Neurocomputing 72 (2009) 3384–3390

Contents lists available at ScienceDirect

Neurocomputing

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

Letters

Feature extraction using fuzzy inverse FDA

Wankou Yang [a,c,*], Jianguo Wang [a,b], Mingwu Ren [a], Lei Zhang [c], Jingyu Yang [a]

[a] School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China
[b] Tangshan College, Network and Education Center, Tangshan 063000, China
[c] Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Article history:
Received 17 February 2008
Received in revised form
8 March 2009
Accepted 8 March 2009
Communicated by T. Heskes
Available online 9 April 2009

Keywords:
FDA
Inverse FDA
Fuzzy IFDA
Feature extraction
Face recognition
Pulse signal recognition

ABSTRACT

This paper proposes a new method of feature extraction and recognition, namely, the fuzzy inverse Fisher discriminant analysis (FIFDA) based on the inverse Fisher discriminant criterion and fuzzy set theory. In the proposed method, a membership degree matrix is calculated using FKNN, then the membership degree is incorporated into the definition of the between-class scatter matrix and within-class scatter matrix to get the fuzzy between-class scatter matrix and fuzzy within-class scatter matrix. Experimental results on the ORL, FERET face databases and pulse signal database show that the new method outperforms Fisherface, fuzzy Fisherface and inverse Fisher discriminant analysis.

1. Introduction

It is well known that FDA is an effective feature extraction method. But, unfortunately, it cannot be applied directly to small size sample (SSS) problems [1], for in these cases, the within-class scatter matrix is singular. As we know, face recognition is a typical small size problem. In order to utilize LDA for face recognition, a number of research works have been done [2–9]. The most popular method, called Fisherface, was build by Swets et al. [2] and Belhumeur et al. [3]. In their methods, PCA is first used to reduce the dimension of the original features space to $N-c$, and the classical FLD is next applied to reduce the dimension to d ($d \leqslant c$). Obviously, in the PCA transform, the small $c-1$ projection components have been thrown away. So some effective discriminatory information may be lost. And PCA step cannot guarantee the transformed within-class scatter matrix still be not singular. Yang and Yang [4] have proven the theoretical foundation for Fisherface and proposed a linear feature extraction method. First, PCA is used to reduce the dimension of feature space to $N-1$ (N denotes the number of training samples), and then the OFLD [5] method is used for the second feature extraction. However, no procedure has been shown to determine the optimal dimensions in the OFLD.

To handle the singularity problem, it is also popular to add a singular value perturbation to within-class scatter matrix to make it nonsingular [6]. A similar but more systematic method is regularized discriminant analysis (RDA) [7]. In RDA, one tries to obtain more reliable estimates of the eigenvalues by correcting the eigenvalue distortion in the sample covariance matrix with a ridge-type regularization. Besides, RDA is also a compromise between LDA and QDA (quadratic discriminant analysis), which allows one to shrink the separate covariances of QDA towards a common covariance as in LDA. Penalized discriminant analysis (PDA) is another regularized version of LDA [8,9]. The goals of PDA are not only to overcome the small sample size problem but also to smooth the coefficients of discriminant vectors for better interpretation. The main problem with RDA and PDA is that they do not scale well. In application such as face recognition, the dimension of covariance matrices is often more than 10 000. It is not practical for RDA and PDA to process such large covariance matrices, especially, when the computing platform is made of PCs. A well-known null subspace method is the LDA+PCA method [10]. When within-class scatter matrix is of full rank, the LDA+PCA method just calculates the maximum eigenvectors of $(S_w)^{-1}S_b$ to form the transformation matrix. Otherwise, a two-stage procedure is employed. First, the data are transformed into null space V_0 of S_w. Second, it tries to maximize the between-class scatter in V_0. Although this method solves small sample size problem, it could be sub-optimal because it maximizes the between-class

* Corresponding author at: School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China.
Tel.: (+)86 (0) 25 84317297x403.
E-mail addresses: wankou_yang@yahoo.com.cn, yangwankou@njust.edu.cn (W. Yang).

0925-2312/$ - see front matter Crown Copyright © 2009 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2009.03.011

scatter in the null space of S_w instead of the original input space. Direct LDA is another null space method that discards the null space of S_b [11]. This is achieved by diagonalizing first S_b and then diagonalizing S_w, which is in the reverse order of conventional simultaneous diagonalization procedure. In Direct LDA, one may also employ S_t instead of S_w. In this way, Direct LDA is actually equivalent to the PCA+LDA. Li et al. [12] further proposed an efficient and robust linear feature extraction method which aims to maximize the following criterion $J = \mathrm{tr}(W^T(S_b - S_w)W)$, which was called maximum margin criterion (MMC). In order to avoid both the singularity and instability critical issues of the within-class scatter matrix S_w when LDA is used in limited sample and high dimensional problems, Thomaz and Gillies [13] proposed a LDA-based approach based on a straightforward covariance selection method for the S_w matrix [14].

Recently, Dai et al. [15,16] develop an inverse Fisher discriminant analysis (IFDA) method. The algorithm modifies the procedure of PCA and derives the regular and irregular information from the within-class scatter matrix by a new criterion, which is called inverse Fisher discriminant criterion. Unfortunately, the IFDA assumes that the same level of typicality of each face to the corresponding class. Inspired by fuzzy Fisherface [19], we propose to incorporate a gradual level of assignment to class being regarded as a membership grade with anticipation that such discrimination helps improve classification results. More specifically, when operating on feature vectors resulting from PCA transformation we complete a fuzzy K-nearest neighbor class assignment that produces the corresponding degree of class membership. By taking advantage of the technology of fuzzy sets [17], a number of studies have been carried out for fuzzy pattern recognition [18–21]. This paper is the improved version of [22].

The organization of this paper is as follows. In Section 2, we review briefly the related works. In Section 3, we propose the idea and describe the new method in detail. In Section 4, experiments with face images data are presented to demonstrate the effectiveness of the new method. Conclusions are summarized in Section 5.

2. Related works

2.1. FDA

Linear (Fisher) discriminant analysis (FDA) is a well-known and popular statistical method in pattern recognition and classification [23]. Suppose there are c known pattern classes $w_1, w_2, \ldots, w_c$, N training samples. $X = \{x_j^i\}$ $(i = 1, 2, \ldots, l_c,\ \ j = 1, 2, \ldots, c)$ is a set of samples with d dimension. l_j is the number of training samples of class j and satisfies $\sum_{i=1}^{c} l_i = N$. Set $\overline{X_j} = (1/l_j)\sum_{i=1}^{l_j} x_j^i$, $\bar{X} = (1/N)\sum_{j=1}^{c}\sum_{i=1}^{l_j} x_j^i$. Let the between-class scatter matrix and the within-class scatter matrix be defined as

$$S_b = \frac{1}{N}\sum_{j=1}^{c} l_j(\overline{X_j} - \overline{X})(\overline{X_j} - \overline{X})^T \tag{1}$$

$$S_w = \frac{1}{N}\sum_{j=1}^{c}\sum_{i=1}^{l_j}(x_j^i - \overline{x_j})(x_j^i - \overline{x_j})^T \tag{2}$$

It is easy to verify that $S_t = S_w + S_b$. Now, the projection matrix W_{FDA} of FDA is chosen as a matrix with orthonormal columns maximizing the following quotient, called Fisher discriminant criterion:

$$W_{FDA} = \arg\max_W \frac{|W^T S_b W|}{|W^T S_w W|} = [w_1, w_2, \ldots, w_m] \tag{3}$$

where $\{w_i | i = 1, 2, \ldots, m\}$ is the set of generalized eigenvectors of $(S_w)^{-1}S_b$ corresponding to the m largest generalized eigenvalues $\{\lambda_i | i = 1, 2, \ldots, m\}$.

2.2. Fuzzy Fisherface

Kwak [19] proposed the fuzzy Fisherface for face recognition via fuzzy set. Given a set of feature vectors, $X = \{x_1, x_2, \ldots, x_{N_i}\}$, a fuzzy "c"-class partition of these vectors specifies the degree of membership of each vector to the classes. The membership matrix $[u_{ij}](i = 1, 2, \ldots, c, j = 1, 2, \ldots, N)$ can be obtained by FKNN [18]; it will be discussed in Section 3.1 in detail. Taking into account the membership grades, the mean vector of each class $\overline{m_i}$ is calculated as follows:

$$\overline{m_i} = \frac{\sum_{j=1}^{N} u_{ij} x_j}{\sum_{j=1}^{N} u_{ij}} \tag{4}$$

The between-class fuzzy scatter matrix SF_b and within-class fuzzy scatter matrix SF_w incorporate the membership values in their calculations

$$SF_b = \sum_{i=1}^{c} N_i(\overline{m_i} - \overline{X})(\overline{m_i} - \overline{X})^T \tag{5}$$

$$SF_w = \sum_{i=1}^{c}\sum_{x_k \in w_i}(x_k - \overline{m_i})(x_k - \overline{m_i})^T \tag{6}$$

The optimal fuzzy projection $W_{F\text{–}FDA}$ follows the expression:

$$W_{F\text{–}FDA} = \arg\max_W \frac{|W^T SF_b W|}{|W^T SF_w W|} \tag{7}$$

Finally, Kwak gave the strategy: PCA plus fuzzy FDA in small sample size case.

2.3. Inverse FDA

From the inverse relationship

$$\arg\max_W \frac{|W^T S_b W|}{|W^T S_w W|} \Leftrightarrow \arg\min_W \frac{|W^T S_w W|}{|W^T S_b W|} \tag{8}$$

Without considering the singularity, the inverse Fisher discriminant criterion, is

$$W_{IFDA} = \arg\min_W \frac{|W^T S_w W|}{|W^T S_b W|} \tag{9}$$

In contrast with LDA, the program using the above inverse Fisher discriminant criterion is called inverse Fisher discriminant analysis [15,16]. Obviously, the Fisher criterion (3) and inverse Fisher criterion (9) are equivalent, provided that the within-class scatter matrix S_w and the between-class scatter S_b are not singular. However, $rank(S_b) \leqslant c - 1$. Thus, the difficulty of SSS problem still exists for IFDA. So the PCA can be exploited to reduce the dimension of the original feature space. First, we apply PCA procedure to lower the dimension from d to d' and get a projection matrix $W_{PCA\_S} \in R^{d \times d'}$ $(W_{PCA\_S} = [u_{i1}, u_{i2}, \ldots, u_{id'}],$ s.t. $u_{ij}^T S_t u_{ij} > 0$, $u_{ij}^T S_b u_{ij} > u_{ij}^T S_w u_{ij})$. Moreover, we project onto the range space of the matrix S_b' and get a projection matrix $W_{proj} \in R^{d' \times d''}$. Second, we use IFDA to find out the feature representation in the lower dimensionality feature space $R^{d''}$ and obtain a transformation matrix W_{IFDA}. Consequently, we have the transformation matrix W_{opt} of IFDA approach as follows:

$$W_{opt} = W_{IFDA}^T \cdot W_{proj}^T \cdot W_{PCA\_S}^T \tag{10}$$