Neurocomputing 73 (2009) 197–203

Contents lists available at ScienceDirect

Neurocomputing

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

Two-dimensional local graph embedding discriminant analysis (2DLGEDA) with its application to face and palm biometrics

Minghua Wan *, Zhihui Lai, Jie Shao, Zhong Jin

School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, PR China

ARTICLE INFO

Available online 26 August 2009

Keywords:
Feature extraction
Graph embedding
Linear discriminant analysis (LDA)
Small sample size (SSS)

ABSTRACT

This paper proposes a novel method, called two-dimensional local graph embedding discriminant analysis (2DLGEDA), for image feature extraction, which can directly extract the optimal projective vectors from two-dimensional image matrices rather than image vectors based on the scatter difference criterion. In graph embedding, the intrinsic graph characterizes the intraclass compactness and connects each data point with its neighboring within the same class, while the penalty graph connects the marginal points and characterizes the interclass separability. The proposed method effectively avoids the singularity problem frequently encountered in the traditional linear discriminant analysis algorithm (LDA) due to the small sample size (SSS) and overcomes the limitations of LDA due to data distribution assumptions and available projection directions. Experimental results on ORL, YALE, FERET face databases and PolyU palmprint database show the effectiveness of the proposed method.

1. Introduction

Biological recognition has recently attracted wide attention of the researchers in biometric authentication. In face recognition, two-dimensional face images are usually transformed into one-dimensional vectors through column by column or row by row concatenation. The resulting image vectors of faces usually lead to a high dimensional image vector space, where it is difficult to evaluate the covariance matrix accurately due to its large size and the relatively small number of training samples.

Two of the most fundamental dimensionality reduction methods are principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2]. PCA is a classical feature extraction and data representation technique widely used in the areas of pattern recognition and computer vision. PCA aims to find a linear mapping, which preserves the total variance by maximizing the trace of feature variance. The optimal projections of PCA is correspond to the first *k*-largest eigenvalues of the data's total variance matrix. Thus, PCA preserves the total variance by maximizing the trace of feature variance, but PCA cannot preserve local information due to pursuing maximal variance. LDA is used to find the optimal set of projection vectors that maximize the determinant of the between-class scatter matrix and at the same time minimize the determinant of the within-class scatter matrix. But, the dimension of vectors is high and the number of observations is small, usually tens or hundreds of samples. An intrinsic limitation of traditional LDA is that it fails to work when the within-class scatter matrix becomes singular, which is known as the SSS problems. Furthermore, class discrimination in LDA is based upon interclass and intraclass scatters, which is optimal only in cases where the data of each class is approximately Gaussian distributed, a property that cannot always be satisfied in real-world applications.

Compared with traditional PCA, 2DPCA [3] extracts image features directly from two-dimensional image matrices rather than one-dimensional vectors so the image matrices do not need to be transformed into vectors. An image covariance matrix is constructed from the original image matrices for feature extraction. The optimal projection axes are its orthogonal eigenvectors corresponding to its largest eigenvalues. Due to the smaller size of image variance matrix than original variance matrix, 2DPCA requires less time to extract image features and achieves a better recognition rate. Li and Yuan [4] extended the idea of using directly image matrix for LDA and presented 2DLDA. Image between class variance matrix and image with-class variance matrix were constructed for 2DLDA. Now $(2D)^2$PCA[5] and $(2D)^2$FLD[6] have been proposed, in which the authors investigated two-directional two-dimensional projections not only in row direction but also in column direction to further reduce the dimension. But, they do not in essence solve the SSS problem since they both use the Fisher discriminant criterion to find a set of optimal discriminant vectors.

He et al. [7,8] proposed locality preserving projections (LPP), which is a linear subspace learning method derived from Laplacian Eigenmap. Laplacianfaces [9] found an embedding that

* Corresponding author.
E-mail addresses: wmh36@mail.njust.edu.cn, wmh36@sina.com (M. Wan), lai_zhi_hui@163.com (Z. Lai), sj012328@163.com (J. Shao), zhong.jin@patternrecognition.cn (Z. Jin).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2009.07.015

preserves local information, and obtains a face subspace that best preserves the essential face manifold structure. The essence of two-dimensional Laplacianfaces [10] is to map nearby points on a manifold to nearby points in a low dimensional space. They construct a similarity matrix of data points, and then minimize the sum of square difference of feature weighted by the similarity matrix elements. The optimal projection axes best preserve the local structure of the underlying distribution in the L^2 Euclidean space. From analysis they found that LPP is connected with PCA and LDA. LPP can be found an embedding that preserves local information. But if the training samples are insufficient and data dimension is high especially for image data, LPP cannot be used directly due to singularity of matrices. Hence, 2DLPP [11,12] was proposed to directly extract the proper features from image matrices based on the locality preserving criterion.

Recently, Yan et al. [13] proposed a general dimensionality reduction framework called graph embedding. In this paper, marginal Fisher analysis (MFA) was combined locality and class label information to represent the intraclass compactness and interclass separability. So, MFA can be viewed as supervised variants of LPP or as localized variants of LDA since it focus on the characterization of intraclass separability and interclass locality. Motivated by the idea of 2DPCA, which operates directly on image matrix, we develop a novel dimensionality reduction algorithm, two-dimensional local graph embedding discriminant analysis (2DLGEDA). In 2DLGEDA, the intrinsic graph is designed to characterize intraclass compactness, and the penalty graph is formulated for interclass separability. In the intrinsic graph, a vertex pair is connected if one vertex is among the k_1-nearest neighbors of the other class and the elements of the pair belong to the same class. In the penalty graph, for each class, the k_2-nearest vertex pairs in which one element is in-class and the other is out-of-class are connected. 2DLGEDA has the following advantages:

- The number of available projection directions is much larger than that of LDA.
- There is no assumption on the data distribution so that it is more general for discriminant analysis.
- Without a prior assumption on data distributions, the interclass margin can better characterize the separability of different classes than the interclass scatter in LDA.
- The difference criterion can avoid the small sample size problem occurred in traditional Fisher discriminant analysis.
- Extracting the features becomes simple and more straightforward. 2DLGEDA directly extracts the optimal projective vectors from two-dimensional face image matrices rather than vectors, reserves useful structural information embedding in the original images.

The rest of the paper is structured as follows: In Section 2 we introduce 2DPCA, 2DLDA and 2DLPP, In Section 3, we propose the idea and describe 2DLGEDA in detail. In Section 4, experiments on ORL, YALE, FERET face databases and PolyU palmprint database are presented to demonstrate the effectiveness of 2DLGEDA. Finally, we give concluding remarks and a discussion of future work in Section 5.

2. Outline of 2DPCA, 2DLDA and 2DLPP

In this section, we repeated three linear projection methods, which appeared recently in the image recognition literature, namely 2DPCA [3], 2DLDA [4] and 2DLPP [11].

Suppose $\omega = [\omega_1, \omega_2, \ldots, \omega_d]$ is an $m \times d$-dimensional matrix, where ω_i is a unitary column vector. Let N denotes the total sample number, and X_i $(i = 1, 2, \ldots, N)$ denotes an $(m \times n)$-dimensional image matrix. Now the linear methods are to project X_i onto ω by the following linear projection:

$$Y_i = \omega^T X_i i = 1, 2, \ldots, N \tag{1}$$

2.1. Two-dimensional PCA

2DPCA seeks a projection direction ω which maximizes the total scatter of the resulting projected samples. Yang et al. [3] chose the following criterion:

$$J(\omega) = tr(S_\omega) = \omega^T G_t \omega \tag{2}$$

where S_ω is the covariance matrix of the projected feature vectors of the training samples and $tr(S_\omega)$ is the trace of S_ω. Let $\overline{X}$ is the total mean image and G_t is the image covariance (scatter) matrix:

$$G_t = \frac{1}{N}\sum_{j-1}^{N}(X_j - \overline{X})(X_j - \overline{X})^T \tag{3}$$

Then the optimal projection matrix $\omega = [\omega_1, \omega_2, \ldots, \omega_d]$ is the orthonormal eigenvectors of G_t corresponding to the first d largest eigenvalues.

Yang et al. [3] selected features on two-dimensional images rather than one-dimensional vectors. Their method reduced computational complexity greatly compared with traditional PCA and also improved recognition rate.

2.2. Two-dimensional LDA

2DPCA was aimed at preserving maximal variance, while 2DLDA was aimed at preserving maximal discrimination. Suppose each of $X_1, X_2, \ldots, X_N$ belongs to one of c classes $B_1, B_2, \ldots, B_c$. The projection direction is chosen as the following:

$$w = \arg\max \frac{\omega^T G_b \omega}{\omega^T G_w \omega} \tag{4}$$

$$G_b = \frac{1}{N}\sum_{i-1}^{c} N_i(\overline{X}_i - \overline{X})(\overline{X}_i - \overline{X})^T \tag{5}$$

$$G_w = \frac{1}{N}\sum_{i-1}^{c}\sum_{X_j \in B_i}(X_j - \overline{X}_i)(X_j - \overline{X}_i)^T \tag{6}$$

where $\overline{X}$ is the total mean image and $\overline{X}_i$ is the mean image of the ith class B_i. N_i is the number of sample images in the ith class B_i. The matrix G_b is called between-class image scatter matrix and G_ω is called within-class image scatter matrix. The optimal projection axes are obtained by solving the generalized eigenvalue problem and they are exactly the orthogonal generalized eigenvectors corresponding to the first d largest generalized eigenvalues of the generalized eigenequation:

$$G_\omega^{-1} G_b w = \lambda\omega \tag{7}$$

2.3. Two-dimensional LPP

Since a node in the nearest-neighbor graph corresponds to an image X_i, the purpose of 2DLPP is to map a node from $(m \times n)$-dimensional image space into a n-dimensional Euclidean space, and to ensure the connected nodes stay as close as possible and the intrinsic geometry of the data and local structure is preserved [14]. The similarity matrix Scan be Gaussian weight or uniform weight of Euclidean distance using k-neighborhood or ε-neighborhood, which was defined as

$$S_{ij} = \begin{cases} 1, & \|X_i - X_j\|^2 < \varepsilon \\ 0, & \text{otherwise} \end{cases}$$