

Face Recognition by Inverse Fisher Discriminant Features

Xiao-Sheng Zhuang¹, Dao-Qing Dai^{1,*}, and P.C. Yuen²

¹ Center for Computer Vision and Department of Mathematics,
Sun Yat-Sen (Zhongshan) University, Guangzhou 510275 China
Tel: (86)(20)8411 3190; Fax: (86)(20)8403 7978
stsdq@mail.sysu.edu.cn

² Department of Computer Science, Hong Kong Baptist University, Hong Kong
pcyuen@comp.hkbu.edu.hk

Abstract. For face recognition task the PCA plus LDA technique is a famous two-pharse framework to deal with high dimensional space and singular cases. In this paper, we examine the theory of this framework: (1) LDA can still fail even after PCA procedure. (2) Some small principal components that might be essential for classification are thrown away after PCA step. (3) The null space of the within-class scatter matrix S_w contains discriminative information for classification. To eliminate these deficiencies of the PCA plus LDA method we thus develop a new framework by introducing an inverse Fisher criterion and adding a constrain in PCA procedure so that the singularity phenomenon will not occur. Experiment results suggest that this new approach works well.

1 Introduction

Face recognition [8, 18] technique has wide applications. Numerous algorithms have been proposed. Among various solutions, the most successful are those appearance-based approaches. Principle component analysis (PCA) and linear discriminant analysis (LDA) are two classic tools widely used in the appearance-based approaches for data reduction and feature extraction. Many state-of-the-art methods, such as Eigenfaces and Fisherfaces [2], are built on these two techniques or their variants. Although successful in many cases, in real-world applications, many LDA-based algorithms suffer from the so-called "small sample size problem" (SSS) [12]. Since SSS problem is common, it is necessary to develop new and more effective algorithms to deal with them. A number of regularization techniques that might alleviate this problem have been suggested [4-7]. Many researchers have been dedicated to searching for more effective discriminant subspaces [15-17].

A well-known approach, called Fisher discriminant analysis (FDA), to avoid the SSS problem was proposed by Belhumeur, Hespanha and Kriegman [2]. This method consists of two steps: PCA plus LDA. The first step is the use of principal

* Corresponding author.

component analysis for dimensionality reduction. The second step is the application of LDA for the transformed data. The basic idea is that after the PCA step the within-class scatter matrix for the transformed data is not singular. Although the effectiveness of this framework in face recognition are obvious, see [2, 9, 13, 18] and the theoretical foundation for this framework has been also laid [16] yet in this paper we find out that (1) LDA can still fail even after the PCA procedure. (2) Some small principal components that might be essential for classification are thrown away after PCA step. (3) The null space of the within-class scatter matrix S_w contains discriminative information for classification.

In this paper, motivated by the success and power of the PCA plus LDA in pattern classification tasks, considering the importance of the information in the null space of the within-class scatter matrix, and in view of the limitation of the PCA step, we propose a new framework for face recognition.

This paper is organized as follows. In Section 2, we start the analysis by briefly reviewing the two latter methods. We point out the deficiency of the PCA plus LDA method. Following that, our new method is introduced and analyzed in Section 3. In section 4, experiments are presented to demonstrate the effectiveness of the new method. Conclusions are summarized in Section 5.

2 The PCA Plus LDA Approach and Its Deficiency

Suppose that there are K classes, labelled as $G_1, G_2, \dots, G_K$. We randomly select n_j samples $X_j^{(i)}$ ($i = 1, 2, \dots, n_j$) from each class $G_j, j = 1, 2, \dots, K$ for training.

Set $N = \sum_{j=1}^K n_j$, $\mu_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_i^{(j)}, j = 1, 2, \dots, K$ and $\mu = \frac{1}{N} \sum_{j=1}^K \sum_{i=1}^{n_j} X_i^{(j)}$. Let the between-class scatter matrix and the within-class scatter matrix be defined by $S_b = \frac{1}{N} \sum_{j=1}^K n_j (\mu_j - \mu)(\mu_j - \mu)^T$, $S_w = \sum_{j=1}^K \sum_{i=1}^{n_j} (X_i^{(j)} - \mu_j)(X_i^{(j)} - \mu_j)^T$, $S_t = S_b + S_w$ is the total scatter matrix.

2.1 The PCA Procedure

PCA is a technique now commonly used for dimensionality reduction in face recognition. The goal of PCA is to find out a linear transformation or projection matrix $W_{PCA} \in \mathbb{R}^{d \times d'}$ that maps the original d -dimensional image space into an d' -dimensional feature space ($d' < d$) and maximize the determinant of the total scatter of the projected samples, i.e.,

$$W_{PCA} = \arg \max_{W \in \mathbb{R}^{d \times d'}} |W^T S_t W|. \quad (1)$$

2.2 The LDA Procedure

The aim of LDA is also to find a projection matrix as in PCA that maximizes the so-called **Fisher criterion**:

$$W_{LDA} = \arg \max_{W \in \mathbb{R}^{d \times d'}} \frac{|W^T S_b W|}{|W^T S_w W|}. \quad (2)$$

2.3 The Deficiency of PCA Plus LDA Approach

When applying the PCA plus LDA approach the following remarks should be considered.

- LDA can still fail even after PCA procedure. For the PCA projected data we get the matrix S'_w , S'_b and S'_t . Then there might exist a direction α such that $\alpha^T S'_t \alpha = \alpha^T S'_b \alpha$ so that $\alpha^T S'_w \alpha = 0$. Hence the matrix S'_w is still singular.
- Some small principal components that might be essential for classification are thrown away after PCA step. Since in PCA step, it just chooses d' eigenvectors corresponding to the first d' largest eigenvalues of S_t . It is very likely that the remainder contains some potential and valuable discriminatory information for the next LDA step.
- The null space of the within-class scatter matrix S_w contains discriminative information for classification. For a projection direct β , if $\beta^T S_w \beta = 0$ and $\beta^T S_b \beta \neq 0$, obviously, the optimization problem (2) is maximized.

3 Inverse Fisher Discriminant Analysis

In this section, we shall develop a new Fisher discriminant analysis algorithm based on **the inverse Fisher criterion**

$$W_{IFDA} = \arg \min_{W \in \mathbb{R}^{d \times d'}} \frac{|W^T S_w W|}{|W^T S_b W|}. \quad (3)$$

In contrast with LDA or FDA, we name the procedure using the above criterion as **the inverse Fisher discriminant analysis (IFDA)**. Obviously, the Fisher criterion (2) and inverse Fisher criterion (3) are equivalent, provided that the within-class scatter matrix S_w and the between-class scatter matrix S_b are not singular. However, we notice that the rank of the between-class scatter matrix $S_b \in \mathbb{R}^{d \times d}$ satisfies $\text{rank}(S_b) \leq K - 1$. Thus, the difficulty of SSS problem still exists for this new criterion.

On the other hand, let us come back to exploit the principle component analysis. For the optimization problem (1), it gives optimal projection vectors that have the largest variance and PCA just selects d' eigenvectors corresponding to the first d' largest eigenvalues of S_t but ignores the smaller ones. If we want to take those eigenvectors into account, we should abandon or modify such criterion for vector selection. Here we present a new criterion by modifying the equation (1) as follow:

$$\begin{aligned} W_{PCA-S} &= \arg \max_{W \in \mathbb{R}^{d \times d'}} |W^T S_t W| \\ &= [w_1 \ w_2 \ \cdots \ w_{d'}] \\ \text{s.t. } &w_i^T S_b w_i > w_i^T S_w w_i, \quad ||w_i|| = 1, \quad i = 1, 2, \dots, d' \end{aligned} \quad (4)$$

We name it as **PCA with selection (PCA_S)**. The reduced matrix $S'_b = W_{PCA-S}^T S_b W_{PCA-S}$ might still be singular. It is obvious that we should not work

in the null space of the reduced within-covariance matrix S'_b . We further project S'_b onto its range space and denote this operation as $W_{proj} \in \mathbb{R}^{d' \times d''}$ ($d'' \leq d'$).

We now introduce our new framework. Firstly, we apply our modified PCA procedure to lower the dimension from d to d' and get a projection matrix $W_{PCA_S} \in \mathbb{R}^{d \times d'}$. Moreover we project onto the range space of the matrix S'_b and get a projection matrix $W_{proj} \in \mathbb{R}^{d' \times d''}$. Finally, we use IFDA to find out the feature representation in the lower dimensionality feature space $\mathbb{R}^{d''}$ and obtain a transformation matrix W_{IFDA} . Consequently, we have the transformation matrix W_{opt} of our new approach as follow

$$W_{opt}^T = W_{IFDA}^T \cdot W_{proj}^T \cdot W_{PCA_S}^T,$$

where W_{PCA_S} is the result of the optimization problem (4) and

$$\begin{aligned} W_{IFDA} &= \arg \min_W \frac{|W^T W_{proj}^T W_{PCA_S}^T S_w W_{PCA_S} W_{proj} W|}{|W^T W_{proj}^T W_{PCA_S}^T S_b W_{PCA_S} W_{proj} W|} \\ &= \arg \min_W \frac{|W^T W_{proj}^T S'_w W_{proj} W|}{|W^T W_{proj}^T S'_b W_{proj} W|} \\ &= \arg \min_W \frac{|W^T S''_w W|}{|W^T S''_b W|} \end{aligned} \quad (5)$$

We call the columns of the transform W_{opt} **the inverse Fisher face (IFFace)** and this new approach as **IFFace** method. Before we go to the end of this part, we make some comments on our new framework.

- Those eigenvectors with respect to the smaller eigenvalues of S_t are taken into account in our modified PCA step.
- Our inverse Fisher criterion can extract discriminant vectors in the null space of S_w rather than just throw them away.

4 Experiment Results

In this section, experiments are designed to evaluate the performance of our new approach: IFFace. Experiment for comparing the performance between Fisher-Face and IFFace is also done.

Two standard databases from the Olivetti Research Laboratory(ORL) and the FERET are selected for evaluation. These databases could be utilized to test moderate variations in pose, illumination and facial expression. The Olivetti set contains 400 images of 40 persons. Each one has 10 images of size 92×112 with variations in pose, illumination and facial expression. For the FERET set we use 432 images of 72 persons. Each person has 6 images whose resolution after cropping is also 92×112 (See Figure 1). Moreover we combine the two to get a new larger set, the ORL FERET, which has 832 images of 112 persons. We implement our IFFace algorithm and test its performance on the above three databases. On ‘Decision Step’, We use the l^2 metric as the distance measure. For



Fig. 1. Example images of two subjects(the first row) and the cropped images(the second row) with the FERET database

the classifier we use the nearest neighbor rule. 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 on each database and average recognition rates are reported.

4.1 Performance of the IFFace Method

We run our algorithm for the ORL database and the FERET database separately. Figure 2 shows the recognition rates from Rank 1 to Rank 10 for different training sample size with ORL in left and FERET in right. From Figure 2, we can see that, when the training sample size is 5, the recognition rates of Rank 5 for both databases are nearly 99%. These results indicate the effectiveness of our new IFFace method in real-world applications.

4.2 Comparison Between IFFace Method and FisherFace Method

As we know, LDA is based on an assumption that all classes are multivariate Gaussian with a common covariance matrix. For ORL database or FERET database, the assumption is reasonable since a great deal of experiments on these

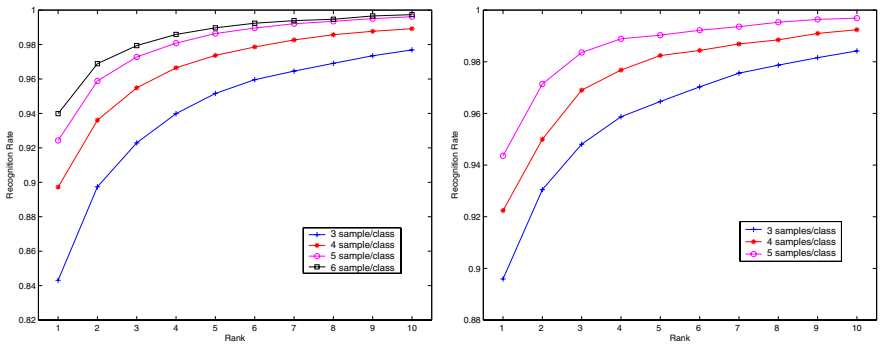


Fig. 2. Recognition rates from Rank 1 to Rank 10 for different training sample per class with ORL database (left) and FERET database (right)

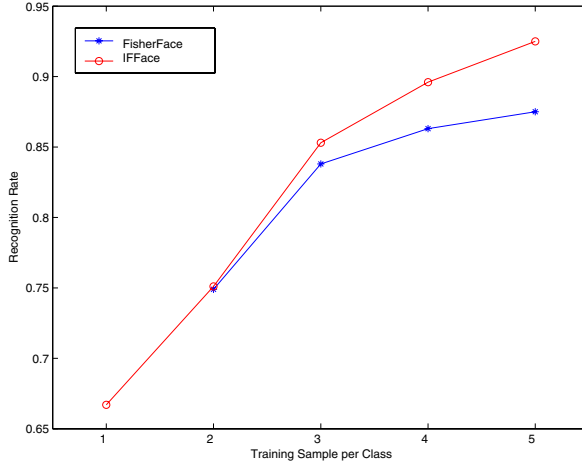


Fig. 3. Comparison between FisherFace and IFFace on the ORL FERET database

two database using FisherFace algorithm have substantiated the efficiency of this two-pharse algorithm. However, when each class has different covariance matrix, this algorithm might not work very well. Therefore, the combination of the two databases would result in a bigger database having different covariance matrix for different classes.

From Figure 3 we can see that IFFace outperforms FisherFace for every number of training sample for each class, take 5, for example, the average recognition rates are 92.5% for IFFace, while for FisherFace it is only 87.6%. This experiment suggests that our IFFace method can work well even in the case that the covariance matrices for different classes are not all the same.

5 Conclusion

In this paper, we proposed a new Fisher discriminant analysis framework: PCA with selection plus IFDA to eliminate deficiencies of the PCA plus LDA method. Based on this framework, we present a new algorithm for face recognition named IFFace method. The algorithm is implemented and experiments are also carried out to evaluate this method. Comparison is made with the PCA plus LDA approach. Further work will be on feature selections and kernel versions.

Acknowledgments

This project is supported in part by grants from NSF of China(Grant No: 60175031, 10231040, 60575004), the Ministry of Education of China, NSF of GuangDong and Sun Yat-Sen University.

References

1. G. Baudat and F. Anouar, Generalized discriminant analysis using a kernel approach, *Neural Computation*, Vol. 12, no. 10(2000), 2385-2404.
2. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, *IEEE Trans. Pattern Analysis and Machine Intelligence.*, Vol. 19(1997), 711-720.
3. L. F. Chen, H. Y. M. Liao, J. C. Lin, M. D. Kao, and G. J. Yu, A new LDA-based face recognition system which can solve the small sample size problem, *Pattern Recognition*, Vol. 33, no. 10(2000), 1713-1726.
4. W. S. Chen, P. C. Yuen, J. Huang and D. Q. Dai, Kernel machine-based one-parameter regularized Fisher discriminant method for face recognition, *IEEE Trans. on Systems, Man and Cybernetics-part B: Cybernetics*, Vol. 35, no. 4(2005), 657-669.
5. D. Q. Dai and P. C. Yuen, Regularized discriminant analysis and its applications to face recognition, *Pattern Recognition*, Vol. 36, no.3(2003), 845-847.
6. D. Q. Dai and P. C. Yuen, A wavelet-based 2-parameter regularization discriminant analysis for face recognition, *Lecture Notes in Computer Science*, Vol. 2688(2003), 137-144.
7. D. Q. Dai and P. C. Yuen, Wavelet based discriminant analysis for face recognition, *Applied Math. and Computation*, 2005(in press), doi: 10.1016/j.amc.2005.07.044
8. A. K. Jain, A. Ross and S. Prabhakar, An introduction to biometric recognition, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, No. 1(2004), 4-20.
9. C. J. Liu and H. Wechsler, A shape- and texture-based enhanced fisher classifier for face recognition, *IEEE Trans. Image Processing*, Vol. 10, no. 4(2001), 598-608.
10. S. Mika, G. Rätsch, J. Weston, B. Schölkopf, A. Smola, and K.-R. Müller, Constructing descriptive and discriminative nonlinear features: rayleigh coefficients in kernel feature spaces, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 25, no. 5(2003), 623-628.
11. I. Pima and M. Aladjem, Regularized discriminant analysis for face recognition, *Pattern Recognition*, Vol. 37(2004), 1945-1948.
12. S. J. Raudys and A. K. Jain, Small sample size effects in statistical pattern recognition: Recommendations for practitioners, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 13(1991), 252-264.
13. D. L. Swets and J. Weng, Using discriminant eigenfeatures for image retrieval, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 18, no. 8(1996), 831-836.
14. J. Yang, A. F. Frangi, J. Y. Yang, D. Zhang, and Z. Jin, KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 27, no. 2(2005), 230-244.
15. J. P. Ye, R. Janardan, C. H. Park, H. Park, An optimization criterion for generalized discriminant analysis on undersampled problems, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26 (8)(2004) 982-994.
16. H. Yu and J. Yang, A direct LDA algorithm for high-dimensional data-with application to face recognition, *Pattern Recognition*, Vol. 34, no. 10(2001), 2067-2070.
17. B. Zhang, H. Zhang, and S. Sam Ge, Face recognition by applying wavelet sub-band representation and kernel associative memory, *IEEE Transactions on Neural Networks*, Vol. 15, No. 1(2004), 166-177.
18. W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld, Face recognition: A literature survey, *ACM Comput. Surv.*, Vol. 35 (4)(2003), 399-459.