Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition 38 (2005) 2192 – 2194
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Rapid and brief communication

Inverse Fisher discriminate criteria for small sample size problem and its application to face recognition

Xiao-Sheng Zhuang, Dao-Qing Dai*

Department of Mathematics, Faculty of Mathematics and Computing, Sun Yat-Sen (Zhongshan) University, Guangzhou 510275, China

Received 1 February 2005; accepted 11 February 2005

Abstract

This paper addresses the small sample size problem in linear discriminant analysis, which occurs in face recognition applications. Belhumeur et al. [IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720] proposed the FisherFace method. We find out that the FisherFace method might fail since after the PCA transform the corresponding within class covariance matrix can still be singular, this phenomenon is verified with the Yale face database. Hence we propose to use an inverse Fisher criteria. Our method works when the number of training images per class is one. Experiment results suggest that this new approach performs well.

Keywords: Linear discriminant analysis; Small sample size problem; Inverse Fisher discriminate criteria; Face recognition

1. Introduction

Face recognition is now a very active field for research [1]. Linear (Fisher) discriminant analysis (LDA) is a well-known and popular statistical method in pattern recognition and classification. In the last decade, LDA approach has been successfully applied in the face recognition technology. Some of the LDA-based face recognition systems have also been developed and encouraging results have been achieved. However, LDA approach suffers from a small sample size problem. A well-known approach, called FisherFace, to avoid the small sample size problem for face recognition was proposed by Belhumeur et al. [2]. This method consists of two steps. The first step is the use of principal component analysis (PCA) for dimension reduction. The second step is the application of LDA for the transformed data. The basic idea is that after the PCA step the covariance matrix for the transformed data is not singular. We find out that this PCA step cannot guarantee the successful application of subsequent LDA, the transformed covariance matrix might still be singular. This phenomenon occurs in the YALE database, see the experiment section for details. In view of this limitation we propose to use a new criteria to deal with the small sample size problem.

2. The inverse Fisher criteria

The basic idea of Fisher discriminate analysis is to maximize the Fisher quotient

$$\max_{W\in\mathbb{R}^d} \frac{W^{\mathrm{T}}C_bW}{W^{\mathrm{T}}C_wW}. \tag{1}$$

The rank of the within class covariance matrix $C_w \in \mathbb{R}^{d\times d}$ satisfies $rank(C_w) \leqslant \min\{d, \# - c\}$, where # is the number of total training samples, c is the number of classes. When small sample size problem occurs, i.e. $\# < d + c$, the matrix C_w is singular, hence the optimization problem (1) is not

* Corresponding author. Tel.: +86 20 8411 3190; fax: +86 20 8403 7978.
E-mail address: stsddq@zsu.edu.cn (D.-Q. Dai).

0031-3203/$30.00 © 2005 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2005.02.011

solvable. Instead we propose the new *inverse Fisher criteria*, minimization of the following quotient:

$$\min_{W \in \mathbb{R}^d} \frac{W^T C_w W}{W^T C_b W}. \tag{2}$$

We notice that the rank $rank(C_b)$ of the between class covariance matrix $C_b \in \mathbb{R}^{d\times d}$ satisfies $rank(C_b) \leqslant c-1$. Thus it can be singular. To solve the optimization problem (2) dimension reduction is needed. We shall not justify the use of (2) in this brief communication and leave it to the full version. Our scheme is as follows.

Inverse Fisher Algorithm

(1) *PCA step*: From the equation $C_t = C_b + C_w$, where C_t is the total covariance matrix, we have the singular value decomposition $C_t = U^T \Lambda U$, where $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_g, 0, \ldots, 0)$ is the diagonal matrix, whose diagonal entries are the eigenvalues of C_t arranged by $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_g > 0$, where g is the rank of C_t. $U = (u_1, u_2, \ldots, u_d)$ is a unitary matrix $U^T U = Id$ so that u_i is eigenvectors corresponding to λ_i for $i = 1, 2, \ldots, g$.
(2) *Eigenvector selection*: Among $\{u_1, u_2, \ldots, u_g\}$ we use the selection rule: for $i = 1, 2, \ldots, g$ if $u_i^T C_b u > u_i^T C_w u_i$ then u_i is selected. We get $p(p \leqslant \min\{g, c-1\})$ eigenvectors $\{u_{i_1}, u_{i_2}, \ldots, u_{i_p}\}$.
(3) *First projection*: $\mathbb{R}^d \to \mathbb{R}^p$ with $T_1 = (u_{i_1}, u_{i_2}, \ldots, u_{i_p}) \in \mathbb{R}^{d\times p}$, $x' = T_1^T x$. The between class covariance matrix $C_b' \in \mathbb{R}^{p\times p}$ of the dimension reduced samples is of full rank p. This projection is different from that used in the FisherFace method [2].
(4) *Inverse Fisher criteria*:

$$\min_{v \in \mathbb{R}^p} \frac{v^T C_w' v}{v^T C_b' v}, \tag{3}$$

where $C_w' \in \mathbb{R}^{p\times p}$ is the within class covariance matrix for x'.
(5) *Eigenvalue problem*: the optimization problem (3) is solved by $(C_b')^{-1} C_w' y = \mu y$, with eigenvalues $0 \leqslant \mu_1 \leqslant \ldots \leqslant \mu_p$, and corresponding normalized eigenvectors $y_1, y_2, \ldots, y_p$.
(6) *Second projection*: $\mathbb{R}^p \to \mathbb{R}^q$ with $T_2 = (y_1, y_2, \ldots, y_q) \in \mathbb{R}^{p\times q}$, where $q \leqslant p$. $x'' = T_2^T x'$.
(7) The *inverse Fisher transform*: $T \in \mathbb{R}^{d\times q}$ is given by $T = T_1 T_2$.

We call the columns of the transform T the inverse Fisher face (*IFFace*).

3. Experiment results

This section reports our experimental results. Three standard databases from Yale University, Olivetti Research Laboratory (ORL) and FERET are selected for evaluation. These databases could be utilized to test moderate variations in pose, illumination and facial expression. The Yale set contains 15 persons, each has 11 images with size 243×320, with different facial expressions and illuminations. The Olivetti set contains 400 images of 40 persons of size 92×112 with variations in pose, illumination and facial expression. For the FERET set we use 432 images of 72 persons, the image resolution is 92×112. As is known, the variations of the ORL database and the FERET database are different, we combine the two to get a new larger set, the ORLFERET, which has 832 images of 112 persons.

Table 1
Success rates of the FisherFace method with YALE database

# of training images	Success rate (%)
2	99.4
5	73.6
8	21.4

Failure rate of FisherFace for the YALE face database: As we pointed out in the introduction, the FisherFace method might not work, that is the within class covariance matrix of the PCA reduced data is still singular. We carry out 500 experiments. In Table 1 we list the failure rates. In the dimension reduction step with PCA, the image dimension is reduced to $d-c-1$. When the number of training images for each class is 2, the success rate of the FisherFace method is about 99.4%. When it becomes 5, the success rate is 73.6%. It becomes worse when the number of training images is 8 and the success rate decreases to 21.4%. This suggests that for the Yale database, strong correlation between images exists, which is in agreement with that of [2], the illumination variation of images of one person lies in a cone.

Recognition performance of the IFFace method: We test the proposed IFFace method with three experiments. We use the L^2 metric. For the classifier we use the nearest neighbor rule with class mean of each class. The recognition rate is calculated as the ratio of number of successful recognition and the total number of test samples. The experiments are repeated 50 times and average recognition rates are reported.

The first two experiments are concerned with performance of IFFace method on the ORL database and FERET database. They are shown in Fig. 1, the left part is for ORL and the right part is for FERET. The average recognition rates with the ORL database change from 74% to 96% when the number of training images per class increases from 2 to 9. For the more challenging FERET database IFFace method has even better performance, it changes from 85% to 94% for training images of 2–5.

In the third experiment, we test with the combined ORLFERET database. When the number of training images per class increases from 1 to 5, the recognition rate is from 66.7% to 92.5%. It is shown in Fig. 2. We notice that the

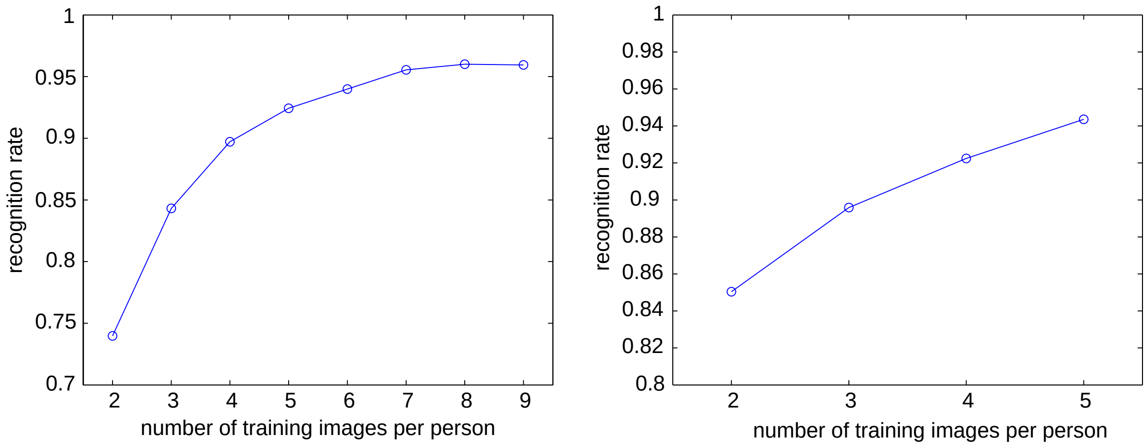

Fig. 1. Recognition rates for the ORL database (left) and FERET database (right).

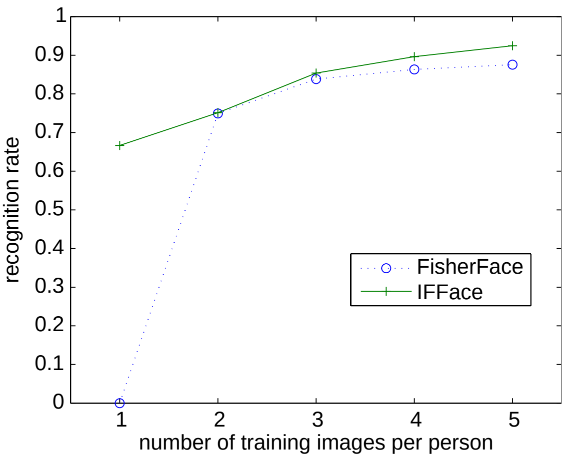

Fig. 2. Performance and comparison on the combined ORLFERET database.

IFFace method works even when the number of training images per class is one.

For the ORLFERET database the FisherFace method works. We compare it with IFFace in Fig. 2. IFFace has better performance. When the number of training images for each class is 5, the average recognition rates of 50 experiments are, respectively, 92.5% and 87.6% for IFFace and FisherFace.

4. Conclusion

In this paper we presented the IFFace method to solve the small sample size problem in face recognition. Compared with the two steps in FisherFace method we used different strategy. Experiments show that the method works well, especially when database have uneven variations for each class.

Acknowledgements

This project is supported in part by NSF of China (Grant No.: 60175031, 10231040), Grants from the Ministry of Education of China and Grants from Sun Yat-Sen University.

References

[1] A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, IEEE Trans. Circuits Syst. Video Technol. 14 (1) (2004) 4–20.

[2] 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.

About the Author—XIAO-SHENG ZHUANG received his Bachelor degree in Mathematics from Sun Yat-Sen (Zhongshan) University, China, in 2003. He is now working for his Master degree in Applied Mathematics. His current research interest includes face recognition and detection.

About the Author—DAO-QING DAI received his Ph.D. in Mathematics from Wuhan University, China, in 1990. He is currently a Professor and Associate Dean of the Faculty of Mathematics and Computing, Sun Yat-Sen (Zhongshan) University. He visited Free University, Berlin, as an Alexander von Humboldt research fellow from 1998 to 1999. He got the Outstanding Research Achievements in Mathematics award from ISAAC (International Society for Analysis, Applications and Computation) in 1999 at Fukuoka, Japan. He served as Programm Chair of Sinobiometrics' 2004. His current research interest includes image processing, wavelet analysis and human face recognition.