PERGAMON

Available at www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT

Pattern Recognition 37 (2004) 381–384

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Rapid and Brief Communication

A novel method for Fisher discriminant analysis

Yong Xu*, Jing-Yu Yang, Zhong Jin

Department of Computer Science, Nanjing University of Science and Technology, Nanjing Jiangsu 210094, People's Republic of China

Received 19 May 2003; accepted 3 June 2003

Abstract

A novel model for Fisher discriminant analysis is developed in this paper. In the new model, maximal Fisher criterion values of discriminant vectors and minimal statistical correlation between feature components extracted by discriminant vectors are simultaneously required. Then the model is transformed into an extreme value problem, in the form of an evaluation function. Based on the evaluation function, optimal discriminant vectors are worked out. Experiments show that the method presented in this paper is comparative to the winner between FSLDA and ULDA.

Keywords: Fisher discriminant analysis; FSLDA (Foley–Sammon linear discriminant analysis); ULDA (uncorrelated linear discriminant analysis)

1. Introduction

Fisher discriminant analysis is a very important method for feature extraction. Foley–Sammon linear discriminant analysis (FSLDA) and uncorrelated linear discriminant analysis (ULDA) are well known as two kinds of Fisher discriminant analysis. Studies show that the Fisher criterion value of each FSLDA discriminant vector is always not less than that of corresponding ULDA discriminant vector. This can be regarded as an advantage for FSLDA, because for a discriminant vector greater Fisher criterion value means more powerful discriminability. On the other hand, ULDA always obtains the uncorrelated feature components, whereas FSLDA usually obtains correlative feature components and sometimes the feature components are highly correlative to each other [2,3]. For discriminant vectors, it appears that the less correlative to each other the feature components extracted by them are, the better the discriminant vectors are [2]. So, in this sense, extracting uncorrelated feature components is an advantage for ULDA. It seems that ideal discriminant vectors not only correspond to maximal Fisher criterion values but also correspond to minimal correlations between extracted feature components. However, unfortunately, neither of ULDA and FSLDA can achieve this kind of ideal discriminant vectors. Moreover, both FSLDA and ULDA are not very efficient. We may suppose that there is a balance between great Fisher criterion values and low correlation between extracted feature components, for discriminant vectors, and under that equilibrium the performance of corresponding discriminant vectors is excellent. In this paper, an attempt for this is made. A novel problem model for Fisher discriminant analysis is proposed. In the novel model, both maximal Fisher criterion values and minimal correlations between extracted feature components are taken as objective functions for optimal discriminant vectors. After the problem model is transformed into an evaluation function, optimal discriminant vectors are worked out. The experiments show that the method proposed in this paper can achieve good effect.

2. Algorithms for three kinds of Fisher discriminant analysis

2.1. Algorithms for FSLDA and ULDA

Our discussion on discriminant vector is based on the Fisher criterion

$$J'(\varphi) = \frac{\varphi^{\mathrm{T}} S_b \varphi}{\varphi^{\mathrm{T}} S_t \varphi}, \tag{1}$$

* Corresponding author. Tel: +86-25-882-7574; fax: +86-25-431-5510.

E-mail addresses: laterfall@sohu.com (Y. Xu), yangjy@mail.njust.edu.cn (J.-Y. Yang).

0031-3203/$30.00
doi:10.1016/S0031-3203(03)00232-2

where S_b and S_t are the between-class scatter matrix and the total population scatter matrix respectively. Suppose S_t is positive definite. The extreme value problem for $J'(\varphi)$ is often discussed with the generalized eigenequation

$$S_b\varphi = \lambda S_t\varphi. \tag{2}$$

The eigenvector corresponding to maximum eigenvalue of Eq. (2), denoted by φ_1, is simultaneously taken as the first FSLDA vector and the first ULDA vector. Based on the Jin's algorithm [2], the eigenvector corresponding to maximum eigenvalue of Eq. (3) can be taken as the ith FSLDA discriminant vector ($i > 1$):

$$MS_b\varphi = \lambda S_t\varphi, \tag{3}$$

where $M = I - D^{\mathrm{T}}(DS_t^{-1}D^{\mathrm{T}})^{-1}DS_t^{-1}$, $D = [\varphi_1 \ \ \varphi_2 \ \ \dots \ \ \varphi_{i-1}]^{\mathrm{T}}$. I is the identity matrix. $\varphi_1, \varphi_2, \dots, \varphi_{i-1}$ are the previous $i-1$ FSLDA discriminant vectors. The eigenvector corresponding to maximum eigenvalue of Eq. (4) can be taken as the ith ULDA discriminant vector ($i > 1$):

$$MS_b\varphi = \lambda S_t\varphi, \tag{4}$$

where $M = I - S_tD^{\mathrm{T}}(DS_tD^{\mathrm{T}})^{-1}D$, $D = [\varphi_1 \ \ \varphi_2 \ \ \dots \ \ \varphi_{i-1}]^{\mathrm{T}}$. I is also the identity matrix. $\varphi_1, \varphi_2, \dots, \varphi_{i-1}$ are the previous $i-1$ ULDA discriminant vectors.

2.2. A novel problem model

First of all, some definitions are necessary to be introduced. As for feature extraction, every discriminant vector corresponds to a feature component. For the same pattern X in the original feature space, the feature component corresponding to φ_i is $y_i = \varphi_i^{\mathrm{T}}X$, and the feature component corresponding to φ_j is $y_j = \varphi_j^{\mathrm{T}}X$. The covariance between y_i and y_j is defined as [3]

$$\mathrm{cov}(y_i, y_j) = E[(y_i - Ey_i)(y_j - Ey_j)] = \varphi_i^{\mathrm{T}}S_t\varphi_j \tag{5}$$

and the correlation coefficient between y_i and y_j is defined by the following formula:

$$\rho(y_i, y_j) = \frac{\mathrm{cov}(y_i, y_j)}{\sqrt{\mathrm{cov}(y_i, y_i)}\sqrt{\mathrm{cov}(y_j, y_j)}} = \frac{\varphi_i^{\mathrm{T}}S_t\varphi_j}{\sqrt{\varphi_i^{\mathrm{T}}S_t\varphi_i}\sqrt{\varphi_j^{\mathrm{T}}S_t\varphi_j}}. \tag{6}$$

In this paper, $\rho(y_i, y_j)$ is also denoted by $f(\varphi_i, \varphi_j)$, which means that the correlation coefficient is defined for the feature components extracted by φ_i and φ_j.

Since both of great Fisher criterion values and low correlations between extracted feature components are important for discriminant vectors, the following problem model is proposed.

The vector corresponding to maximum eigenvalue of generalized eigenequation (2) is selected as the first discriminant vector. The kth discriminant vector is worked out from the following ($k > 1$):

$$\begin{aligned} &\max J'(\varphi), \\ &\min f_1^2(\varphi, \varphi_1), \\ &\cdot \\ &\cdot \\ &\cdot \\ &\min f_{k-1}^2(\varphi, \varphi_{k-1}), \end{aligned} \tag{7}$$

where

$$f_j(\varphi, \varphi_j) = \frac{\varphi_j^{\mathrm{T}}S_t\varphi}{\sqrt{\varphi_j^{\mathrm{T}}S_t\varphi_j}\sqrt{\varphi^{\mathrm{T}}S_t\varphi}}, \quad j = 1, 2, \dots, k-1.$$

The greater $f_j^2(\varphi, \varphi_j)$ is, the higher the correlation between the feature components extracted by φ and φ_j is. This model requires that the feature component extracted by the kth discriminant vector has the lowest correlation with anyone of those extracted by the previous $k-1$ discriminant vectors, meantime the kth discriminant vector is the available vector corresponding to maximal Fisher criterion value. By contrast with ULDA, the discriminant vectors obtained from this model may correspond to greater Fisher criterion values, and different from FSLDA, the feature components extracted by the discriminant vectors from the model are always only little correlative to each other.

2.3. Solution to the novel problem model

Both $J'(\varphi)$ and $f_j^2(\varphi, \varphi_j)$vary from 0 to 1. We transform the above problem model into an evaluation function

$$P(\varphi) = r_0J'(\varphi) - \sum_{j=1}^{k-1} r_jf_j^2(\varphi, \varphi_j), \tag{8}$$

where $r_j > 0$ ($j = 0, 1, 2, \dots, k-1$) is weighting coefficient and $\sum_{j=0}^{k-1} r_j = 1$. It is certain that the greater $J'(\varphi)$ is and the smaller $f_j^2(\varphi, \varphi_j)$ is, the greater $P(\varphi)$ is. So, the kth discriminant vector should be the vector corresponding to maximal $P(\varphi)$. Because of

$$P(\varphi) = r_0\frac{\varphi^{\mathrm{T}}S_b\varphi}{\varphi^{\mathrm{T}}S_t\varphi} - \sum_{j=1}^{k-1} r_j\frac{(\varphi_j^{\mathrm{T}}S_t\varphi)^2/\varphi_j^{\mathrm{T}}S_t\varphi_j}{\varphi^{\mathrm{T}}S_t\varphi} \tag{9}$$

and $P(\mu\varphi) = P(\varphi)$, where μ is an arbitrary constant and not zero, the extreme value problem for $P(\varphi)$ can be transformed into the constrained maximization problem

$$\begin{aligned} &\max P'(\varphi), \\ &\varphi^{\mathrm{T}}S_t\varphi = 1, \end{aligned} \tag{10}$$

where $P'(\varphi) = r_0\varphi^{\mathrm{T}}S_b\varphi - \sum_{j=1}^{k-1} r_j(\varphi_j^{\mathrm{T}}S_t\varphi)^2/\varphi_j^{\mathrm{T}}S_t\varphi_j$. By the method of Lagrange multipliers, the following function can be obtained:

$$L(\varphi) = P'(\varphi) - \lambda(\varphi^{\mathrm{T}}S_t\varphi - 1), \tag{11}$$

where λ is multiplier. When $P(\varphi)$ corresponds to maximum value, the partial derivative of $L(\varphi)$ with respect to φ will be equal to zero, i.e.

$$2r_0 S_b\varphi - 2\lambda S_t\varphi - 2\sum_{j=1}^{k-1} r_j S_t\varphi_j(\varphi_j^{\mathrm{T}} S_t\varphi)/(\varphi_j^{\mathrm{T}} S_t\varphi_j) = 0. \tag{12}$$

Caused by left-hand multiplier φ^{T}, the following can be derived from Eq. (12):

$$\lambda = r_0\frac{\varphi^{\mathrm{T}} S_b\varphi}{\varphi^{\mathrm{T}} S_t\varphi} - \sum_{j=1}^{k-1} r_j\frac{(\varphi_j^{\mathrm{T}} S_t\varphi)^2/\varphi_j^{\mathrm{T}} S_t\varphi_j}{\varphi^{\mathrm{T}} S_t\varphi}. \tag{13}$$

Maximizing λ is the sake. It is sure that $S_t\varphi_j(\varphi_j^{\mathrm{T}} S_t\varphi) = S_t(\varphi_j\varphi_j^{\mathrm{T}})S_t\varphi$. So, Eq. (14) is achieved:

$$M\varphi = \lambda S_t\varphi, \tag{14}$$

where

$$M = r_0 S_b - \sum_{j=1}^{k-1} r_j S_t(\varphi_j\varphi_j^{\mathrm{T}})S_t/(\varphi_j^{\mathrm{T}} S_t\varphi_j). \tag{15}$$

So, solving for the kth discriminant vector can be presented by Theorem 1.

Theorem 1. *The kth ($K > 1$) discriminant vector is the vector corresponding to maximum eigenvalue of generalized eigenequation* (14).

3. Experiments

In the following two experiments, $r_j = 1/k$ ($j = 0, 1, 2, \ldots, k-1$) is adopted for solving the kth discriminant vector. Consequently, r_j can be neglected in computing. In addition, we directly use the results of $\sum_{j=1}^{k-2} S_t(\varphi_j\varphi_j^{\mathrm{T}})S_t/(\varphi_j^{\mathrm{T}} S_t\varphi_j)$, obtained in the previous steps, to economize computational time. As a result, the computational complexity of M in our method is not greater than those of FSLDA and ULDA, without matrix inverse in the formulation.

3.1. Experiment on CENPARMI handwritten numeral database

An experiment on Concordia University CENPARMI handwritten numeral database is performed. The two well-known kinds of features, 256-dimensional Gabor transformation feature and 121-dimensional Legendre moment feature [1,4], are used in the experiment. There are 10 classes, i.e. 10 digits (from 0 to 9), and 600 samples for each. The training samples and testing samples are 4000 and 2000, respectively [3]. In this experiment, 30 FSLDA discriminant vectors corresponding to positive Fisher criterion values, all the available ULDA discriminant vectors and all the available discriminant vectors coming from problem model (7), corresponding to positive Fisher criterion values, are worked out. Feature extractions are respectively performed based on these three kinds of discriminant vectors, and then the classifications, according to the minimum distance classifier, are respectively performed based on the three kinds of extracted features.

Table 1
The classification error rates, on CENPARMI, based on the three kinds of discriminant vectors

	FSLDA	ULDA	Method presented in this paper
Gabor	27.4	19.9	19.8
Legendre	27.0	14.1	15.0

Table 1 indicates that the classification error rates achieved by FSLDA are the highest. The classification results of ULDA are good. Based on ULDA, the classification error rate on Gabor feature is 19.9%, and that on Legendre feature is 14.1%. The classification error rates achieved by our method are comparative to ULDA. Based on our method, the classification error rates on Gabor feature and Legendre feature are 19.8% and 15.0%, respectively.

3.2. Experiment on ORL face image database

ORL face image database is a widely used database [2]. In our experiment, the first five images of each subject are for training and the others are for testing and every image is treated as a vector. It is notable that S_t is singular and Eqs. (2)–(4) and (14) cannot be solved directly [2]. Each significant discriminant vector φ must satisfy $\varphi^{\mathrm{T}} S_b\varphi > 0$, according to Fisher's idea, so, the following approach is adopted. Firstly, all non-zero eigenvectors of S_b, corresponding to non-zero eigenvalues of S_b, are worked out. Suppose matrix P consists of all the non-zero eigenvectors. Let $\hat{S}_b = P^{\mathrm{T}} S_b P$, $\hat{S}_t = P^{\mathrm{T}} S_t P$. Secondly, based on the eigenequation $\hat{S}_b\varphi = \lambda\hat{S}_t\varphi$, the first discriminant vector of three kinds of discriminant vectors can be obtained. Then, the other discriminant vectors of three kinds of discriminant vectors can be successively worked out according to respective algorithms (S_b and S_t in the original algorithms are replaced by $\hat{S}_b$ and $\hat{S}_t$, respectively). New feature space can be formed based on two transformations. For every original pattern X, the first transformation is defined by $X' = P^{\mathrm{T}}X$. Then the second transformation is performed to form new feature Y, based on the formula $Y = \Phi^{\mathrm{T}}X'$, where Φ consists of all the discriminant vectors of some discriminant analysis method. In this experiment, the minimum distance classifier is also used to classify.

Table 2 shows the numbers of erroneous classification samples resulted from three kinds of discriminant

Table 2
The classification results, on ORL, based on the three kinds of discriminant vectors

	FSLDA	ULDA	Our method
Number of erroneous classification samples	29	30	28

analysis. The number of erroneous classification samples of our method is 28, and those of ULDA and FSLDA are 30 and 29, respectively.

References

[1] S.X. Liao, M. Pawlak, On image analysis by moments, IEEE Trans. Pattern Anal. Mach. Intel. 18 (3) (1996) 254–266.

[2] Z. Jin, J.Y. Yang, Z.S. Hu, Z. Lou, Face recognition based on the uncorrelated discriminant transformation, Pattern Recognition 34 (7) (2001) 1405–1416.

[3] J. Yang, J.Y. Yang, D. Zhang, What's wrong with Fisher criterion? Pattern Recognition 35 (2002) 2665–2668.

[4] H. Yoshihiko, et al., Recognition of handwritten numerals using Gabor features, Proceedings of the 13th ICPR (1996), pp. 250–253.

About the Author—YONG XU was born in Sichuan, China, in 1972. He received his B.S. 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 an author of eight scientific papers in pattern recognition and image processing. His current interests include face recognition and detection, handwritten character recognition and image processing.

About the Author—JING-YU YANG received the B.S. degree in Computer Science from NUST, Nanjing, China. He is the author of over 100 scientific papers in computer vision, pattern recognition, and artificial intelligence. His current research interests are in the areas of pattern recognition, image processing and artificial intelligence, and expert system.

About the Author—ZHONG JIN was born in Jiangsu, China, in 1961. He received his B.S., M.S. and Ph.D. degree in 1982, 1984 and 1999, respectively. He is now a professor in the Department of Computer Science at NUST. He is the author of over 10 scientific papers in pattern recognition, image processing, and artificial intelligence. His current interests are in the areas of pattern recognition, video image processing, face recognition.