Pattern Recognition Letters 33 (2012) 1157–1162

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

A two-stage linear discriminant analysis for face-recognition

Alok Sharma [a,b,c,*], Kuldip K. Paliwal [a]

[a] Signal Processing Lab, Griffith University, Australia
[b] School of Engineering & Physics, University of the South Pacific, Fiji
[c] Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, University of Tokyo, Japan

ARTICLE INFO

Article history:
Received 20 September 2011
Available online 10 February 2012
Communicated by K.A. Toh

Keywords:
Two-stage linear discriminant analysis
Small sample size problem
Classification accuracy

ABSTRACT

A two-stage linear discriminant analysis technique is proposed that utilizes both the null space and range space information of scatter matrices. The technique regularizes both the between-class scatter and within-class scatter matrices to extract the discriminant information. The regularization is conducted in parallel to give two orientation matrices. These orientation matrices are concatenated to form the final orientation matrix. The proposed technique is shown to provide better classification performance on face recognition datasets than the other techniques.

1. Introduction

Linear discriminant analysis (LDA) is a well known technique for dimensionality reduction and feature extraction (Duda et al., 2000; Sharma and Paliwal, 2006, 2008, 2010, 2012; Chen et al., 2000; Lu et al., 2003a,b, 2005; Yang et al., 2003; Yu and Yang, 2001; Swets and Weng, 1996; Belhumeur et al., 1997; Ye, 2005; Guo et al., 2007; Thomaz et al., 2005; Huang et al., 2002; Tian et al., 1986; Zhao et al., 2003; Jiang et al., 2008; Gao and Davis, 2006; Paliwal and Sharma, 2010, 2011; Mandal et al., 2010). Dimensionality reduction plays crucial role in the face recognition problem. It is generally applied for improving robustness (or generalization capability) and reducing computational complexity of the face recognition classifier. In the LDA technique, the orientation matrix $\boldsymbol{W}$ is computed from the eigenvalue decomposition (EVD) of $\boldsymbol{S}_W^{-1}\boldsymbol{S}_B$ (Duda et al., 2000), where $\boldsymbol{S}_W \in \Re^{d\times d}$ is within-class scatter matrix, $\boldsymbol{S}_B \in \Re^{d\times d}$ is between-class scatter matrix and d is the dimensionality of feature space. In the face recognition problem, the matrix $\boldsymbol{S}_W$ becomes singular and its inverse computation becomes impossible. Several techniques are reported in the literature that overcome this drawback of LDA (Chen et al., 2000; Lu et al., 2003a,b, 2005; Yang et al., 2003; Yu and Yang, 2001; Swets and Weng, 1996; Belhumeur et al., 1997; Ye, 2005; Guo et al., 2007; Thomaz et al., 2005; Sharma and Paliwal, 2010, 2012; Huang et al., 2002; Tian et al., 1986; Zhao et al., 2003; Jiang et al., 2008; Paliwal and Sharma, 2010, 2011; Mandal et al., 2010).

In LDA, there are four informative spaces namely, null space of $\boldsymbol{S}_W\left(\boldsymbol{S}_W^{null}\right)$, range space of $\boldsymbol{S}_W(\boldsymbol{S}_W^{range})$, null space of $\boldsymbol{S}_B\left(\boldsymbol{S}_B^{null}\right)$ and range space of $\boldsymbol{S}_B$ ($\boldsymbol{S}_B^{range}$). All these four individual spaces have significant discriminant information (refer Appendix I for empirical demonstration). To approximate the inverse computation of $\boldsymbol{S}_W$, different combinations of these spaces are used in the literature for finding the orientation matrix $\boldsymbol{W}$. For an instance the pseudo-inverse technique (Tian et al., 1986) uses $\boldsymbol{S}_W^{range}$ and $\boldsymbol{S}_B^{range}$ to compute the orientation matrix. The regularized LDA technique (Zhao et al., 2003) uses $\boldsymbol{S}_W^{null}$, $\boldsymbol{S}_W^{range}$ and $\boldsymbol{S}_B^{range}$. However, due to the use of small value of regularization parameter (compared to the large eigenvalues of $\boldsymbol{S}_W$), the $\boldsymbol{S}_W^{range}$ gets de-emphasize in the inverse operation of $\boldsymbol{S}_W$. Therefore, the influential spaces in the regularized LDA technique are $\boldsymbol{S}_W^{null}$ and $\boldsymbol{S}_B^{range}$. Similarly, the null LDA technique (Chen et al., 2000) uses $\boldsymbol{S}_W^{null}$ and $\boldsymbol{S}_B^{range}$. These techniques basically utilize two spaces in the orientation matrix computation and discard the other two spaces. Since the individual spaces contribute crucial discriminant information for classification, discarding some spaces would sacrifice the classification performance of the classifier. Theoretically, if all the four spaces can be inherited appropriately in the computation of orientation matrix $\boldsymbol{W}$ then the classification performance can be improved further.

In this paper, we exploit ways of utilizing all the four spaces. The inclusion of all the spaces of scatter matrices is done in two analyses. Fig. 1 illustrates the proposed strategy. The orientation matrix can be computed from the input data by carrying out two discriminant analyses in parallel. In the first analysis, the orientation matrix $\boldsymbol{W}_1$ is computed by retaining top eigenvalues and eigenvectors of $\boldsymbol{S}_W'^{-1}S_B$, where non-singular matrix $\boldsymbol{S}'$ is the approximation of singular matrix $\boldsymbol{S}$. This will retain $\boldsymbol{S}_W^{null}$ and $\boldsymbol{S}_B^{range}$. In the second analysis, the orientation matrix $\boldsymbol{W}_2$ is obtained by retaining top eigenvalues and eigenvectors of $\boldsymbol{S}_B'^{-1}\boldsymbol{S}_W$. This will retain $\boldsymbol{S}_W^{range}$ and $\boldsymbol{S}_B^{null}$. The orientation matrices obtained by these two analyses are

* Corresponding author at: Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, University of Tokyo, Japan.
E-mail addresses: aloks@ims.u-tokyo.ac.jp, sharma_al@usp.ac.fj (A. Sharma).

0167-8655/$ - see front matter
doi:10.1016/j.patrec.2012.02.001

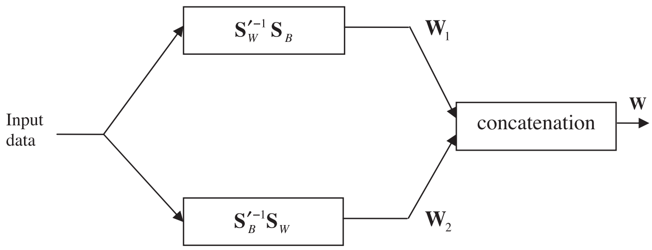

Fig. 1. The proposed strategy.

concatenated to get the final orientation matrix $\boldsymbol{W}$, i.e., $\boldsymbol{W} = [\boldsymbol{W}_1, \boldsymbol{W}_2]$. For brevity we call the proposed technique the two-stage LDA technique. The non-singular approximation $\boldsymbol{S}'$ of singular matrix $\boldsymbol{S}$ can be evaluated in two ways: (1) using regularized LDA technique (Zhao et al., 2003) where $\boldsymbol{S}' = \boldsymbol{S} + \alpha\boldsymbol{I}$ (α is the regularization parameter); and, (2) using extrapolation technique (Jiang et al., 2008; Sharma and Paliwal, 2010) where eigenvalues of $\boldsymbol{S}$ are extrapolated by applying curve fitting or some criterion function. In this paper we show that the resulting orientation matrix $\boldsymbol{W}$ provides better classification results than other existing techniques.

2. Notations and descriptions

Let us denote the n linearly independent training samples (or feature vectors) in d-dimensional space by $\mathcal{X} = \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, having class labels $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, where $\omega_i \in \{1, 2, \ldots, c\}$ and c is the number of classes. The set $\mathcal{X}$ can be subdivided into c subsets $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_c$ where each subset $\mathcal{X}_j$ belongs to a particular class label and consists of n_j number of samples such that:

$$n = \sum_{j=1}^{c} n_j$$

and $\mathcal{X}_j \subset \mathcal{X}$ and $\mathcal{X}_1 \cup \mathcal{X}_2 \cup \cdots \cup \mathcal{X}_c = \mathcal{X}$.

Let $\boldsymbol{\mu}_j$ be the centroid of $\mathcal{X}_j$ and $\boldsymbol{\mu}$ be the centroid of $\mathcal{X}$, then the between class scatter matrix $\boldsymbol{S}_B$, within-class scatter matrix $\boldsymbol{S}_W$ and total-scatter matrix $\boldsymbol{S}_T$ are defined as (Duda et al., 2000)

$$\boldsymbol{S}_B = \sum_{j=1}^{c} \boldsymbol{n}_j (\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^{\mathrm{T}} \quad (1)$$

$$\boldsymbol{S}_W = \sum_{j=1}^{c} \boldsymbol{S}_j \quad (2)$$

where

$$\boldsymbol{S}_j = \sum_{\boldsymbol{x} \in \mathcal{X}_j} (\boldsymbol{x} - \boldsymbol{\mu}_j)(\boldsymbol{x} - \boldsymbol{\mu}_j)^{\mathrm{T}}$$

and

$$\boldsymbol{S}_T = \sum_{\boldsymbol{x} \in \mathcal{X}} (\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} \quad (3)$$

Since in the face recognition task $d > n$, the scatter matrices $\boldsymbol{S}_B$, $\boldsymbol{S}_W$ and $\boldsymbol{S}_T$ will be singular with ranks $r_b = c - 1$, $r_w = n - c$ and $r_t = n - 1$, respectively. The null space of $\boldsymbol{S}_T$ carries no discriminative information (Huang et al., 2002), therefore, the dimensionality can be reduced from d-dimensional space to $r_t = n - 1$ dimensional space by applying principal component analysis (PCA) as a pre-processing step to remove the null space of $\boldsymbol{S}_T$. This would make the technique computationally faster. The range space of total scatter matrix $\boldsymbol{U}_{TR} \in \Re^{d \times r_t}$ will be used as a transformation. This will give us transformed within-class scatter matrix $\widehat{\boldsymbol{S}}_W \in \Re^{r_t \times r_t}$ and transformed between-class scatter matrix $\widehat{\boldsymbol{S}}_B \in \Re^{r_t \times r_t}$. These matrices can be decomposed as

$$\widehat{\boldsymbol{S}}_W = \boldsymbol{U}_W \boldsymbol{D}_W^2 \boldsymbol{U}_W^{\mathrm{T}} \quad (4)$$

and

$$\widehat{\boldsymbol{S}}_B = \boldsymbol{U}_B \boldsymbol{D}_B^2 \boldsymbol{U}_B^{\mathrm{T}} \quad (5)$$

where $\boldsymbol{D}_W \in \Re^{r_t \times r_t}$ and $\boldsymbol{D}_B \in \Re^{r_t \times r_t}$ are diagonal matrices whose elements (arranged in descending order) are the square-root of the eigenvalues of $\widehat{\boldsymbol{S}}_W$ and $\widehat{\boldsymbol{S}}_B$, respectively; and $\boldsymbol{U}_W \in \Re^{r_t \times r_t}$ and $\boldsymbol{U}_B \in \Re^{r_t \times r_t}$ are orthogonal matrices consisting of the corresponding eigenvectors as columns. Since the rank of $\widehat{\boldsymbol{S}}_W$ is r_w, the matrix $\boldsymbol{U}_W$ can be formed as $\boldsymbol{U}_W = [\boldsymbol{U}_{WR}, \boldsymbol{U}_{WN}]$ where $\boldsymbol{U}_{WR} \in \Re^{r_t \times r_w}$ corresponds to the range space of $\widehat{\boldsymbol{S}}_W$ and $\boldsymbol{U}_{WN} \in \Re^{r_t \times (r_t - r_w)}$ corresponds to the null space of $\widehat{\boldsymbol{S}}_W$. In a similar way, we can write $\boldsymbol{U}_B = [\boldsymbol{U}_{BR}, \boldsymbol{U}_{BN}]$ where $\boldsymbol{U}_{BR} \in \Re^{r_t \times r_b}$ corresponds to the range space of $\widehat{\boldsymbol{S}}_B$ and $\boldsymbol{U}_{BN} \in \Re^{r_t \times (r_t - r_b)}$ corresponds to the null space of $\widehat{\boldsymbol{S}}_B$.

3. Two-stage LDA technique

It is well known in the literature that the null space of $\widehat{S}_W$ contains crucial information for classification (Chen et al., 2000; Ye, 2005). The null space based LDA techniques retain the null space information of $\widehat{S}_W$, however, they discard the range space information of $\widehat{\boldsymbol{S}}_W$. It has been seen that the range space information of $\widehat{\boldsymbol{S}}_W$ is also important for classification (Swets and Weng, 1996; Belhumeur et al., 1997) and by discarding it could penalize classification performance. Some techniques (e.g. Guo et al., 2007; Zhao et al., 2003; Jiang et al., 2008; Sharma and Paliwal, 2010) estimates non-singular within-class scatter matrix $\widehat{\boldsymbol{S}}'_W$ by adding a small positive constant (known as regularization parameter) to the eigenvalues of $\widehat{\boldsymbol{S}}_W$ (Guo et al., 2007; Zhao et al., 2003) or by extrapolating the eigenvalues of $\widehat{\boldsymbol{S}}_W$ in its null space (Jiang et al., 2008; Sharma and Paliwal, 2010). Thereafter, obtaining the eigenvectors corresponding to the top eigenvalues of $\widehat{\boldsymbol{S}}'^{-1}_W \widehat{\boldsymbol{S}}_B$. In these techniques the null space information of $\widehat{\boldsymbol{S}}_W$ and the range space information of $\widehat{\boldsymbol{S}}_B$ are effectively retained. Although, the range space information of $\widehat{\boldsymbol{S}}_W$ is utilized in these techniques, it has very less influence as it is de-emphasized in the inverse operation of $\widehat{\boldsymbol{S}}_W$ (see Fig. 2). Nonetheless, theoretically the latter implementation would contain more information than the former techniques. To see the qualitative contribution of $\widehat{\boldsymbol{S}}'^{-1}_W \widehat{\boldsymbol{S}}_B$ in obtaining the orientation matrix, we decompose $\widehat{\boldsymbol{S}}'_W$ into its eigenvalues and eigenvectors as

$$\widehat{\boldsymbol{S}}'_W = \boldsymbol{U}_W \widehat{\boldsymbol{D}}_W^2 \boldsymbol{U}_W^{\mathrm{T}} \quad (6)$$

where diagonal matrix $\widehat{\boldsymbol{D}}_W = \begin{bmatrix} \Sigma_W & 0 \\ 0 & \widehat{\Sigma}_W \end{bmatrix}$, $\Sigma_W \in \Re^{r_w \times r_w}$ and $\widehat{\Sigma}_W \in \Re^{(r_t - r_w) \times (r_t - r_w)}$ is the estimation or regularization of eigenvalues Σ_W.

From Eq. (5), $\widehat{\boldsymbol{S}}_B$ can be formed as

$$\widehat{\boldsymbol{S}}_B = [\boldsymbol{U}_{BR}, \boldsymbol{U}_{BN}] \begin{bmatrix} \Sigma_B^2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_{BR}^{\mathrm{T}} \\ \boldsymbol{U}_{BN}^{\mathrm{T}} \end{bmatrix} = \boldsymbol{U}_{BR} \Sigma_B^2 \boldsymbol{U}_{BR}^{\mathrm{T}} \quad (7)$$

where $\Sigma_B \in \Re^{r_b \times r_b}$.

From Eqs. (6) and (7), we can write

$$\widehat{\boldsymbol{S}}'^{-1}_W \widehat{\boldsymbol{S}}_B = \boldsymbol{U}_W \widehat{\boldsymbol{D}}_W^{-2} \boldsymbol{U}_W^{\mathrm{T}} \boldsymbol{U}_{BR} \Sigma_B^2 \boldsymbol{U}_{BR}^{\mathrm{T}} \quad (8)$$

The EVD of Eq. (8) can be computed and the range space information of $\widehat{\boldsymbol{S}}'^{-1}_W \widehat{\boldsymbol{S}}_B$ can be used in the formation of orientation matrix. Three things can be observed here:

(1) The null space of $\widehat{\boldsymbol{S}}_B$ is discarded.
(2) The range space information of within-class scatter matrix in the inverse operation is de-emphasized.
(3) The null space of the product $\widehat{\boldsymbol{S}}'^{-1}_W \widehat{\boldsymbol{S}}_B$ is discarded.