dimensional PCA for efficient face representation and recognition, J. Neurocomput. 69 (1–3) (2005) 224–231.
[3] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cognitive Neurosci. 3 (1) (1991) 71–86.
[4] J. Yang, D. Zhang, A.F. Frangi, J.Y. Yang, Two-dimensional PCA: a new approach to appearance based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (1) (2004) 131–137.