Neurocomputing 134 (2014) 46–52

Contents lists available at ScienceDirect

Neurocomputing

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

Kernel inverse Fisher discriminant analysis for face recognition

Zhongxi Sun [a,b], Jun Li [a], Changyin Sun [a,*]

[a] School of Automation, Southeast University, Nanjing 210096, China
[b] College of Science, Hohai University, Nanjing 210098, China

ARTICLE INFO

Article history:
Received 9 June 2012
Received in revised form
7 November 2012
Accepted 31 December 2012
Available online 22 January 2014

Keywords:
Kernel IFDA
KPCA
Feature extraction
Face recognition

ABSTRACT

In this paper, we present a new nonlinear feature extraction method for face recognition. The proposed method incorporates the kernel trick with inverse Fisher discriminant analysis and develops a two-phase kernel inverse Fisher discriminant analysis criterion – KPCA plus IFDA. In the proposed method, we first apply the nonlinear kernel trick to map the original face samples into an implicit feature space and then perform inverse Fisher discriminant analysis in the feature space to produce nonlinear discriminating features. In implementation, kernel IFDA seeks nonlinear discriminating features by minimizing the inverse Fisher discriminant quotient and overcome the singularity problem by projective transformation of scatter matrices. Experimental results on ORL, FERET and AR face databases demonstrate the effectiveness of the proposed method.

1. Introduction

Face recognition is one of the most active research topics in pattern recognition and feature extraction plays an important role in face recognition. Fisher discriminant analysis (FDA), also known as linear discriminant analysis (LDA), is one of the classical feature extraction methods in pattern recognition In order to utilize FDA for face recognition, many new extension techniques have been developed [2–18]. The objective of FDA is to seek a linearly optimal projection by maximizing the ratio of the determinants of the between-class and within-class scatter matrices of the projected samples. The optimal projection can be obtained by applying the eigen-decomposition on the scatter matrices. Although successful in many cases, there still exist two key problems in using FDA for face recognition. One is small size sample problem [1] derived from the singularity of the within-class scatter matrix. As we know, the objective function of FDA requires the nonsingularity of one of the scatter matrices. But in many practical face recognition tasks, the dimensionality of face image is usually very high and generally exceeds the number of available training samples and, hence, the within-class scatter matrix is often singular, which gives rise to a problem of unstable numerical computation.

In recently years, many extensions of FDA have been developed to alleviate this problem. A popular method called Fisherface, was put forth by Swets et al. [2] and Belhumeur et al. [3]. In their method, an intermediate dimensionality reduction stage using PCA is first applied for making the within-class scatter matrix nonsingular before the application of the classical LDA. However, applying PCA for dimensionality reduction may result in the loss of some important discriminative information [4–7]. Artificially adding some small perturbation to the within-class scatter matrix to make it full rank [8] was also a solution to this problem. To handle the singularity problem, regularized discriminant analysis (RDA) [9] adopts a similarly ridge-type regularization to correct the eigenvalue distortion of sample covariance matrix for obtaining more reliable estimates of the eigenvalues. Penalized discriminant analysis (PDA) [10,11] is another regularized version of LDA, which tries to smooth the coefficients of discriminant vectors for better interpretation. But, both RDA and PDA do not scale well. Direct LDA (DLDA) [12,13] focuses on between-class scatter matrix and extracts the eigenvectors corresponding to the smallest eigenvalues of the within-class scatter matrix. As the smallest eigenvalues are very sensitive to noise, it is difficult for DLDA to scale the extracted features. The FDA method can perform well in the subspace. However, a common problem of previous FDA extensions is that they may lose some important discriminant information [14,15] and the preprocessing step to reduce the dimensionality using the PCA cannot guarantee the nonsingularity of the transformed within-class scatter matrix. In view of this, Dai et al. [16,17] adopted the ratio of the determinants of both the between-class scatter and the within-class scatter as a discriminant criterion and developed a new method called the inverse Fisher discriminant analysis (IFDA). The algorithm of their new method modifies the procedure of PCA and derives the regular and irregular information from the within-class scatter matrix. Based on this new criterion and fuzzy set theory, Yang et al. [18] proposed a fuzzy version of IFDA for feature extraction and recognition.

Moreover, the classical FDA and its extensions, as linear methods, cannot well deliver good performance when face

* Corresponding author.
E-mail address: cysun@seu.edu.cn (C. Sun).

0925-2312/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.12.075

patterns are subject to large changes under pose, illumination, and facial expression variations, which results in a highly nonconvex and complex distribution. Thus, it is reasonable to assume a better solution to this inherent nonlinear problem. Since the nonlinear kernel-based learning machines have achieved a great success in support vector machine (SVM) [19], recently kernel based nonlinear analysis [20–24,26] has attracted great interest in the areas of pattern recognition and machine learning. To tackle with the nonlinear problem, Scholkopf et al. [20] pioneered to combine the kernel trick with PCA and developed the earliest kernel PCA (KPCA) method. However, as with PCA, KPCA only captures the overall variance of all patterns, which are inadequate for discriminating purposes. Yang et al. [21] developed a framework of KPCA plus unsupervised discriminant projection. In their method, they first used nonlinear kernel mapping to map the data into an implicit feature space and then sought a linear transformation that could maximize the nonlocal scatter to the local scatter in the feature space. As FDA handles the within and between-class variations separately, it is better for discrimination. Along the same outline of KPCA for PCA, Mika et al. [22] combined the kernel trick with FDA for two-class cases and Baudat et al. [23] then developed a general kernel-based Fisher discriminant analysis (KFDA) for multiple-class problems. However, due to the small sample size problem, the within-class scatter in the KFDA feature space still suffers from the singularity problem. Baudat et al. [23] exploited the QR decomposition algorithm to handle this problem of numerical computation. Yang et al. [24] adopted the PCA plus LDA framework applied in the Fisherface method to avoid the problem. While Chen et al. [25] and Yang et al. [14] showed that the null space of the within-class scatter matrix contained valuable discriminatory information. For kernel-based methods, this irregular discriminant information discarded from the within-class scatter matrix [17,26] may be important for the small sample size problem.

Motivated by the success of kernel methods in pattern regression and classification tasks, we propose a new kernel inverse Fisher discriminant analysis method (kernel IFDA) to overcome both the matrix singularity problem and the nonlinear problem for face recognition. We first nonlinearly map the original input space into an implicit high-dimensional feature space, where the "kernel trick" is employed to construct the implicit feature space. Then we perform inverse FDA in this feature space to find optimal linear projection. Meanwhile, we will utilize KPCA to reduce the dimension of feature space and derive the regular and irregular information from the within-class scatter matrix.

The remainder of this paper is organized as follows. Section 2 briefly gives an overview of related works about FDA, inverse FDA and KPCA. Section 3 describes the idea of the proposed method and gives the algorithm in detail. Experimental results comparing with other methods are given in Section 4. Conclusions are presented in Section 5.

2. Related works

2.1. FDA and IFDA

FDA [3] is a classical linear subspace method and is widely used for feature extraction. Suppose $w_1, w_2, \ldots, w_c$ are c known pattern classes and $\{x_i^j\}$ $(i = 1, 2, \ldots, l_c, \quad j = 1, 2, \ldots, c)$ are the training samples, where l_j denotes the number of jth class training samples and satisfies $\sum_{j=1}^{c} l_j = N$. The definitions of the between-class scatter matrix and the within-class scatter matrix are

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

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

where $\overline{X}_j = (1/N)\sum_{i=1}^{l_j} x_i^j$ and $\overline{X} = (1/N)\sum_{j=1}^{c}\sum_{i=1}^{l_j} x_i^j$, respectively, represent the mean vector of jth class and the global mean vector. Obviously

$$S_t = S_b + S_w \tag{3}$$

The projection matrix W_{FDA} of FDA with orthonormal columns can be given by maximizing the following Fisher discriminant quotient:

$$W_{FDA} = \arg\max_W \frac{W^T S_b W}{W^T S_w W} = [w_1, w_2, \cdots, w_d] \tag{4}$$

where $w_i, i = 1, 2, \ldots, d$ are the generalized eigenvectors corresponding to the d generalized maximum eigenvalues $\lambda_i, i = 1, 2, \ldots, d$ of the eigenstructure.

$$S_b w_i = \lambda_i S_w w_i, \quad i = 1, 2, \ldots, d \tag{5}$$

FDA seeks the projection matrix W_{FDA} of m directions such that the between-class scatter is maximized and the within-class scatter minimized simultaneously when the training samples $\{x_i^j\}$ are projected onto this set of directions. In FDA, if the within-class scatter matrix S_w and the between-class scatter matrix S_b are nonsingular, we have the equivalent relation

$$\arg\max_W \frac{W^T S_b W}{W^T S_w W} \Leftrightarrow \arg\min_W \frac{W^T S_w W}{W^T S_b W} \tag{6}$$

Similar to the Fisher discriminant analysis, the inverse Fisher discriminant analysis (IFDA) criterion [17] can be defined as follows:

$$W_{IFDA} = \arg\min_W \frac{W^T S_w W}{W^T S_b W} \tag{7}$$

2.2. KPCA

Kernel PCA (KPCA) [20] is a nonlinear extension of PCA with the kernel trick. The idea of KPCA is to map the input space into a higher dimensional feature space by a nonlinear transformation and then perform PCA in this implicit feature space for acquiring more intrinsic information of the original input data.

Suppose that $\varphi : x \in R^n \rightarrow \varphi(x) \in F$ is a nonlinear mapping that maps the original input space R^n into a Hilbert space F. Let $x_1, x_2, \ldots, x_M \in R^n$ and $\varphi(x_1), \varphi(x_2), \ldots, \varphi(x_M) \in F$ represent M training samples and mapped samples, respectively, we can obtain the covariance matrix in the implicit feature space F as follows:

$$S_t^{\varphi} = \frac{1}{M}\sum_{j=1}^{M}(\varphi(x_j) - \overline{m}_{\varphi})(\varphi(x_j) - \overline{m}_{\varphi})^T \tag{8}$$

where $\overline{m}_{\varphi} = 1/M\sum_{j=1}^{M}\varphi(x_j)$ denotes the mean of all the mapped samples.

The eigenvalues $\lambda \geq 0$ and corresponding eigenvectors β can be obtained by solving the eigensystem $\lambda\beta = S_t^{\varphi}\beta$. According to the reproducing kernel theory [19] and the method proposed in [20], any β lying in the feature space can be spanned by $\varphi(x_1)$, $\varphi(x_2), \ldots, \varphi(x_M)$. That is, there exist expansion coefficients $\alpha_i (i = 1, 2, \ldots, M)$ such that

$$\beta = \sum_{i=1}^{M}\alpha_i\varphi(x_i) \tag{9}$$

Let

$$\tilde{K}_{ij} = \varphi(x_i)^T\varphi(x_j) = (\varphi(x_i)\varphi(x_j)) = k(x_i, y_j) \tag{10}$$

Obviously, $\tilde{K} = (\tilde{K}_{ij})$ is a $M \times M$ matrix composed of dot product in the feature space$F$ and it can be centralized by the following