Available online at www.sciencedirect.com
SCIENCE DIRECT®
Neurocomputing 69 (2006) 1683–1687
NEUROCOMPUTING
www.elsevier.com/locate/neucom

Letters

# A novel dimensionality-reduction approach for face recognition

Fengxi Song[a,b,*], David Zhang[c], Jingyu Yang[d]

[a] *Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Sili, Shenzhen 518055, PR China*
[b] *New Star Research Institute of Applied Technology, Hefei, PR China*
[c] *Department of Computing, Hong Kong Polytechnic University, Hong Kong, PR China*
[d] *Department of Computer Science, Nanjing University of Science & Technology, Nanjing, PR China*

Received 4 January 2006; received in revised form 24 January 2006; accepted 25 January 2006
Communicated by R.W. Newcomb

**Abstract**

In this paper, we propose a novel dimensionality-reduction method—Fisher discriminant with Schur decomposition (FDS). Similar to Foley–Sammon discriminant analysis (FSD), FDS is an improvement of Fisher discriminant analysis (FDA) in that it eliminates linear dependences among discriminant vectors. In comparison with FSD, FDS is very simple in theory and realization. Experimental results conducted on two benchmark face-image databases, i.e. ORL and AR, demonstrate that FDS is highly effective and efficient in reducing dimensionalities of facial image spaces. Especially when the size of a database is large, FDS can even outperform the state-of-the-art facial feature extraction methods such as the null space method.


*Keywords:* Fisher discriminant analysis; Dimensionality reduction; Orthogonal discriminant vectors; Face recognition

## 1. Introduction

Face recognition is one of the hottest research areas in pattern recognition. In face recognition, the dimensionality of an input space is so high that the input space has to be compressed into a low-dimensional feature space before classification. The reason is that direct recognition on a high-dimensional input space might arouse the so-called overtraining problem and heavy burden of computation. Among numerous dimensionality-reduction techniques, the Fisher discriminant analysis (FDA) [4] is most popular. The essence of FDA is to learn a set of Fisher discriminant vectors by solving an optimization model.

However, Fisher discriminant vectors are usually linearly dependent. Many practices suggest that it will be helpful to eliminate linear dependences among discriminant vectors by making them orthogonal to each other. Based on this fact, Foley and Sammon [5] proposed a set of optimal discriminant vectors, which shared the idea with FDA, but were subjected to orthogonality constraints for binary classification tasks. Duchene and Leclercq [3] extended the concept of optimal discriminant vectors to multiple-class recognition problems. The dimensionality-reduction or feature-extraction method based on optimal discriminant vectors is well known as Foley–Sammon discriminant analysis (FSD).

FSD is very effective for reducing dimensionalities of input spaces in face recognition. But, its calculation procedure for discriminant vectors is extraordinarily time-consuming.

The rest of this paper is organized as follows. In Section 2, the concepts of FDA and FSD are briefly reviewed. In Section 3, the dimensionality-reduction method of FDS is presented. In Section 4, the performance of FDS is experimentally evaluated on two benchmark face-image databases, ORL and AR. At last, a brief conclusion is offered in Section 5.

## 2. FDA and FSD

In theory, the Fisher discriminant matrix whose column vectors are Fisher discriminant vectors is an optimal

*Corresponding author.
*E-mail address:* songfengxi@yahoo.com (F. Song).

0925-2312/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2006.01.016

solution of the following optimization model:

$$\max_{W \in R^{d\times r}} J(W) = \frac{|W^T S_B W|}{|W^T S_W W|}, \tag{1}$$

where $W \in R^{d\times r}$ is an arbitrary matrix, $S_B$ and $S_W$ are between- and within-class scatter matrices defined as in [4], and $|A|$ is the determinant of a square matrix $A$.

It has been proved that if $S_W$ is nonsingular, the matrix composed of unit eigenvectors of the matrix $S_W^{-1}S_B$ corresponding to the first $r$ largest eigenvalues is an optimal solution of the model (1) [9]. The unit eigenvector matrix is the Fisher discriminant matrix commonly used in practice. Thus, discriminant vectors of FDA can be obtained by performing eigenanalysis on the matrix $S_W^{-1}S_B$. Suppose $\mathbf{w}_1,\ldots,\mathbf{w}_r$ to be Fisher discriminant vectors corresponding to eigenvalues $\lambda_1,\ldots,\lambda_r$, respectively. It follows that

$$J([\mathbf{w}_1,\ldots,\mathbf{w}_r]) = \frac{|[\mathbf{w}_1,\ldots,\mathbf{w}_r]^T S_B [\mathbf{w}_1,\ldots,\mathbf{w}_r]|}{|[\mathbf{w}_1,\ldots,\mathbf{w}_r]^T S_W [\mathbf{w}_1,\ldots,\mathbf{w}_r]|} = \prod_{j=1}^{r} \lambda_j = \max_{W \in R^{d\times r}} J(W). \tag{2}$$

Since the matrix $S_W^{-1}S_B$ is usually asymmetric, the Fisher discriminant vectors, i.e. column vectors of the eigenvector matrix of $S_W^{-1}S_B$ are not necessarily orthogonal to each other. In order to eliminate linear dependences in Fisher discriminant vectors, FSD calculates its discriminant vector as follows. The first discriminant vector of FSD is taken as the first Fisher discriminant vector. After the first $k$ discriminant vectors of FSD $\mathbf{w}_1,\ldots,\mathbf{w}_k$ have been calculated, the $(k+1)$th discriminant vector $\mathbf{w}_{k+1}$ is one of the optimal solutions for the following optimization model:

$$\max_{\mathbf{w}_i^T\mathbf{w}=0, i=1,\ldots,k} \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}}. \tag{3}$$

Algorithms for calculation of optimal discriminant vectors are usually complex. The summary of a simple algorithm proposed by Jin et al. [7] is as follows. The first discriminant vector of FSD is taken as a unit eigenvector of the matrix $S_W^{-1}S_B$ corresponding to the largest eigenvalue. After the first $k$ discriminant vectors of FSD $\mathbf{w}_1,\ldots,\mathbf{w}_k$ have been calculated, the $(k+1)$th discriminant vector $\mathbf{w}_{k+1}$ is taken as a unit eigenvector of the matrix $S_W^{-1}PS_B$ corresponding to the largest eigenvalue. The matrix $P$ is calculated by the formula

$$P = I - D^T(DS_W^{-1}D^T)^{-1}DS_W^{-1}, \tag{4}$$

where I is an identity matrix, and $D = [\mathbf{w}_1,\ldots,\mathbf{w}_k]^T$.

## 3. Fisher discriminant with Schur decomposition (FDS)

It is well known that the Schur decomposition is a natural extension of and a good substitution for eigenanalysis when the matrix identified to be analyzed is asymmetric [6]. Let $A$ be a real square matrix, its Schur decomposition is $A = UTU^T$. Here, $U$ is an orthogonal matrix, and $T$ is a quasi-upper-diagonal matrix with the real eigenvalues of the matrix $A$ on the diagonal and the complex eigenvalues in $2\times 2$ blocks on the diagonal.[1]

Instead of performing eigenanalysis on the matrix $S_W^{-1}S_B$ in FDA, Schur decomposition is carried out on $S_W^{-1}S_B$ in FDS. Suppose the Schur decomposition of $S_W^{-1}S_B$ to be $UTU^T$, and $\mathbf{u}_1,\ldots,\mathbf{u}_d$ to be all column vectors of the matrix $U$, i.e. Schur vectors of $S_W^{-1}S_B$. It is obvious that $\mathbf{u}_1,\ldots,\mathbf{u}_d$ are orthogonal to each other. Assume $\mathbf{u}_1,\ldots,\mathbf{u}_r$ to be Schur vectors of $S_W^{-1}S_B$ corresponding to the first $r$ largest real eigenvalues. Thus, we can compress a high-dimensional input space $R^d$ into a low-dimensional feature space $R^r$ by mapping $x \mapsto V^T x$. Here, $V = [\mathbf{u}_1,\ldots,\mathbf{u}_r]$ is the discriminant matrix of FDS, and $\mathbf{u}_1,\ldots,\mathbf{u}_r$ are corresponding discriminant vectors.

Similar to FDA, the discriminant matrix of FDS is also an optimal solution of the optimization model (1). The following theorem reveals the fact:

**Theorem.** *Suppose* $\mathbf{u}_1,\ldots,\mathbf{u}_r$ *to be discriminant vectors of FDS. Thus, we have*

$$J([\mathbf{u}_1,\ldots,\mathbf{u}_r]) = \frac{|[\mathbf{u}_1,\ldots,\mathbf{u}_r]^T S_B [\mathbf{u}_1,\ldots,\mathbf{u}_r]|}{|[\mathbf{u}_1,\ldots,\mathbf{u}_r]^T S_W [\mathbf{u}_1,\ldots,\mathbf{u}_r]|} = \max_{W \in R^{d\times r}} J(W). \tag{5}$$

**Proof.** Since $\mathbf{u}_1,\ldots,\mathbf{u}_r$ are Schur vectors of the matrix $S_W^{-1}S_B$ corresponding to the first $r$ largest real eigenvalues, we have

$$S_W^{-1}S_B\mathbf{u}_j = \lambda_j\mathbf{u}_j \quad j = 1,\ldots,r, \tag{6}$$

where $\lambda_j$ is the $j$th largest real eigenvalue of the matrix $S_W^{-1}S_B$. From the formula (6), it follows that

$$[\mathbf{u}_1,\ldots,\mathbf{u}_r]^T S_B [\mathbf{u}_1,\ldots,\mathbf{u}_r] = [\mathbf{u}_1,\ldots,\mathbf{u}_r]^T S_W [\mathbf{u}_1,\ldots,\mathbf{u}_r]\,\mathrm{diag}(\lambda_1,\ldots,\lambda_r). \tag{7}$$

Thus, we have

$$J([\mathbf{u}_1,\ldots,\mathbf{u}_r]) = \frac{|[\mathbf{u}_1,\ldots,\mathbf{u}_r]^T S_B [\mathbf{u}_1,\ldots,\mathbf{u}_r]|}{|[\mathbf{u}_1,\ldots,\mathbf{u}_r]^T S_W [\mathbf{u}_1,\ldots,\mathbf{u}_r]|} = \prod_{j=1}^{r} \lambda_j = \max_{W \in R^{d\times r}} J(W). \tag{8}$$

## 4. Evaluation of FDS in face recognition

Since face recognition is a typical small-sample-size (S3) problem, like FDA, FSD and FDS cannot be used to compress facial image spaces directly. Similar to Fisherfaces, an intermediate space is introduced first, i.e. a facial image space is first compressed by a PCA transformation. Then, the largest $N-l$ principal components are

[1] See help of the function SCHUR in Matlab.