Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Pattern Recognition 38 (2005) 1121 – 1124

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Rapid and brief communication

Two-dimensional FLD for face recognition☆

Huilin Xiong, M.N.S. Swamy*, M.O. Ahmad

Department of Electrical and Computer Engineering, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8

Received 24 November 2004; accepted 8 December 2004

Abstract

This paper presents a new scheme of face image feature extraction, namely, the two-dimensional Fisher linear discriminant. Experiments on the ORL and the UMIST face databases show that the new scheme outperforms the PCA and the conventional PCA + FLD schemes, not only in its computational efficiency, but also in its performance for the task of face recognition.

Keywords: Fisher criterion; Principal component analysis (PCA); Linear discriminant analysis (LDA)

1. Introduction

Fisher linear discriminant (FLD), sometimes known as linear discriminant analysis (LDA), has been widely used in pattern recognition for feature extraction and dimension reduction. The objective of the FLD 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. A difficulty in using the FLD method for face recognition is the very high-dimensional nature of the image vector. For an image of size 112×92, the commonly used image size in face recognition, the dimension of the vector space is 10304, and the size of the scatter matrices 10304×10304. Obviously, it is quite difficult to handle matrices of such large size. Moreover, the within-class scatter matrix is always singular, making the direct implementation of the classical FLD algorithm impossible.

The traditional solution to this problem is to utilize the principal component analysis (PCA) as a pre-processing step aiming to reduce the dimensionality of the vector space. After all the image vectors are projected into the subspace consisting of the "principal components", the FLD algorithm can perform well in the subspace. However, since the projection criterion of the PCA and that of the FLD are essentially different, the pre-processing procedure to reduce the dimensionality using the PCA could result in the loss of some important discriminatory information for the FLD algorithm that follows the PCA. Actually, Chen et al. [1] have shown that the null space of the within-class scatter matrix contains valuable discriminatory information. In view of this, the so-called direct LDA (DLDA) algorithms [1,2] have been proposed to avoid the possible loss of useful information.

An alternative way to handle the above problem is to directly project the image matrix under a specific projection criterion, rather than using the stretched image vector. Yang et al. [3] have shown that a "two-dimensional" PCA (2DPCA) can be constructed in a straightforward manner based on the image matrix projection. The size of the scatter matrices for the 2DPCA scheme is either only $m \times m$ or $n \times n$ for an image of size $m \times n$, instead of the size $mn \times mn$ in the classic PCA scheme. Therefore, the 2DPCA scheme

☆ This work was partially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

* Corresponding author. Tel.: +1 514 848 2424x3091; fax: +1 514 848 2802.

E-mail addresses: hlxiong@ece.concordia.ca (H. Xiong), swamy@ece.concordia.ca (M.N.S. Swamy), omair@ece.concordia.ca (M.O. Ahmad).

0031-3203/$30.00
doi:10.1016/j.patcog.2004.12.003

is much faster in computation than the conventional PCA algorithm. The most interesting fact shown in Ref. [3] is that the features extracted by the 2DPCA scheme are more efficient than those extracted by the conventional PCA scheme for the task of face recognition and image reconstruction.

In this paper, we present a two-dimensional FLD scheme, called 2DFLD, which employs an alternate to the commonly used measure of the Fisher discriminant scalar, and derive a matrix-projection-based FLD algorithm. Experimental results show that the 2DFLD feature extraction scheme outperforms the conventional PCA and PCA + FLD scheme not only in terms of the computational efficiency, but also in terms of the performance for face recognition.

2. Two-dimensional FLD for image feature extraction

We project an $m \times n$ image matrix X onto an m-dimensional vector space through the transformation $y = X\alpha$, where α is an n-dimensional vector, and y the m-dimensional projected vector. Our goal is to find the optimal projection direction α so that the projected vectors in the m-dimensional space reaches its maximum class separability.

2.1. An alternative Fisher criterion

The conventional Fisher criterion is not convenient for the theoretical analysis in the case of our matrix-based projection. Instead, in this paper, we adopt the alternative measure of class separability given by

$$J = \frac{\mathrm{tr}(S_b)}{\mathrm{tr}(S_w)}, \tag{1}$$

where "tr" denotes the trace of a matrix, and S_b and S_w, respectively, denote the between-class and within-class matrices. The measure J in Eq. (1) is also a well-known Fisher scalar for measuring class separability, actually, it is called "criteria J_4" in Ref. [4].

Suppose $\{X_i\}_{i=1}^{N}$ are the training images, which contain C classes (subjects), and the ith class C_i has n_i samples ($\sum_{i=1}^{C} n_i = N$). The images, all $m \times n$ matrices, are projected into a m-dimensional vector space $y_i = X_i\alpha$. In the projection space, the measure of the class separability of the projected images is calculated by

$$J(\alpha) = \frac{\mathrm{tr}(S_b^{\alpha})}{\mathrm{tr}(S_w^{\alpha})}, \tag{2}$$

where

$$S_b^{\alpha} = \frac{1}{N} \sum_{i=1}^{C} n_i (\overline{y}^i - \overline{y})(\overline{y}^i - \overline{y})^{\mathrm{T}},$$

$$S_w^{\alpha} = \frac{1}{N} \sum_{i=1}^{C} \sum_{j \in C_i} (y_j - \overline{y}^i)(y_j - \overline{y}^i)^{\mathrm{T}}$$

in which $\overline{y}$ and $\overline{y}^i$, respectively, denote the global mean vector and the mean vector of the ith class in the projection space.

It is easy to verify that $\mathrm{tr}(S_b^{\alpha}) = \alpha^{\mathrm{T}} G_b \alpha$ and $\mathrm{tr}(S_w^{\alpha}) = \alpha^{\mathrm{T}} G_w \alpha$, where

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

$$G_w = \frac{1}{N} \sum_{i=1}^{C} \sum_{j \in C_i} (X_j - \overline{X}^i)^{\mathrm{T}} (X_j - \overline{X}^i)$$

in which $\overline{X}$ and $\overline{X}^i$, respectively, represent the global and the ith class mean images.

We call matrices G_b and G_w, image between-class scatter matrix and image within-class scatter matrix, respectively. Note that the size of the image scatter matrices is only $m \times m$, much smaller than that of the scatter matrices whose sizes are $mn \times mn$ in the conventional FLD algorithm. Using the image scatter matrices, the two-dimensional Fisher criterion given by Eq. (1) can be expressed as

$$J(\alpha) = \frac{\alpha^{\mathrm{T}} G_b \alpha}{\alpha^{\mathrm{T}} G_w \alpha}. \tag{3}$$

2.2. Two-dimensional FLD feature extraction

The goal of our 2DFLD scheme is to find the optimal projection direction α in order to maximize (3). Obviously, the optimal projection direction α_{opt} is the eigenvector corresponding to the maximum eigenvalue of the eigenstructure:

$$G_b \alpha = \lambda G_w \alpha. \tag{4}$$

It is not difficult to handle the above eigenproblem directly, since the size of the matrix G_b or G_w is only $m \times m$. In practice, one optimal projective direction is not enough to extract sufficient discriminatory features. We usually need to project the image data onto a set of orthogonal directions, namely, $\alpha_1, \alpha_2, \ldots, \alpha_k$, which maximize the criterion (3). These projection directions can be selected as the k eigenvectors corresponding to the first k largest eigenvalues of the eigenstructure (4).

Suppose $\{\alpha_i\}_{i=1}^{k}$ are the optimal projective directions. Given an image X, all the projections of the image matrix in the k directions make up a mk-dimensional vector y, which is our 2DFLD feature vector

$$y^{\mathrm{T}} = (y_1^{\mathrm{T}}, y_2^{\mathrm{T}}, \ldots, y_k^{\mathrm{T}}) = (\alpha_1^{\mathrm{T}}, \alpha_2^{\mathrm{T}}, \ldots, \alpha_k^{\mathrm{T}}) X^{\mathrm{T}}.$$

3. Experimental results

Two face image databases, namely, the ORL database and UMIST database, are used to compare the proposed 2DFLD approach with the following algorithms: the PCA,

Fig. 1. Ten sample images of a subject in the UMIST database.

the PCA + FLD [5], the direct LDA (DLDA) [2], and the 2DPCA [3] schemes. 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 UMIST database is used to examine the performance of the algorithms when the angle of rotation of the facial images is quite large.

The ORL database (http://www.cam-orl.co.uk) contains 40 persons, each having 10 different images. The UMIST face image database, available at http://images.ee.umist.ac.uk/danny/database.html, contains 20 subjects, each covering a wide range of multiviewed face images, from profile to frontal views. We select 25 pictures from each subject to construct our UMIST input images. Note that there are six subjects in which the image number is less than 25, and a few "mirror" images generated by exchanging the right and left pixels of some existing frontal pictures are added to the subject in order to compliment the image number in the subject up to 25. Therefore, the total image number of the UMIST data set is 500. Fig. 1 shows ten sample images of one of the subject.

In the experiments, 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. Since k, 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 p 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.

Each experiment is repeated 20 times. The average recognition error rates of the different algorithms on the test sets are, respectively, summarized in Table 1 for the ORL database, and Table 2 for the UMIST database. Table 3 shows the average CPU time (CPU: PIII 800, RAM: 128MB) taken for feature extraction on the ORL database under the Matlab 6.1 platform for the various algorithms. We see from these tables that the proposed 2DFLD approach is more efficient than the PCA and the PCA + FLD schemes, not only in terms of computation times, but also in terms of the error rate for face recognition. The proposed 2DFLD approach also performs better than the DLDA and 2DPCA schemes in terms of the error rate, although the 2DFLD scheme is comparable to the 2DPCA scheme in terms of the computational efficiency for the feature extraction.

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

p	2	3	4	5	6
PCA	18.42	11.57	8.13	5.68	4.78
PCA + FLD	21.41	14.14	10.58	9.02	7.03
DLDA	19.98	11.80	8.23	5.45	4.03
2DPCA	15.81	10.68	7.10	4.73	3.78
2DFLD	13.41	8.34	6.54	4.60	3.70

Table 2
Comparison of different approaches in terms of error rates (%) on the UMIST database

p	2	3	4	5	6	7	8
PCA	32.08	24.70	16.94	14.16	13.25	9.51	6.40
PCA + FLD	29.90	18.30	14.24	10.98	9.33	8.26	5.43
DLDA	28.28	18.34	11.62	9.74	8.99	6.04	3.79
2DPCA	22.76	16.42	10.39	7.98	6.86	4.47	3.07
2DFLD	19.70	12.75	7.11	5.91	5.46	3.08	1.76

Table 3
Comparison of the average CPU time (s) for feature extraction on the ORL database

p	2	3	4	5	6
PCA	57.49	77.42	106.72	74.34	66.37
PCA + FLD	34.43	67.68	90.44	140.31	164.43
DLDA	33.44	33.29	32.63	37.36	36.98
2DPCA	3.17	4.30	4.10	5.11	5.28
2DFLD	3.91	4.40	4.78	5.87	6.26

4. Conclusion

In this paper, we have presented a new scheme for image feature extraction, namely, the two dimensional FLD (2DFLD), which is based on a straightforward projection of the image matrix along with an alternative Fisher criterion. Since the size of the scatter matrices in the 2DFLD scheme is much smaller than those in the conventional FLD and PCA schemes, the computation of the feature extraction using the proposed 2DFLD scheme is much more efficient than that of the PCA and PCA + FLD schemes. More importantly, experimental results on the ORL and UMIST databases have shown that the 2DFLD features outperform the features extracted by the PCA, 2DPCA, PCA+FLD, and DLDA schemes in the performance of face recognition.

References

[1] L. Chen, H. Liao, M. Ko, J. Lin, G. Yu, A new LDA-based face recognition which can solve the small sample size problem, Pattern Recogn. 33 (10) (2000) 1713–1726.
[2] H. Yu, J. Yang, A direct LDA algorithm for high-dimensional data–with application to face recognition, Pattern Recogn. 34 (2001) 2067–2070.
[3] 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.
[4] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, San Diego, 1990.
[5] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.