Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition 38 (2005) 1125–1129
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Rapid and brief communication

Two-dimensional discriminant transform for face recognition

Jian Yang[a,b,*], David Zhang[a], Xu Yong[b], Jing-yu Yang[b]

[a]*Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong*
[b]*Department of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, PR China*

Received 13 October 2004; accepted 4 November 2004

Abstract

This paper develops a new image feature extraction and recognition method coined *two-dimensional linear discriminant analysis* (2DLDA). 2DLDA provides a sequentially optimal image compression mechanism, making the discriminant information compact into the up-left corner of the image. Also, 2DLDA suggests a feature selection strategy to select the most discriminative features from the corner. 2DLDA is tested and evaluated using the AT&T face database. The experimental results show 2DLDA is more effective and computationally more efficient than the current LDA algorithms for face feature extraction and recognition.

Keywords: Fisher linear discriminant analysis (FLD or LDA); Fisherfaces; Feature extraction; Face recognition; Two-dimensional data analysis

1. Introduction

Fisher linear discriminant analysis (LDA) has been successfully applied to face recognition area in the past few years. LDA is a 1D-data-based feature extraction technique, so, 2D image matrices must be converted into 1D image vectors before the application of LDA. Since the resulting image vectors are high dimensional, LDA usually encounters the small sample size (S3) problem in which the within-class scatter matrix becomes singular and thus the traditional LDA algorithm fails to use. To address this problem, a number of extended LDA algorithms have been suggested. Among them, the most popular one is to use PCA for dimension reduction prior to performing LDA [1,2]. This method has a computational complexity of $\mathcal{O}(M^3)$. When the training sample size M is large, the computation requirement of this method is still considerable.

To avoid the S3 problem LDA encounters, Liu [3] suggested a 2D image matrix-based linear discriminant technique. His idea is to perform LDA directly based on image matrices, while overleaping the process of turning image matrices into vectors. Thus, the difficulty resulting from high-dimensionality is artfully avoided. As a further development of Liu's method, the *uncorrelated image matrix-based linear discriminant analysis* (IMLDA) technique was proposed recently [4]. IMLDA has an advantage to eliminate the correlation between discriminant feature vectors so that it is more effective than Liu's method for face recognition [4].

A drawback of IMLDA is that it needs more coefficients than LDA for image representation. Thus, IMLDA 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 method coined *two-dimensional linear discriminant analysis* (2DLDA) to overcome the disadvantage of IMLDA. The initial idea of 2DLDA is to perform IMLDA twice: the first one is in horizontal direction and the second is in vertical

* Corresponding author. Tel.: +852 2766 7312; fax: +852 2774 0842.

E-mail addresses: csjyang@comp.polyu.edu.hk (J. Yang), csdzhang@comp.polyu.edu.hk (D. Zhang), laterfall@sina.com (X. Yong), yangjy@mail.njust.edu.cn (J.-y. Yang).

0031-3203/$30.00 © 2005 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2004.11.019

direction. After the two sequential IMLDA transforms, the discriminant information is compacted into the up-left corner of the image. A feature selection mechanism is followed to select the most discriminative features from the corner. The effectiveness of the proposed method is verified using AT&T database.

2. Outline of IMLDA

Suppose there are c known pattern classes. M is the total number of training samples, and M_i is the number of training samples in class i. In class i, the jth training image is denoted by an $m \times n$ matrix $\mathbf{A}_j^{(i)}$. The mean image of training samples in class i is denoted by $\bar{\mathbf{A}}^{(i)}$ and the mean image of all training sample is $\bar{\mathbf{A}}$.

Based on the given training image samples (image matrices), the *image between-class scatter matrix* and *image within-class scatter matrix* can be constructed by

$$\mathbf{G}_b = \frac{1}{M}\sum_{i=1}^{c} M_i(\bar{\mathbf{A}}_i - \bar{\mathbf{A}})^{\mathrm{T}}(\bar{\mathbf{A}}_i - \bar{\mathbf{A}}), \tag{1}$$

$$\mathbf{G}_w = \frac{1}{M}\sum_{i=1}^{c}\sum_{j=1}^{M_i}(\mathbf{A}_j^{(i)} - \bar{\mathbf{A}}^{(i)})^{\mathrm{T}}(\mathbf{A}_j^{(i)} - \bar{\mathbf{A}}^{(i)}). \tag{2}$$

By the definition, it is easy to verify that $\mathbf{G}_b$ and $\mathbf{G}_w$ are both $n \times n$ nonnegative definite matrices. It should be mentioned that in face recognition problems, $\mathbf{G}_w$ is usually invertible unless that there is only one training sample per class.

The generalized Fisher criterion can be defined by

$$J(\boldsymbol{\varphi}) = \frac{\boldsymbol{\varphi}^{\mathrm{T}}\mathbf{G}_b\boldsymbol{\varphi}}{\boldsymbol{\varphi}^{\mathrm{T}}\mathbf{G}_w\boldsymbol{\varphi}}. \tag{3}$$

It is easy to find a vector $\mathbf{u}^*$ to maximize the Rayleigh quotient function $J(\boldsymbol{\varphi})$. After the projection of samples onto $\mathbf{u}^*$, the ratio of the between-class scatter to the within-class scatter is maximized. So, the vector $\mathbf{u}^*$ is called the optimal image projection direction. Generally, a single projection direction is not enough for the discrimination of multiclass problems so that we need a set of discriminant vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_q$, which maximize the generalized Fisher criterion and satisfy $\mathbf{G}_t$-orthogonal constraints, i.e.,

$$\mathbf{u}_i^{\mathrm{T}}\mathbf{G}_t\mathbf{u}_j = 0, \quad \text{where } \mathbf{G}_t = \mathbf{G}_b + \mathbf{G}_w,\ i \neq j,\quad i, j = 1, \ldots, q. \tag{4}$$

The role of these constraints is to make the resulting discriminant feature vectors uncorrelated and thereby more discriminative for classification [4].

Actually, the discriminant feature vectors subject to the above constraints can be selected as the $\mathbf{G}_t$-orthogonal generalized eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_q$ of $\mathbf{G}_b$ and $\mathbf{G}_w$ corresponding to q largest generalized eigenvalues, i.e., $\mathbf{G}_b\mathbf{u}_j = \lambda_j\mathbf{G}_w\mathbf{u}_j$, where $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_q$. These eigenvectors can be calculated using the algorithm suggested in Ref. [4]. The obtained eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_q$ are used for image feature extraction. Let

$$\mathbf{B} = \mathbf{AU}, \quad \text{where } \mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_q). \tag{5}$$

The resulting feature matrix $\mathbf{B}$ is used to represent image $\mathbf{A}$ for classification.

3. 2DLDA

3.1. Idea

IMLDA 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 LDA 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 IMLDA. Our idea is simple, just to perform IMLDA 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 $\mathbf{A}$, we obtain its feature matrix $\mathbf{B}$ after the first IMLDA transform. Then, we transpose $\mathbf{B}$ and input $\mathbf{B}^{\mathrm{T}}$ into IMLDA, and determine the transform matrix $\mathbf{V}$. Projecting $\mathbf{B}^{\mathrm{T}}$ onto $\mathbf{V}$, we obtain $\mathbf{C}^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{V}$. The resulting feature matrix is $\mathbf{C} = \mathbf{V}^{\mathrm{T}}\mathbf{B}$. This process is illustrated in Fig. 1.

In the whole process, the first IMLDA transform $\mathbf{B} = \mathbf{AU}$ performs the compression of 2D-data in horizontal direction, making the discriminant information pack into a small number of columns. While the second IMLDA transform $\mathbf{C} = \mathbf{V}^{\mathrm{T}}\mathbf{B}$ performs the compression of 2D-data in vertical direction, eliminating the correlations between columns of image $\mathbf{B}$ and making its discriminant information further compact into a small number of rows. Ultimately, the discriminant information of the whole image is packed into the up-left corner of the image matrix.

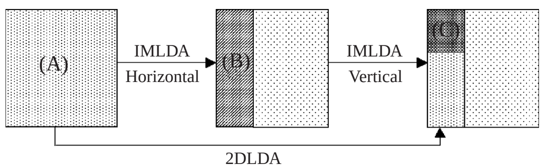

Fig. 1. Illustration of 2DLDA transform.

Table 1
The computational complexity of Fisherfaces [1], Enhanced Fisher Model (EFM) [2], Direct LDA [5], and the proposed 2DLDA

Method	Fisherfaces [1]	EFM [2]	Direct LDA [5]	2DLDA
Computational complexity	$\mathcal{O}(M^3)$	$\mathcal{O}(M^3)$	$\mathcal{O}(c^3)$	$\mathcal{O}(l^3)$, where $l=\max\{m,n\}$

3.2. Transform

Now, let us present the detailed implementation of 2DLDA. After the first IMLDA transform in horizontal direction, we get the feature matrix $\mathbf{B}$ of sample $\mathbf{A}$ using Eq. (5). Constructing the *image between-class* and *within-class scatter matrices* $\mathbf{H}_b$ and $\mathbf{H}_w$ based on $\mathbf{B}^{\mathrm{T}}$, we have

$$\mathbf{H}_b = \frac{1}{M}\sum_{i=1}^{c} M_i(\bar{\mathbf{B}}_i - \bar{\mathbf{B}})(\bar{\mathbf{B}}_i - \bar{\mathbf{B}})^{\mathrm{T}}, \tag{6}$$

$$\mathbf{H}_w = \frac{1}{M}\sum_{i=1}^{c}\sum_{j=1}^{M_i}(\mathbf{B}_j^{(i)} - \bar{\mathbf{B}}^{(i)})(\mathbf{B}_j^{(i)} - \bar{\mathbf{B}}^{(i)})^{\mathrm{T}}, \tag{7}$$

where $\mathbf{B}_j^{(i)} = \mathbf{A}_j^{(i)}\mathbf{U}$, $\bar{\mathbf{B}}^{(i)} = \bar{\mathbf{A}}^{(i)}\mathbf{U}$, and $\bar{\mathbf{B}} = \bar{\mathbf{A}}\mathbf{U}$.

It is easy to show $\mathbf{H}_b$ and $\mathbf{H}_w$ are both $m\times m$ nonnegative definite matrices. Generally, $\mathbf{H}_w$ is invertible. Let $\mathbf{H}_t = \mathbf{H}_b + \mathbf{H}_w$. Suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are the $\mathbf{H}_t$-orthogonal generalized eigenvectors of $\mathbf{H}_b$ and $\mathbf{H}_w$ corresponding to p largest eigenvalues $\mu_1, \mu_2, \ldots, \mu_p$. Let $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p)$. We get the IMLDA feature matrix of $\mathbf{B}^{\mathrm{T}}$ by

$$\mathbf{C}^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{V}. \tag{8}$$

Thus

$$\mathbf{C} = \mathbf{V}^{\mathrm{T}}\mathbf{B} = \mathbf{V}^{\mathrm{T}}\mathbf{A}\mathbf{U}. \tag{9}$$

The resulting feature matrix $\mathbf{C}$ is a $p\times q$ matrix, which is much smaller than the IMLDA feature matrix $\mathbf{B}$ and the original image $\mathbf{A}$ since p and q are always chosen much smaller than m and n.

3.3. Feature selection strategy

Since the up-left corner feature matrix $\mathbf{C} = (c_{ij})_{p\times q}$ (after the 2DLDA transform) contains most of image discriminant information, it is reasonable to choose the most discriminative features from it for recognition purpose. Here, we suggest a simple way to derive features from $\mathbf{C}$.

Based on the physical meaning of the generalized Fisher criterion, the discrimination power of the jth column of $\mathbf{C}$ can be characterized by λ_j (the jth largest eigenvalue of $\mathbf{G}_b\boldsymbol{\varphi} = \lambda\mathbf{G}_w\boldsymbol{\varphi}$), while the discrimination power of the ith row of $\mathbf{C}$ can be characterized by μ_i (the ith largest eigenvalue of $\mathbf{H}_b\boldsymbol{\varphi} = \mu\mathbf{H}_w\boldsymbol{\varphi}$). So, we can measure the discrimination power of the component c_{ij} by

$$\gamma_{ij} = \mu_i\lambda_j. \tag{10}$$

We will rank $\mathbf{C}$'s components according to their corresponding discrimination power and arrange them into a feature vector for classification.

3.4. Computational advantages

The computational complexity of Fisherfaces [1], Enhanced Fisher Model (EFM) [2], Direct LDA [5], and the proposed 2DLDA are listed in Table 1. Obviously, the computation requirements of Fisherfaces and EFM increase cubically with the increase of the training sample size M, while the computation requirement of Direct LDA does with the increase of the class number c. Whereas, the computation scale of 2DLDA only depends on the size of image, i.e., the number of rows (m) and columns (n) of image matrix. For large-scaled databases where m or n is far less than M and c, 2DLDA is computationally more efficient than other methods mentioned above.

4. Experiments

The proposed method is tested using the standard AT&T database (http://www.uk.research.att.com/facedatabase.html). This database contains images from 40 individuals, each providing 10 different images. In our experiments, we split the whole database into two parts evenly. One part is used for training and the other part is for testing. In order to make full use of the available data and to evaluate the generalization power of algorithms more accurately, we adopt a cross-validation strategy and run the system 20 times. In each time, five face images from each person are randomly selected as training samples, and the rest is for testing. Fisherfaces [1], EFM [2], Direct LDA [5], and the proposed 2DLDA are used for feature extraction. Note that for 2DLDA, we choose $p = q = 8$. Finally, a nearest-neighbour classifier is employed for classification. The average recognition rate across 20 tests of each method over the variation of dimensions is plotted in Fig. 2. Fig. 2 shows 2DLDA consistently outperforms others irrespective of the variation in dimensions.

In addition, the average CPU time consumed for training and testing, the top recognition rates and the corresponding dimensions of the foregoing four methods and IMLDA [4] are given in Table 2. 2DLDA achieves its maximal recognition rate of 96.4% using only 25 features and, it needs less CPU time compared to other methods. Although

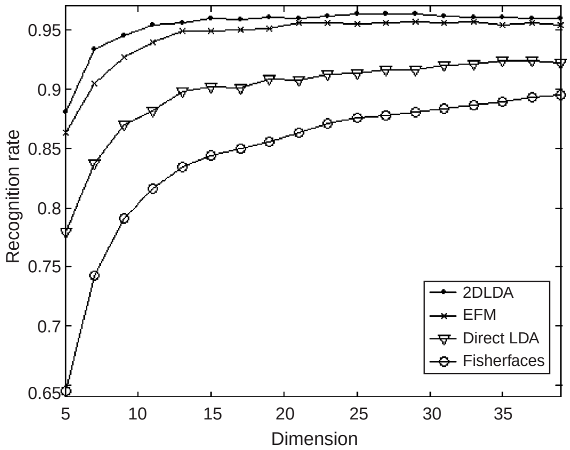

Fig. 2. The average recognition rate (%) across 20 tests of each method over the variation of dimensions.

Table 2
The average CPU time (s) consumed for training and testing, the top recognition rates (%) and the corresponding dimensions of the five methods (CPU: PIII 800 MHz, RAM: 256M)

Method	Fisherfaces	Direct LDA	EFM	IMLDA	2DLDA
Recognition rate	89.5	92.4	95.7	95.8	96.4
Dimension	39	35	29	224	25
CPU time	65.01	43.17	52.79	45.42	35.08

IMLDA is faster than 2DLDA for training, it still needs more time for the whole process (training and testing) because it costs more computation using 224 features for classification. It should be noted that the speed difference between 2DLDA and other methods would become more significant with the increase of face database scale (e.g. the database with more training sample size and class number).

Acknowledgements

This research was supported by the CERG fund from the HKSAR Government and the central fund from the Hong Kong Polytechnic University, the National Science Foundation of China under Grant No. 60332010, No. 60472060, and No. 60473039. Dr. Yang was supported by China and the Hong Kong Polytechnic University Postdoctoral Fellowships.

References

[1] P.N. Belhumeur, J.P. Hespanha, D.J. Kriengman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Machine Intell. 19 (7) (1997) 711–720.

[2] C.J. Liu, H. Wechsler, Robust coding schemes for indexing and retrieval from large face databases, IEEE Trans. Image Process. 9 (1) (2000) 132–137.

[3] K. Liu, Y.-Q. Cheng, J.-Y. Yang, et al., Algebraic feature extraction for image recognition based on an optimal discriminant criterion, Pattern Recognition 26 (6) (1993) 903–911.

[4] J. Yang, J.-y. Yang, A.F. Frangi, D. Zhang, Uncorrelated projection discriminant analysis and its application to face image feature extraction, Internat. J. Pattern Recogn. Artif. Intell. 17 (8) (2003) 1325–1347.

[5] H. Yu, J. Yang, A direct LDA algorithm for high-dimensional data—with application to face recognition, Pattern Recognition 34 (10) (2001) 2067–2070.

About the Author—JIAN YANG was born in Jiangsu, China, June 1973. He obtained his Bachelor of Science in Mathematics at the Xuzhou Normal University in 1995. He then continued to complete a Masters of Science degree in Applied Mathematics at the Changsha Railway University in 1998 and his Ph.D. at the Nanjing University of Science and Technology (NUST) in the Department of Computer Science on the subject of Pattern Recognition and Intelligence Systems in 2002. From March to September in 2002, he worked as a research assistant at department of computing, Hong Kong Polytechnic University. From January to December in 2003, he was a postdoctoral researcher at the University of Zaragoza and affiliated with the Division of Bioengineering of the Aragon Institute of Engineering Research (I3A). Now, he is a research associate at Hong Kong Polytechnic University with biometrics center and postdoctoral research fellow at NUST. He is the author of more than 30 scientific papers in pattern recognition and computer vision. His current research interests include pattern recognition, computer vision and machine learning.

About the Author—DAVID ZHANG graduated in computer science from the Peking University in 1974 and received his M.Sc. and Ph.D. degrees in computer science & engineering from the Harbin Institute of Technology (HIT) in 1983 and 1985, respectively. He received his second Ph.D. in electrical and computer engineering at the University of Waterloo, Ontario, Canada, in 1994. After that, he was an associate professor at the City University of Hong Kong and a professor at the Hong Kong Polytechnic University. Currently, he is a founder and director of the Biometrics Technology Centre supported by the UGC of the Government of the Hong Kong SAR. He is the Founder and Editor-in-Chief of the International Journal of Image and Graphics, and an Associate Editor in some international journals such as IEEE Trans. on SMC-C, Pattern Recognition, and International Journal of Pattern Recognition and Artificial Intelligence. His research interests include automated biometrics-based identification, neural systems and applications, and image processing and pattern recognition. So far, he has published over 180 articles as well as ten books, and won numerous prizes.

About the Author—YONG XU was born in Sichuan, China, in 1972. He received his B.S. degree and M.S. degree in 1994 and 1997, respectively. Now, he is working for his Ph.D. degree in Pattern Recognition and Intelligence System. He is the author of more than ten scientific papers in pattern recognition and image processing. His current interests include face recognition and detection, character recognition and image processing.

About the Author—JING-YU YANG received the B.S. Degree in Computer Science from Nanjing University of Science and Technology (NUST), Nanjing, China. From 1982 to 1984 he was a visiting scientist at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. From 1993 to 1994 he was a visiting professor at the Department of Computer Science, Missuria University. And in 1998, he acted as a visiting professor at Concordia University in Canada. He is currently a professor and Chairman in the department of Computer Science at NUST. He is the author of over 300 scientific papers in computer vision, pattern recognition, and artificial intelligence. He has won more than 20 provincial awards and national awards. His current research interests are in the areas of pattern recognition, robot vision, image processing, data fusion, and artificial intelligence.