Applied Mathematics and Computation 217 (2011) 9659–9668

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

Feature extraction using two-dimensional local graph embedding based on maximum margin criterion

Minghua Wan [a,c,*], Zhihui Lai [b,c], Zhong Jin [c]

[a] School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China
[b] Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518005, China
[c] School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China

ARTICLE INFO

Keywords:
Feature extraction
Two-dimensional Fisherfaces (2DLDA)
Two-dimensional Laplacianfaces (2DLPP)
Two-dimensional Local Graph Embedding Discriminant Analysis (2DLGEDA)
Maximum margin criterion (MMC)

ABSTRACT

In this paper, we propose a novel method for image feature extraction, namely the two-dimensional local graph embedding, which is based on maximum margin criterion and thus not necessary to convert the image matrix into high-dimensional image vector and directly avoid computing the inverse matrix in the discriminant criterion. This method directly learns the optimal projective vectors from 2D image matrices by simultaneously considering local graph embedding and maximum margin criterion. The proposed method avoids huge feature matrix problem in Eigenfaces, Fisherfaces, Laplacianfaces, maximum margin criterion (MMC) and inverse matrix in 2D Fisherfaces, 2D Laplacianfaces and 2D Local Graph Embedding Discriminant Analysis (2DLGEDA) so that computational time would be saved for feature extraction. Experimental results on the Yale and the USPS databases show the effectiveness of the proposed method under various experimental conditions.

1. Introduction

Techniques for dimensionality reduction in linear and nonlinear learning tasks have attracted much attention in the areas of pattern recognition and computer vision. Linear dimensionality reduction seeks to find a meaningful low dimensional subspace from a high-dimensional input space. The derived subspace can provide a compact representation of the input data. Two of the most fundamental linear dimensionality reduction methods are principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2].

PCA aims to find a linear mapping, which preserves the total variance by maximizing the trace of feature covariance matrix. The optimal projections of PCA are corresponding to the first k-largest Eigenvalues of the data's total covariance matrix. LDA is used to find the optimal set of projection vectors that maximize the ratio of the between-class scatter matrix and at the same time minimize the determinant of the within-class scatter matrix. But, since 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 small sample size (SSS) problems. So far many effective and efficient methods [4–14] have been explored to solve the problem. To avoid the singularity problem of LDA, Li et al. [3] used the difference of both between-class scatter and within-class scatter as discriminant

* Corresponding author at: School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China.
E-mail addresses: wmh36@mail.njust.edu.cn, wmh36@sina.com (M. Wan).

0096-3003/$ - see front matter
doi:10.1016/j.amc.2011.04.050

criterion, which is called maximum margin criterion (MMC). Since the inverse matrix does not need to be computed, the SSS problem in traditional LDA is alleviated.

In some applications such as face and handwriting digit recognition, 2D images are usually transformed into 1D vector through column by column or row by row concatenation. On the one hand, the image-to-vector transform procedure may cause the loss of some useful structural information embedding in the original images. On the other hand, the resulting 1D image vectors of faces or handwriting digit usually lead to a high dimensional image vector space. To overcome the problem of high dimension, 2DPCA [15,16] extracts image features directly from 2D image matrices rather than 1D vectors and the image matrices do not need to be transformed into vectors. The optimal projection axes are the orthogonal Eigenvectors corresponding to the largest Eigenvalues. Due to the smaller size of image covariance matrix than original covariance matrix, 2DPCA requires less time to learn the optimal projections and achieves a better recognition rate. Li et al. [17–20] extended the idea of directly using image matrix for LDA and presented 2DLDA. Image between-class covariance matrix and image within-class covariance matrix were constructed for 2DLDA. Now $(2D)^2$PCA [21], $(2D)^2$FLD [22] and $(2D)^2$ PCALDA [23] have been proposed, in which the authors investigated two-directional two-dimensional projections to further reduce the dimension not only in row direction but also in column direction.

PCA, LDA and theirs extension methods have been successfully applied to some linear data. However, they fail to explore the essential structure of the data with nonlinear distribution. In order to overcome this problem, many nonlinear feature extraction methods including kernel-based techniques and manifold learning based techniques have been developed. Kernel-based technique is implicitly mapping the observed patterns into potentially much high dimensional feature space by a kernel trick Since ti is possible that the nonlinear data will be linearly separable in the kernel space. The widely used kernel techniques are kernel principal component analysis (KPCA) [24] and kernel Fisher discriminant analysis (KFDA) [25], which can be viewed as the kernel versions of PCA and LDA. KPCA and KFDA have been proved to be effective in some real world applications. The kernel based methods can improve the linear discriminability at the cost of increasing dimensions and the high computational cost. Furthermore, due to introducing the kernel trick, how to select different kernels and how to assign the optimal parameters in kernels remain unclear. In most of the cases, experience still plays an important role.

Unlike kernel-based methods, manifold learning-based methods are straightforward in finding the inherent nonlinear structure hidden in the observe space. In the past few years many manifold learning-based algorithms have been presented. Among them, isometric feature mapping (ISOMAP) [26], locally linear embedding (LLE) [27,28] and Laplacian Eigenmap (LE) [29,30] are widely used. They have yielded impressive results on artificial and real world data sets. Salakhutdinov and Hinton [42] used the reconstruction error as a regularizer and fine-tuned a deep nonlinear encoder network to learn a similarity metric for nearest-neighbor classification.

In the real world, nonlinear data include non-Gaussian and manifold-value data. We usually deal with non-Gaussian data from local patches because it can be viewed locally Gaussian and acurved manifold-value data which can be viewed locally Euclidean [31,32]. Recently, He et al. [33,34] proposed locality preserving projections (LPP), which is a linear subspace learning method derived from Laplacian Eigenmap. Laplacianfaces aim to find an embedding space by preserved local information and detected the essential face manifold structure. The optimal projection axes 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 do not use the class label information and it is a unsupervised method. Xu et al. [43] propose three novel solution schemes to solve the small sample size (SSS) problem. Chen et al. proposed the local discriminant embedding (LDE) [36] for feature extraction and recognition. It combined locality and class label information to represent the intraclass compactness and interclass separability. LDE take advantage of the partial structural information of classes and neighborhoods of samples. However, it is difficult to decide the number of nearest neighbors of each sample and the number of nearest point pairs from different classes in LDE. But if the training samples are insufficient and data dimension is high, especially for image data, LPP and LDE cannot be used directly due to singularity of within-class scatter matrices. Hence, 2D-LPP [35,37,38] was proposed to directly extract the proper features from image matrices based on the locality preserving criterion. Recently, some variant versions of 2DLPP such as two dimensional local graph embedding discriminant analysis (2DLGEDA) [39] and two dimensional discriminant locality preserving projection (2DDLPP) [40] were also proposed to improve the performance of the 2DLPP.

However, the computational cost of 2DLPP and theirs extension methods are high because they involve dense matrix Eigen-decomposition and singularity of within-class scatter matrices. So, in this paper we present two-dimensional local graph embedding based on maximum margin criterion. Therefore, computational time would be saved for feature extraction since it is not necessary to convert the image matrix into high-dimensional image vector and can avoid computing the inverse matrix. This method directly computes the optimal projective vectors from 2D image matrices by simultaneously considering local graph embedding [41] and maximum margin criterion techniques. 2DLDA, 2DLPP and 2DLGEDA must compute the inverse matrix, while the proposed method avoids this computation successfully by the virtue of trace difference, which potentially saves computational time on learning procedures.

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

2. Related works

Now let us consider a set of N sample images$X_1, X_2, \ldots, X_N$ taken from an $(m \times n)$-dimensional image space. We design a linear transformation, which maps the original $(m \times n)$-dimensional image space into an $n \times d$-dimensional feature space. Let$\Omega = [\omega_1, \omega_2, \ldots, \omega_d]$ is an $n \times d$-dimensional matrix, where ω_i is a unitary column vector. The method proposed here is to project each image X_i, an $(m \times n)$ matrix, onto Ω by the following transformation:

$$Y_i = X_i\Omega, \quad i = 1, 2, \ldots, N. \tag{1}$$

Then we get a $m \times d$-dimensional projected feature Y_i for each image X_i.

2.1. Two-dimensional Laplacianfaces (2DLPP) [33]

Let $G = \{X,S\}$ be a complete undirected weighted graph with vertex set X and similarity matrix $S \in R^{N\times N}$. Since a node in the nearest-neighbor graph corresponds to an image X_i, the purpose of 2DLPP is to ensure the connected nodes stay as close as possible and the intrinsic geometry of the data and local structure is preserved. The similarity matrix S can be Gaussian weight or uniform weight of Euclidean distance using k-neighborhood or ε-neighborhood, which was defined as:

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

Hence, the objective function of 2DLPP is defined as:

$$\min \sum_{i,j} \|Y_i - Y_j\|^2 S_{ij}, \tag{3}$$

where $Y_i = \omega^T X_i$ and $\|\bullet\|$ represents the L_2 norm. After some matrix analysis steps, the minimization problem of Eq. (3) becomes:

$$\begin{aligned} &\arg\min_{\omega} \omega^T X(L \otimes I_n)X^T\omega, \\ &\text{s.t. } \omega^T X(D \otimes I_n)X^T\omega = 1, \end{aligned} \tag{4}$$

where $X = [X_1, X_2, \ldots, X_N]$ is the training space of size $N(m \times n)$, and D is a diagonal matrix whose entries are column or row sums of S. $L = D - S$ is the Laplacian matrix, and I_n is an identity matrix of order n, operator $\otimes$ is the Kronecher product of the matrices.

The optimal d projection vectors that minimizes the objective function is computed by the minimum Eigenvalue solutions to the generalized Eigenvalue problem

$$X(L \otimes I_n)X^T\omega = \lambda X(D \otimes I_n)X^T\omega. \tag{5}$$

2.2. Two-dimensional Local Graph Embedding Discriminant Analysis (2DLGEDA) [39]

2DLGEDA was proposed as a supervised extension of 2DLPP that works directly on 2D images. The goal of 2DLGEDA is to preserve the 2D image within-class compactness and maximize the between-class separability. 2D image within-class compactness is characterized from the intrinsic graph W_{ij}^w by the term:

$$S_w = \sum_{i=1}^{N}\sum_{j=1}^{N} \|Y_i - Y_j\|^2 W_{ij}^c = 2\omega^T X(L^w \otimes I_n)X^T\omega, \tag{6}$$

where

$$W_{ij}^w = \begin{cases} 1, & X_i \in N_{k_w}^+(X_j) \text{ or } X_j \in N_{k_w}^+(X_i), \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

where $X = [X_1, X_2, \ldots, X_N]$ is the 2D image training sample matrix of size $N(m \times n)$, and D^w is a diagonal matrix whose entries are column or row sums of $W^w, N_{k_w}^+(X_i)$ indicates the samples in the k_w nearest neighbors of X_i in the same class, and ω denotes the projections vector,$L^w = D^w - W^w$.

Similarly, 2D image between-class separability is characterized from the between-class graph $W_{ij}^b$by the term:

$$S_b = \sum_{i=1}^{N}\sum_{j=1}^{N} \|Y_i - Y_j\|^2 W_{ij}^b = 2\omega^T X(L^b \otimes I_n)X^T\omega, \tag{8}$$

where

$$W_{ij}^b = \begin{cases} 1, & if\,(i,j) \in P_{k_b}(\pi_t)\ or\ (j,i) \in P_{k_b}(\pi_t), \\ 0, & \text{otherwise}, \end{cases} \tag{9}$$

where $P_{k_b}(\pi_t)$ is a set of data pairs that are in the k_b nearest pairs among the set $\{(i,j)|i \in \pi_t, j \notin \pi_t\}$, where π_t denotes the index set of tth class and t varied from 1 to c. D^b is a diagonal matrix whose entries are column or row sums of W^b, and $L^b = D^b - W^b$.

Finally, the criterion of 2DLGEDA is formally similar to the Fisher criterion since they are both Reyleigh quotients and the optimal projections can be obtained from solving the generalized Eigen-equation:

$$X^T(L^b \otimes I_n)X\omega = \lambda X^T(L^w \otimes I_n)X\omega, \tag{10}$$

where λ is generalized Eigenvalue corresponding to the Eigenvector ω. Then, the optimal transformation matrix of 2DLGEDA is composed of the Eigenvectors associated with the d top Eigenvalues.

3. Two-dimensional local graph embedding based on maximum margin criterion

3.1. Fundamentals

For a given sample X_i, the class label of the sample X_i is assumed to be π_t, $t \in \{1,2,\ldots,c\}$. We can divide the other measurements into two groups: measurements in the same class with X_i and measurements from different classes with X_i.

We select k_1 nearest neighbors with respect to X_i from measurements in the same class with X_i and k_2 nearest neighbors with respect to X_i from measurements in different classes with X_i. By putting $k_1$nearest neighbors and k_2 nearest neighbors together, we can build the local graph embedding for the measurement X_i as $\widetilde{X}_i = \left[X_i, X_i^1, \ldots, X_i^{k_1}, X_{i1}, \ldots, X_{ik_2}\right]$.

For each local graph embedding, the corresponding output in the low dimensional space is denoted by $\widetilde{Y}_i = \left[Y_i, Y_i^1, \ldots, Y_i^{k_1}, Y_{i1}, \ldots, Y_{ik_2}\right]$. In the low-dimensional space, we expect that distances between the given measurement and the neighbor measurements of a same class are as small as possible, while distances between the given measurement and the neighbor measurements of different classes are as large as possible.

For each local graph embedding in the low-dimensional subspace, we expect that distances between Y_i and the neighbor measurements of the same class are as small as possible, so we have:

$$\begin{aligned} J_c(\omega) &= \min \sum_{i=1}^{N}\sum_{j=1}^{N} \left\|Y_i - Y_i^j\right\|^2 = \min \sum_{i=1}^{N}\sum_{j=1}^{N} \|Y_i - Y_j\|^2 W_{ij}^c = \min \sum_{i=1}^{N}\sum_{j=1}^{N} \|\omega^T X_i - \omega^T X_j\|^2 W_{ij}^c \\ &= 2\omega^T X((D^c - W^c) \otimes I_n)X^T\omega = 2\omega^T X(L^c \otimes I_n)X^T\omega, \end{aligned} \tag{11}$$

where

$$W_{ij}^c = \begin{cases} 1, & if\ X_j\ is\ in\ the\ K_c\ nearest\ from\ same\ class\ of\ X_i, \\ 0, & otherwise. \end{cases} \tag{12}$$

where $L^c = D^c - W^c$.

Meanwhile, we expect that distances between Y_i and the neighbor measurements of different classes are as large as possible, so we have

$$\begin{aligned} J_p(\omega) &= \max \sum_{i=1}^{N}\sum_{q=1}^{k_2} \|Y_i - Y_{iq}\|^2 = \max \sum_{i=1}^{N}\sum_{j=1}^{N} \|Y_i - Y_j\|^2 W_{ij}^p = \max \sum_{i=1}^{N}\sum_{j=1}^{N} \|\omega^T X_i - \omega^T X_j\|^2 W_{ij}^p \\ &= 2\omega^T X((D^p - W^p) \otimes I_n)X^T\omega = 2\omega^T X(L^p \otimes I_n)X^T\omega, \end{aligned} \tag{13}$$

where

$$W_{ij}^p = \begin{cases} 1, & if\ X_j\ is\ in\ the\ K_p nearest\ from\ different\ classes\ of\ X_i \\ 0, & otherwise. \end{cases} \tag{14}$$

where $L^p = D^p - W^p$.

Since the local graph embedding formed by the local neighborhood can be regarded approximately linear, an optimization objective function can be devised to minimize the difference between the intraclass compactness scatter and the interclass separability scatter as follows:

$$\begin{aligned} J(\omega) &= \min\left(\sum_{j=1}^{k_1} \left\|Y_i - Y_i^j\right\|^2 - \alpha \sum_{q=1}^{k_2} \|Y_i - Y_{iq}\|^2\right) = \min\left(\sum_{i=1}^{N}\sum_{j=1}^{N} \|Y_i - Y_j\|^2 W_{ij}^c - \alpha \sum_{i=1}^{N}\sum_{j=1}^{N} \|Y_i - Y_j\|^2 W_{ij}^p\right) \\ &= \min\left(\sum_{i=1}^{N}\sum_{j=1}^{N} \|\omega^T X_i - \omega^T X_j\|^2 W_{ij}^c - \alpha \sum_{i=1}^{N}\sum_{j=1}^{N} \|\omega^T X_i - \omega^T X_j\|^2 W_{ij}^p\right) \\ &= tr(\omega^T X(D^c - W^c)X^T\omega - \alpha\omega^T X(D^p - W^p)X^T\omega) = tr(\omega^T X(L^c \otimes I_n)X^T\omega - \alpha\omega^T X(L^p \otimes I_n)X^T\omega), \end{aligned} \tag{15}$$

whereα is an adjustable parameter which balances $J_c(\omega)$ and $J_p(\omega)$.

Then we can easily find that Ω consists of the Eigen vectors associated with top d Eigenvalues of the above Eigen-equation.

After training by the proposed method, the feature matrix of each image and a transformation matrix are obtained. Then a nearest-neighbor classifier can be used for classification.

Given two imagesX_1, X_2 represented by the proposed method feature matrices $Y_1 = (y_1^1, y_1^2, \ldots, y_1^d)$ and $Y_2 = (y_2^1, y_2^2, \ldots, y_2^d)$, then the dissimilarity $d(Y_1, Y_2)$ is defined as:

$$d(Y_1, Y_2) = \sum_{k=1}^{d} \left\| Y_1^k - Y_2^k \right\|^2. \tag{16}$$

If the feature matrices of training images are Y_1, Y_2, ..., Y_N (N is the total number of training images), and each image is assigned to a class π_t. Then for a given test image$Y$, if $d(Y, Y_{\pi_l}) = \min_j d(Y, Y_j)$ and $Y_j \in \pi_j$, the resulting decision is$Y \in \pi_l$.

3.2. The outline of the proposed method

The proposed method based feature extraction algorithm can be summarized as follows:

Step 1: Construct the similarity matrix W_{ij}^c and W_{ij}^p using Eqs. (12) and (14).
Step 2: Calculate the intraclass compactness scatter $J_c(\omega)$ using Eq. (11). Calculate the intraclass compactness scatter $J_p(\omega)$ using Eq. (13).
Step 3: Extract the sample feature using Eq. (15).
Step 4: Projecting all samples onto the obtained optimal discriminant vectors and yielding the projected Eigenvectors using Eq. (1).
Step 5: Classifying the projected Eigenvectors with a classifier using Eq. (16).

4. Experiments

To evaluate the proposed method algorithm, we compared it with the PCA, LDA, LPP, MMC, 2DPCA, 2DLDA, 2DLPP and 2DLGEDA algorithms on two databases: the Yale face databases and the USPS database. When the projection matrix was computed from the training part, all the images including training part and the test part were projected to feature space. nearest neighborhood classifier with Euclidean distance was used in all the experiments. The experiments were carried out on a PC (CPU: P4 2.8 GHz, RAM: 1024 MB).

Fig. 1. Sample images of one person in the Yale database.

Table 1
The maximal average recognition rates (percent) of each method on the Yale face database and the corresponding dimensions when the 2, 3, 4, 5, 6 samples per class are randomly selected for training and the remaining 9, 8, 7, 6, 5 images respectively for test.

Method	Number of training images				
	2	3	4	5	6
PCA % (dim)	78.49 (29)	81.47 (40)	85.37 (36)	85.96 (40)	87.01 (46)
LDA % (dim)	81.93 (14)	85.61 (14)	88.30 (14)	88.84 (14)	89.36 (14)
LPP % (dim)	81.45 (22)	85.97 (24)	88.57 (21)	89.00 (18)	90.40 (21)
MMC % (dim)	81.29 (21)	83.72 (14)	86.99 (14)	87.20 (14)	88.29 (14)
2DPCA % (dim)	88.67 (50 × 36)	90.58 (50 × 35)	91.05 (50 × 35)	92.00 (50 × 35)	93.73 (50 × 35)
2DLDA % (dim)	88.37 (50 × 15)	89.75 (50 × 39)	93.14 (50 × 20)	93.22 (50 × 20)	94.67 (50 × 3)
2DLPP % (dim)	83.93 (50 × 6)	86.00 (50 × 4)	94.10 (50 × 3)	93.11 (50 × 11)	93.47 (50 × 14)
2DLGEDA % (dim)	89.70 (50 × 39)	91.61 (50 × 38)	93.30 (50 × 38)	93.96 (50 × 38)	94.68 (50 × 37)
The proposed method % (dim)	**90.90** (50 × 36)	**92.08** (50 × 36)	**94.23** (50 × 38)	**94.49** (50 × 37)	**95.08** (50 × 37)

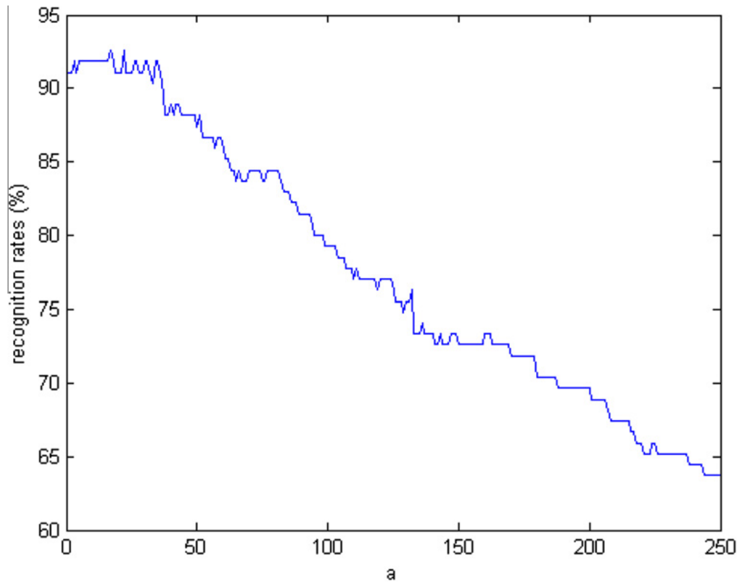

Fig. 2. The recognition rates (%) of the proposed method when the 2 images per person are randomly selected for training on the Yale face database with varied α.

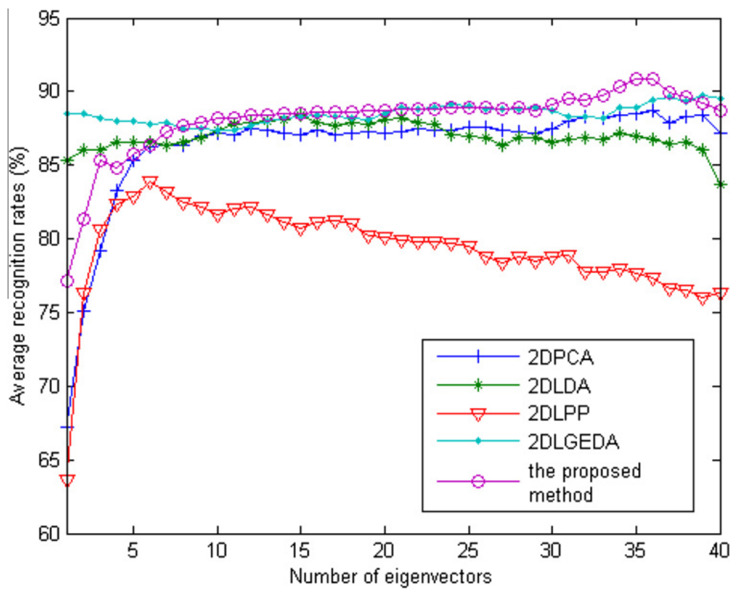

Fig. 3. The average recognition rates (%) of 2DPCA, 2DLDA, 2DLPP, 2DLGEDA and the proposed method versus the dimensions when the 2 images per person are randomly selected for training on the Yale face database. The dimension here is the number of Eigenvectors.

4.1. Experiment on the Yale face Database

The Yale face database (http://www.cvc.yale.edu/projects/yalefaces/yalefaces.html) contains 165 images of 15 individuals (each person providing 11 different images) under various facial expressions and lighting conditions (i.e., center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised, and winking). In our experiments, each image was manually cropped and resized to 50×40 pixels. Fig. 1 shows sample images of one person on the Yale database. In the experiments, l images (l varies from 2 to 6) are randomly selected from the image gallery of each individual to form the training sample set. The remaining $N - l$ images are used for test. For each$l$, we independently run 10 times. We varied K_c from 1 to $l - 1$ andK_p from 2 to 20 with step 2.

For feature extraction, we used, respectively, PCA, LDA, LPP, MMC, 2DPCA, 2DLDA, 2DLGEDA 2DLPP and the proposed method. In the PCA phase of LDA, LPP and MMC, we keep 90 percent image energy. The maximal average recognition rate of each method and the corresponding dimension are given in Table 1 when the 2, 3, 4, 5, 6 samples per class are randomly selected for training and the remaining 9, 8, 7, 6, 5 images are respectively for test. Table 1 presents the top recognition accuracy of the each method, which corresponds to different number of images per person used for training.

To find how the weight parameter α affects the recognition performance, we changed α from 1 to 250 with step 1. Fig. 2 displays the recognition rates with varied parameter α by carrying out the proposed method. The recognition rates (%) of the proposed method when the 2 images per person are randomly selected for training on the Yale face database with varied α. And this experiment run 1 times. From Fig. 2, it can be found that the proposed method obtains the best recognition rate is 92.56% when α=16.

Fig. 3 shows the average recognition rates (%) of 2DPCA, 2DLDA, 2DLPP, 2DLGEDA and the proposed method versus the dimensions when the 2 images per person are randomly selected for training on the Yale face database. From the figure, it is observed that the proposed method outperformed 2DPCA, 2DLDA, 2DLPP and 2DLGEDA methods comprehensively.

In addition, the average CPU time consumed for training, test and classification, and the maximal average recognition rates of the foregoing nine methods are given in Table 2. The proposed method achieves its maximal recognition rate of 90.90%, and it needs less CPU time compared to other methods.

4.2. Experiments on USPS handwriting database

The USPS handwriting digital data include 10 classes from "0" to "9". Each class has 1100 examples. In our experiment, we select a subset from the original database. We cropped each image to be size of 16×16. There are 100 images for each class in the subset and the total number is 1000. Fig. 4 displays a subset of digital "2" from original USPS handwriting digital

Table 2
The average CPU time (s) consumed for training, test and classification, and the maximal average recognition rates (%) when the 2 images per person are randomly selected for training on the Yale face database.

Methods	PCA	LDA	LPP	MMC	2DPCA	2DLDA	2DLPP	2DLGEDA	The proposed method
Recognition rate (%)	78.49	81.93	81.45	81.29	88.67	88.37	83.93	89.70	**90.90**
dim	(29)	(14)	(22)	(21)	(50×36)	(50×15)	(50×6)	(50×39)	(50×36)
CPU time (s)	0.155	0.151	0.152	0.140	0.0606	0.0589	0.0592	0.0587	**0.0559**

Fig. 4. The sample digital images "2" from USPS handwriting database.

Table 3
The maximal average recognition rates (percent) of each method on the USPS database and the corresponding dimensions when the 20, 30, 40, 50, 60 samples per class are randomly selected for training and the remaining 80, 70, 60, 50, 40 images respectively for test.

Method	Number of training images				
	20	30	40	50	60
PCA % (dim)	80.88 (20)	84.56 (20)	86.72 (29)	87.96 (26)	88.9 (27)
LDA % (dim)	82.72 (7)	85.83 (9)	86.80 (8)	88.00 (9)	88.57 (9)
LPP % (dim)	78.93 (25)	82.75 (14)	85.70 (29)	86.78 (13)	88.82 (17)
MMC % (dim)	79.85 (30)	83.74 (27)	86.43 (27)	87.88 (27)	89.40 (27)
2DPCA % (dim)	81.56 (16×3)	85.41 (16×3)	87.83 (16×3)	88.82 (16×4)	89.98 (16×4)
2DLDA % (dim)	78.01 (16×15)	81.80 (16×14)	84.60 (16×1)	86.04 (16×1)	87.32 (16×1)
2DLPP % (dim)	77.51 (16×5)	79.66 (16×4)	84.82 (16×4)	86.86 (16×5)	85.35 (16×5)
2DLGEDA % (dim)	81.75 (16×14)	83.93 (16×15)	87.80 (16×14)	88.96 (16×14)	89.43 (16×14)
The proposed method % (dim)	**82.34** (16×3)	**85.93** (16×3)	**88.45** (16×3)	**89.80** (16×3)	**90.25** (16×3)

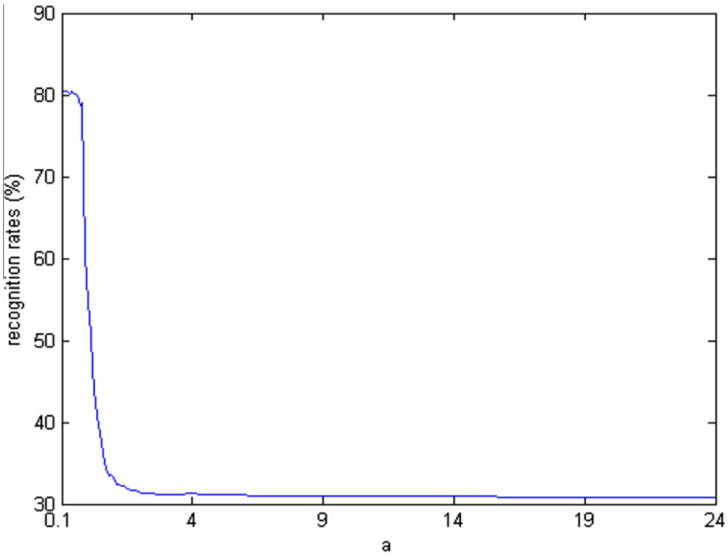

Fig. 5. The recognition rates (%) of the proposed method when the 20 images per class are randomly selected for training on the USPS database with varied α.

database. In the experiments, l images (l varies from 20 to 60) are randomly selected from the image gallery of each individual to form the training sample set. The remaining $N - l$ images are used for test. For each l, we independently run 10 times. We varied$K_c = l - 1$ andK_p from 5 to 50 with step 5.

In this experiment, we first perform PCA with keep 90 percent image energy on the data and then apply LDA, LPP and MMC on the PCA subspace. Table 3 shows the best maximal average recognition rates and the corresponding dimensions after carrying out PCA, LDA, LPP, MMC, 2DPCA, 2DLDA, 2DLPP, 2DLGEDA and the proposed method.

In this experiment, we also test the impact of αon the performance, which can be found in Fig. 5. We varied α from 0.1 to 24 with step 0.1. And this experiment also run 1 times to get α. Fig. 5 displays the recognition rates with varied parameter α by carrying out the proposed method. Then it can be found that when α equals to 0.4, the recognition rate gains the best value 80.38 %.

Fig. 6 shows the average recognition rates (%) of 2DPCA, 2DLDA, 2DLPP, 2DLGEDA and the proposed method versus the dimensions when the 20 images per class are randomly selected for training on the USPS database. It is observed from the Fig. 6 that the proposed method's performance is far better than 2DPCA, 2DLDA, 2DLPP and 2DLGEDA method.

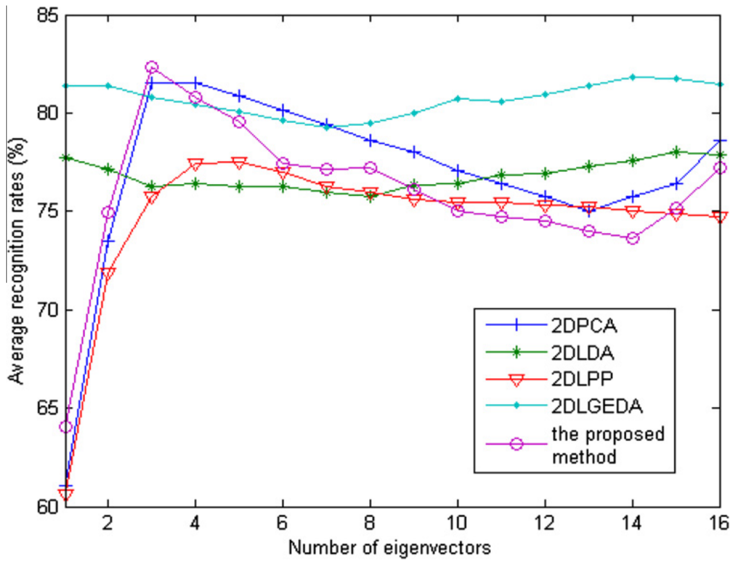

Fig. 6. The average recognition rates (%) of 2DPCA, 2DLDA, 2DLPP, 2DLGEDA and the proposed method versus the dimensions when the 20 images per class are randomly selected for training on the USPS database. The dimension here is the number of Eigenvectors.

Table 4
The average CPU time (s) consumed for training, test and classification, and the maximal average recognition rates (%) when the 20 images per class are randomly selected for training on the USPS database.

Methods	PCA	LDA	LPP	MMC	2DPCA	2DLDA	2DLPP	2DLGEDA	The proposed method
Recognition rate (%)	80.88	82.72	78.93	79.85	81.56	78.01	77.51	81.75	**82.34**
dim	(20)	(7)	(25)	(30)	(16 × 3)	(16 × 15)	(16 × 5)	(16 × 14)	(16 × 3)
CPU time (s)	1.542	1.504	1.478	1.014	0.610	0.598	0.587	0.601	**0.548**

In addition, the average CPU time consumed for training, test and classification, and the top recognition rates of the foregoing nine methods are given in Table 4. The proposed method achieves its top recognition rate of 82.34%, and it needs less CPU time compared to other methods.

4.3. Observations and evaluations of the experimental results

The above experiments showed that the top recognition rate of the proposed method is always higher than other methods. From the experiments we can draw the following conclusions in details:

- Simple and more straightforward. 2D feature extraction methods such as 2DPCA, 2DLDA, 2DPCA and the proposed method can directly extract the optimal projective vectors from 2D face image matrices rather than 1D feature extraction methods such as PCA, LDA, LPP, MMC, and reserve useful structural information embedding in the original images. And the proposed method consistently outperforms PCA, LDA, LPP, MMC, 2DPCA, 2DLDA, 2DLPP and 2DLGEDA inspire of the variation of dimensions, which are shown in Tables 1 and 3.
- Efficient computation. 2DLDA, 2DLPP and 2DLGEDA must compute the inverse matrix of discriminant criterion, while the proposed method avoids this computation successfully by virtue of trace difference, which saves much computational time on feature extraction. The proposed method needs less CPU time compared to other methods which are shown in Tables 2 and 4.
- The average recognition rates (%) of the proposed method versus the dimensions is better than other methods, which are shown in Fig. 3 and Fig. 6.
- How to select parameter α, K_c andK_p is still an open problems in feature extraction.

5. Conclusions

In pattern recognition, feature extraction techniques are widely employed to reduce the dimensionality of data and enhance the discriminatory information. In this paper, we proposed a new method for feature extraction and recognition, namely the two-dimensional local graph embedding based on maximum margin criterion, which can directly extract the optimal projective vectors from 2D image matrices by simultaneously considering local graph embedding and difference criterion techniques. We adopt the difference of the intraclass compactness scatter matrix and the interclass separability scatter matrix according to local graph embedding. The experiments conducted on the Yale face databases and handwriting digital recognition on the USPS database indicates the effectiveness of the proposed method. In the future, we will make more tests on other types of data and decide the optimal parameter α, K_c and K_p.

Acknowledgements

This work is partially supported by the National Science Foundation of China under Grant No. 60632050, 90820306, 60873151, 60973098, 60963002, 61005008 and 61005005.

References

[1] M. Turk, A.P. Pentland, Face recognition using Eigenfaces, IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit. (CVPR'91), June 3-6,(1991), Maui, Hawaii, pp. 586–591.
[2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriengman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intel. 19 (7) (1997) 711–720.
[3] H. Li, T. Jiang, K. Zhang, Efficient and robust feature extraction by maximum margin criterion, IEEE Trans. Neural Netw. 17 (1) (2006) 1157–1165.
[4] P. Howland, J. Wang, H. Park, Solving the small sample size problem in face recognition using generalized discriminant analysis, Pattern Recognit. 39 (2006) 277–287.
[5] W. Zheng, L. Zhao, C. Zou, Foley–Sammon optimal discriminant vectors using kernel approach, IEEE Trans. Neural Netw. 16 (1) (2005) 1–9.
[6] W. Zheng, L. Zhao, C. Zou, An efficient algorithm to solve the small sample size problem for LDA, Pattern Recognit. 37 (2004) 1077–1079.
[7] J.H. Friedman, Regularized discriminant analysis, J. Am. Statist. Assoc. 84 (405) (1989) 165–175.
[8] G.H. Golub, C.F. Van Loan, Matrix Computations, third ed., The Johns Hopkins University Press, Baltimore, MD, 1996.
[9] P.N. Belhumeour, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intel. 19 (7) (1997) 711–720.
[10] D.L. Swets, J. Weng, Using discriminant Eigenfeatures for image retrieval, IEEE Trans. Pattern Anal. Mach. Intel. 18 (8) (1996) 831–836.

[11] P. Howland, M. Jeon, H. Park, Structure preserving dimension reduction for clustered text data based on the generalized singular value decomposition, SIAM J. Matrix Anal. Appl. 25 (1) (2003) 165–179.
[12] J. Ye, R. Janardan, C.H. Park, H. Park, An optimization criterion for generalized discriminant analysis on undersampled problems, IEEE Trans. Pattern Anal. Mach. Intel. 26 (8) (2004) 982–994.
[13] J. Ye, Q. Li, A two-stage linear discriminant analysis via QR-decomposition, IEEE Trans. Pattern Anal. Mach. Intel. 27 (6) (2005) 929–941.
[14] L.-F. Chen, X. Hong-Yuan, M. Liao, M.-T. Ko, J.-C. Lin, G.-J. Yu, A new LDA-based face recognition system which can solve the small sample size problem, Pattern Recognit. 33 (2000) 1713–1726.
[15] J. Yang, D. Zhang, A.F. Frangi, J. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intel. 26 (1) (2004) 131–137.
[16] J. Yang, J.Y. Yang, From image vector to matrix: a straightforward image projection technique—IMPCA vs. PCA, Pattern Recognit. 35 (9) (2002) 1997–1999.
[17] M. Li, B. Yuan, 2D-LDA: a statistical linear discriminant analysis for image matrix, Pattern Recognit. Lett. 26 (2005) 527–532.
[18] J. Yang, D. Zhang, Y. Xu, J.J. Yang, Two-dimensional discriminant transform for face recognition, Pattern Recognit. 38 (2005) 1125–1129.
[19] H.L. Xiong, M.N.S. Swanmy, M.O. Ahmad, Two-dimensional FLD for face recognition, Pattern Recognit. 38 (2005) 1121–1124.
[20] X.Y. Jing, H.S. Wong, D. Zhang, Face recognition based on 2D Fisherface approach, Pattern Recognit. 39 (2006) 707–710.
[21] D.Q. Zhang, Z.H. Zhou, $(2D)^2$PCA: two-directional two-dimensional PCA for efficient face representation and recognition, Neurocomputing 69 (2005) 224–231.
[22] P. Nagabhushan, D.S. Guru, B.H. Shekar, (2D)2FLD: an efficient approach for appearance based object recognition, Neurocomputing 69 (2006) 934–940.
[23] Yongfeng Qi, Jiashu Zhang, $(2D)^2$PCALDA: an efficient approach for face recognition, Applied Math. Comput. 213 (2009) 1–7.
[24] B. Scholkopf, A. Smola, K.R. Muller, Nonlinear component analysis as a kernel Eigenvalue problem, Neural Comput. 10 (5) (1998) 1299–1319.
[25] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, A. Smola, K.-R. Muller, Constructing descriptive and discriminative nonlinear features: Rayleigh coefficients in kernel feature spaces, IEEE Trans. Pattern Anal. Mach. Intel. 25 (5) (2003) 623–628.
[26] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
[27] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[28] L.K. Saul, S.T. Roweis, Think globally, fit locally: unsupervised learning of low dimensional manifolds, J. Mach. Learn. Res. (4) (2003) 119–155.
[29] M. Belkin, P. Niyogi, Laplacian Eigenmaps and spectral techniques for embedding and clustering, In Processing, December 3-6, 2001, Vancouver, Canada.
[30] M. Belkin, P. Niyogi, Laplacian Eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.
[31] M. Kirby, L. Sirovich, Application of the KL procedure for the characterization of human faces, IEEE Trans. Pattern Anal. Mach. Intel. 12 (1) (1990) 103–108.
[32] J.M. Lee, Riemannian Manifolds: An Introduction to Curvature, Springer, Berlin, 1997.
[33] X. He and P. Niyogi, Locality Preserving Projections, In Proceedings of the 17th Annual Conference on Neural Information Processing Systems, December 8–13, 2003, Vancouver and Whistler, Canada.
[34] X. He, S. Yan, Y. Hu, P. Niyogi, H.-J. Zhang, Face Recognition Using Laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intel. 27 (3) (2005) 328–340.
[35] Ben, Yang, Shiu, Pal, Two-dimensional Laplacianfaces method for face recognition, Pattern Recognit. 41 (10) (2008) 3237–3243.
[36] H.T. Chen, H.W. Chang, T.L. Liu, Local discriminant embedding and its variants, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2005), vol. 2, pp. 846–853 (2005).
[37] D. Hu, G. Feng, Z. Zhou, Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition, Pattern Recognit. 40 (1) (2007) 339–342.
[38] S. Chen, H. Zhao, M. Kong, B. Luo, 2DLPP: a two-dimensional extension of locality preserving projections, Neurocomputing 70 (4–6) (2007) 912–921.
[39] M. Wan, Z. Lai, J. Shao, Z. Jin, Two-dimensional local graph embedding discriminant analysis (2DLGEDA) with its application to face and Palm Biometrics, Neurocomputing 73 (2009) 193–203.
[40] Yong Xu, Ge Feng, Yingnan Zhao, One improvement to two-dimensional locality preserving projection method for use with face recognition, Neurocomputing 73 (2009) 245–249.
[41] S. Yan, D. Xu, B. Zhang, and H.-J. Zhang, Graph Embedding: A General Framework for Dimensionality Reduction, IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit. (CVPR'05), vol. 2, pp. 830–837(2005) San Diego.
[42] R.R. Salakhutdinov, and G.E. Hinton, Learning a non-linear embedding by preserving class neighbourhood structure. In AI and Statistics, 2007.
[43] Y. Xu, A. Zhong, J. Yang, D. zhang, LPP solution schemes for use with face recognition, Pattern Recognition 43 (12) (2010) 4165–4176.