Information Sciences 274 (2014) 55–69

Contents lists available at ScienceDirect

Information Sciences

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

Feature extraction using two-dimensional maximum embedding difference

Minghua Wan [a,b,*], Ming Li [a], Guowei Yang [a], Shan Gai [a], Zhong Jin [c]

[a] School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China
[b] School of Information Science and Engineering, Southeast University, Nanjing 210096, China
[c] School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China

ARTICLE INFO

Article history:
Received 30 July 2012
Received in revised form 25 November 2013
Accepted 9 February 2014
Available online 11 March 2014

Keywords:
Feature extraction
Intra-class compactness graph
Margin separability graph
Inter-class separability graph
Difference criterion

ABSTRACT

In this paper we propose a novel method combining graph embedding and difference criterion techniques for image feature extraction, namely two-dimensional maximum embedding difference (2DMED). This method directly extracts the optimal projective vectors from 2D image matrices by simultaneously considering characteristic that is the intra-class compactness graph, the margin graph and inter-class separability graph, respectively. In this method, it is not necessary to convert the image matrix into high-dimensional image vector so that much computational time would be saved. In addition, the proposed method preserves the manifold reconstruction relationships in the low-dimensional subspace. Experimental results on the ORL, Yale face and USPS database show the effectiveness of the proposed method.

1. Introduction

Over past few decades, many dimensionality reduction and feature extraction methods have been developed [1–10] in the fields of image processing, computer vision, and pattern recognition.

In many applications such as face and handwritten character recognition, 2D image matrices are usually transformed into 1D image vectors through column by column or row by row concatenation. On the one hand, the image-to-vector transformation procedure may cause loss of useful structural information embedded in the original images. On the other hand, the resulting 1D image vectors of face images usually lead to a high dimensional image vector space and covariance matrix, which will potentially cause a heavy computational burden. This also usually causes by the small sample size (SSS) problem, which cannot be implemented because of the matrix singularity.

To overcome those problems, Yang et al. proposed two-dimensional principal component analysis (2DPCA) [11,12] which directly extract image features from 2D image matrices and thus the image matrices do not need to be transformed into vectors. Jing et al. [13], Li and Yuan [14], Xiong et al. [15] and Yang et al. [16] extended the idea and presented two-dimensional linear discriminant analysis (2DLDA) using image matrices. Recently, $(2D)^2$PCA [17], $(2D)^2$FLD [18] and $(2D)^2$PCALDA [19] have been proposed, in which the authors investigated two-directional two-dimensional projections along not only in row direction but also in column direction to further reduce the dimension. There are lots of methods successful applying to linear data, such as PCA, LDA and their extensions involve probabilistic principal component analysis (PPCA) [20], mixture

* Corresponding author at: School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China. Tel.: +86 18970856727.
E-mail address: wmh36@sina.com (M. Wan).

http://dx.doi.org/10.1016/j.ins.2014.02.145
0020-0255/

of PCA [21], independent component analysis (ICA) [22], and incremental PCA [23]. However, they may fail to explore the essential structure of data with nonlinear distribution [24,25].

In real world applications, the nonlinearity encountered mostly appears in non-Gaussian or manifold-value data. For non-Gaussian or manifold-value data, we usually deal with it from local patches because non-Gaussian or manifold-value data can be viewed as locally. So, they can be viewed as locally Euclidean [26,27]. Recently, kernel-based techniques and manifold learning-based have been developed to deal with nonlinear data. Two most common kernel-based techniques are kernel principal component analysis (KPCA) [28] and kernel linear discriminant analysis (KLDA) [29], which can be viewed as the kernel versions of PCA and LDA. However, kernel-based methods improve the linear discriminability at the cost of high computational requirements with increasing the dimensions. Furthermore, the way of how to select the most suitable kernel and assign the optimal parameters in kernel techniques is remain unclear. Manifold learning-based techniques are used to find the intrinsic nonlinear structure of data hidden in the high dimensional observation space. He et al. [30] and He and Niyogi [31] proposed locality preserving projections (LPP), which is a linear subspace learning method derived from Laplacian Eigenmaps [32,33]. LPP finds an embedding space that preserves local information and detects the best essential face manifold structure. 2DLPP [34–36] methods are proposed to directly extract the proper features from image matrices based on the locality preserving criterion. More recently, some variant versions of 2DLPP such as two-dimensional local graph embedding discriminant analysis (2DLGEDA) [37] and two-dimensional discriminant locality preserving projection [38] were also proposed to improve the performance of 2DLPP.

However, the computational cost of LPP and its 2D extensions is high because they involve eigen-decomposition of dense matrices. And the possible singularity of within (or intra)-class scatter matrix is statistical properties which are used to obtain the estimates of the covariance matrices in the same class, particularly when the image size is large. Therefore, in this paper we present a two-dimensional maximum embedding difference (2DMED) method for feature extraction. This method does not need to convert image matrix into high-dimensional image vector and avoids the computation of inverse matrix so that additional computation time can be saved to a certain extent. The idea of 2DMED is very similar to marginal Fisher analysis (MFA) [12], but MFA cannot avoid the singularity problem using the Fisher discriminant criterion to find a set of optimal discriminant projection vectors. All of the 2DLDA, 2DLPP, and 2DLGEDA need to compute inverse matrices, while the proposed 2DMED method successfully avoids this computation by the virtue of difference trace. 2DMED saves significant amount of computational and processing time for features extraction and obtains the optimal projective vectors from 2D image matrices by simultaneously considering characteristic that is the intra-class compactness, the margin and inter-class separability using difference criterion techniques [39], respectively. This projection is a transformation of data points from one axis system to another, and is an identical process to axis transformations in graphics.

The proposed method is very similar to two-dimensional maximum scatter difference (2DMSD) [40] and two-dimensional local graph embedding based on maximum margin criterion (2DLGE/MMC) [41] method, both of which adopt the difference of both between (or inter)-class scatter matrix and within (or intra)-class scatter matrix as discriminant criterion. The between (or inter)-class scatter matrix is statistical properties which are used to make the estimates of the covariance matrices in the different class. However, 2DMSD failed to deal with nonlinear data and 2DLGE/MMC did not consider the margin separability characteristic which may degrade recognition rates.

The main novelties of our new method come from the following perspectives:

1. The proposed method uses three graphs to extract the optimal discriminant projective vectors without assuming the particular form of class densities.
2. The proposed method avoids the small sample size problem since it does not need to compute any matrix inversion.
3. The proposed method is capable of having more feature dimensions, which is not limited to (compared to previous upper bound) $c-1$ (where c is the number of different classes), as in LDA.

The rest of the paper is organized as follows: In Section 2 we briefly review the basic 2DMSD, 2DLPP, 2DLGEDA, and 2DLGE/MMC methods, respectively. In Section 3, we introduce the idea, the theoretical analysis, and the outline of the proposed method in detail. In Section 4, experiments on ORL, Yale face databases and USPS database are presented to demonstrate the effectiveness of 2DMED. In addition, we also present overall observations and give some discussions in Section 4. At last, we draw concluding remarks and a discussion of future work in Section 5.

2. Related works

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

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

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

The symbols used in this section are listed in Table 1.

Table 1
List of symbols.

Symbol	Description
N	The number of samples
X	The set of X_i
$m \times n$	The dimensional of X_i
Y_i	The feature matrix of X_i
Ω	The projected matrix
ω_i	The unitary column vector
c	The number of classes
B_i	The label of ith class
$\overline{X}_i$	The class mean matrix of training samples in ith class
N_i	The total number of training samples in ith class
G_b	The between-class scatter matrix
G_w	The within-class scatter matrix
S	The similarity matrix
$W_{ij}^{w}(W_{ij}^{c})$	The intra-class weight
W_{ij}^{b}	The margin weight
W_{ij}^{p}	The inter-class weight
D	The diagonal matrix
$\otimes$	The Kronecher product of matrix
I_n	The identity matrix of order n
L	The Laplacian matrix
λ	The generalized eigenvalue
d	The feature matrix dimension

2.1. Two-dimensional maximum scatter difference (2DMSD) [40]

MMC and 2DMSD are based on the difference between between-class scatter matrix and within-class scatter matrix. However, MMC is obtained the optimal projection subspaces from 1D image vectors whereas 2DMSD can directly extracts the optimal projection subspaces from 2D image matrices. 2DMSD is aimed at preserving maximum 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 by discriminatory criteria as the following:

$$J(\omega) = \arg\max_{\omega}[\omega^T(G_b - G_w)\omega] \tag{2}$$

$$\text{where} \quad G_b = \frac{1}{N}\sum_{i=1}^{c} N_i(\overline{X}_i - \overline{X})(\overline{X}_i - \overline{X})^T \tag{3}$$

$$G_w = \frac{1}{N}\sum_{i=1}^{c}\sum_{X_j \in B_i}(X_j - \overline{X}_i)(X_j - \overline{X}_i)^T \tag{4}$$

where $\overline{X}_i$ is the average image of all the sample images in 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 scatter matrix and G_w is called within-class scatter matrix. $\Omega = [\omega_1, \omega_2, \ldots, \omega_d]$ are exactly the orthogonal generalized eigenvectors corresponding to the first d largest generalized eigenvalues of matrix $G_b - G_w$.

2.2. Two-dimensional Laplacianfaces (2DLPP) [36]

Let $G = \{X, S\}$ denote the complete undirected weighted graph with the vertex-set X and the similarity matrix $S \in R^{N \times N}$. Since each node of 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 to preserve the intrinsic geometry of the data and local structures. The similarity matrix S can be Gaussian weight or uniform weight of Euclidean distance using k-neighborhood or ε-neighborhood, which is defined as:

$$S_{ij} = \begin{cases} 1, & \|X_i - X_j\|^2 < \varepsilon \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

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

$$\min \sum_{i,j}\|Y_i - Y_j\|^2 S_{ij} \tag{6}$$

where $Y_i = \omega^T X_i$ and $\|\cdot\|$ represents the L_2 norm. After some matrix analysis steps, the minimization problem of Eq. (6) becomes:

$$\begin{aligned} &\arg\min_{\omega} \ \omega^T X (L \otimes I_n) X^T \omega \\ &\text{s.t.} \quad \omega^T X (D \otimes I_n) X^T \omega = 1 \end{aligned} \tag{7}$$

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

The optimal d projection vectors that minimize the objective function by the minimum eigenvalue solution to the generalized eigenvalue problem:

$$X(L \otimes I_n)X^T \omega = \lambda X(D \otimes I_n)X^T \omega \tag{8}$$

2.3. Two-dimensional local graph embedding discriminant analysis (2DLGEDA) [37]

2DLGEDA is proposed as a supervised extension of 2DLPP which can directly work on 2D image matrices.

The goal of 2DLGEDA is to preserve graphs to characterize 2D image intra-class compactness and margin separability. 2D image intra-class scatter matrix is characterized in the intrinsic graph W_{ij}^w by the term:

$$S_w = \sum_{i=1}^{N} \sum_{j=1}^{N} \|Y_i - Y_j\|^2 W_{ij}^w = 2\omega^T X (L^w \otimes I_n) X^T \omega \tag{9}$$

where

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

where D^w is the diagonal matrix whose entries are column or row sums of W^w, I_n is an identity matrix of order n, operator $\otimes$ is the Kronecker product of matrix, $N_{k_w}^+(X_i)$ indicates the samples in the k_w nearest neighbors of X_i in the same class, and ω denotes the projection vector, $L^w = D^w - W^w$.

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

$$S_b = \sum_{i=1}^{N} \sum_{j=1}^{N} \|Y_i - Y_j\|^2 W_{ij}^b = 2\omega^T X (L^b \otimes I_n) X^T \omega \tag{10}$$

where

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

where $P_{k_b}(\pi_t)$ is a set of data pairs in the k_b nearest pairs among a set $\{(i,j) | i \in \pi_t, j \notin \pi_t\}$, and π_t denotes the index set of tth class and t is varied from 1 to c. D^b is the 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 by solving the generalized eigen-equation:

$$X^T (L^b \otimes I_n) X \omega = \lambda X^T (L^w \otimes I_n) X \omega \tag{11}$$

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 first d larger eigenvalues.

2.4. Two-dimensional local graph embedding based on maximum margin criterion (2DLGE/MMC) [41]

2DLGE/MMC is adopted the difference of both intra-class compactness and the inter-class separability as discriminant criterion. 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 intra-class scatter matrix and the inter-class scatter matrix as follows:

$$\begin{aligned} J(\omega) &= \min \left(\sum_{j=1}^{k_1} \|Y_i - Y_i^j\|^2 - \alpha \sum_{q=1}^{k_2} \|Y_i - Y_{iq}\|^2 \right) = \min \left(\sum_{i=1}^{N} \sum_{j=1}^{N} \|Y_i - Y_j\|^2 W_{ij}^c - \alpha \sum_{i=1}^{N} \sum_{j=1}^{N} \|Y_i - Y_j\|^2 W_{ij}^p \right) \\ &= tr(\omega^T X (L^c \otimes I_n) X^T \omega - \alpha \omega^T X (L^p \otimes I_n) X^T \omega) \end{aligned} \tag{12}$$

where α is an adjustable parameter which balances $\omega^T X (L^c \otimes I_n) X^T \omega$ and $\omega^T X (L^p \otimes I_n) X^T \omega$.

Table 2
The main characteristics of different methods.

Characteristics	Method				
	2DMSD	2DLPP	2DLGEDA	2DLGE/MMC	2DMED
Criterion	Difference	Fisher	Fisher	Difference	Difference
Data	Linear	Nonlinear	Nonlinear	Nonlinear	Nonlinear
Image space	Global	Local	Local	Local	Local
Class information	Supervised	Unsupervised	Supervised	Supervised	Supervised
Intra-class graph	None	S_{ij}	W_{ij}^{w}	W_{ij}^{c}	W_{ij}^{c}
Margin graph	None	None	W_{ij}^{b}	None	W_{ij}^{o}
Inter-class graph	None	None	None	W_{ij}^{p}	W_{ij}^{p}

$$\text{where}\quad W_{ij}^{c} = \begin{cases} 1, & \text{if } X_j \text{ is in the } K_c \text{ nearest from same class of } X_i \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

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

Then we can easily find that Ω consists of the eigenvectors associated with top d eigenvalues of the above eigen-equation.

At last, the main characteristics of different methods are listed in Table 2.

Compared with other methods, 2DMED has several advantages including avoiding the SSS problem and reducing the computational costs by difference criteria. On the other hand, 2DMED is still a local and supervised method. At last, 2DMED fully captures the geometrical structure on data manifolds by defining three graphs W_{ij}^{c}, W_{ij}^{o} and W_{ij}^{p}.

3. Two-dimensional maximum embedding difference (2DMED)

3.1. Foundations and problem statement

In Fig. 1, suppose there are two pattern classes and we preserve graphs to characterize the adjacency relationships of intra-class compactness, inter-class separability and margin-class separability, respectively.

Graph-I has the graph characteristic of intra-class compactness, where a vertex pair is connected if one vertex is among the k_1-nearest neighbors of the other and the elements of the pair belong to the same class.

Graph-II has the graph characteristic of inter-class separability, where the k_2-nearest vertex pairs in a way that one in-class element and the other out-of-class element are connected.

Graph-III has the graph characteristic of margin-class separability, where the k_3 marginal points from different classes are connected.

2DLPP characterizes intra-class compactness by graph-I, which finds an embedding that preserves local information and detects the intrinsic image manifold structure. However, 2DLPP does not consider the adjacency relationships between different classes such as graph-IIand graph-III.

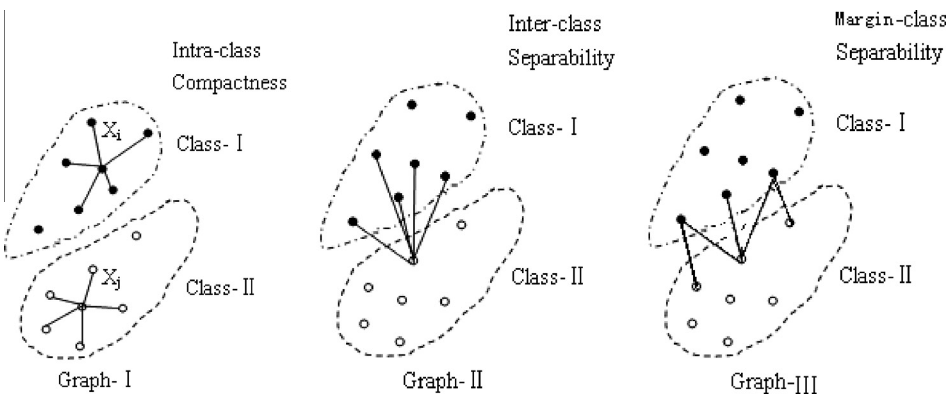

Fig. 1. The adjacency relationships graph characteristic of intra-class compactness, inter-class separability and margin-class separability, respectively.

To overcome this problem, 2DLGE/MMC and 2DLGEDA are proposed to characterize inter-class separability and margin-class separability according to graph-IIand graph-III, respectively. 2DLGE/MMC and 2DLGEDA are supervised method while 2DLPP is unsupervised method. However, 2DLGE/MMC is to use the difference criterion while 2DLGEDA is to adopt the Fisher criterion.

However, 2DLGE/MMC does not consider the margin-class separability and 2DLGEDA does not consider inter-class separability which might cause degradation of recognition rate. Therefore, 2DMED not only contains inter-class separability, but also considers margin-class separability. After the linear transformation, the considered points in the same class are placed as close as possible, while those between different classes are placed as far as possible. We formulate this problem as a constrained optimization problem, in which the global optimum can be effectively and efficiently obtained. Compared with 2DLPP, 2DLGEDA and 2DLGE/MMC, our method is able to extract more discriminative features by considering three graphs.

3.2. The proposed scheme and algorithm

Let $G = \{X, W\}$ denote the complete undirected weighted graph with the vertex-set X and the similarity matrix $W \in R^{N\times N}$. It is easy to see that each element of W measures the similarity of a pair of vertices. For supervised learning problems, the class label for each sample X_i is assumed to be π_t, $t \in \{1,2,\ldots,c\}$.

The diagonal matrix D and the Laplacian matrix L of a graph G are defined as:

$$L = D - W, \quad D = diag(D_{11}, \ldots, D_{NN}), \quad D_{ii} = \sum_{j\neq i} W_{ij}, \quad \forall i \tag{15}$$

We preserve graph characteristic to calculate similarity matrix in Fig. 1, and then we propose the 2DMED algorithm for feature extraction.

The similarity matrix in the intra-class compactness graph can be defined as follows:

$$W^c_{ij} = \begin{cases} 1, & \text{if } (i,j) \in \pi_c, \text{ and } i \in N^+_{K_c}(j) \text{ or } j \in N^+_{K_c}(i) \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

where $N^+_{K_c}(i)$ indicates the index set of the K_c nearest neighbors of X_i in the same class.

The similarity matrix in the margin separability graph can be defined as follows:

$$W^p_{ij} = \begin{cases} 1, & \text{if } (i,j) \in P_{K_p}(\pi_t) \text{ or } (j,i) \in P_{K_p}(\pi_t) \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

where $P_{K_p}(c)$ is a set of data pairs that are in the K_p nearest pairs among the set $\{(i,j) | i \in \pi_t, j \notin \pi_t\}$.

The similarity matrix in the inter-class separability graph can be defined as follows:

$$W^o_{ij} = \begin{cases} 1, & \text{if } i \in N^-_{K_o}(j) \text{ or } j \in N^-_{K_o}(i) \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

where $N^-_{K_o}(i)$ indicates the index set out of the K_o nearest neighbors of X_i or in the K_o nearest neighbors of X_i but not in the same class.

Due to introducing the similarity matrix W^c_{ij}, the intra-class scatter matrix from the intra-class graph can be expressed to:

$$J_c(\omega) = \sum_i \sum_j \|Y_i - Y_j\|^2 W^c_{ij} = \sum_i \sum_j \|\omega^T x_i - \omega^T x_j\|^2 W^c_{ij} = 2\omega^T X((D^c - W^c) \otimes I_n) X^T \omega = 2\omega^T X(L^c \otimes I_n) X^T \omega \tag{19}$$

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

After characterizing the intra-class scatter matrix, the margin separability scatter matrix can be characterized by the following expression:

$$J_p(\omega) = \sum_i \sum_j \|Y_i - Y_j\|^2 W^p_{ij} = \sum_i \sum_j \|\omega^T x_i - \omega^T x_j\|^2 W^p_{ij} = 2\omega^T X((D^p - W^p) \otimes I_n) X^T \omega = 2\omega^T X(L^p \otimes I_n) X^T \omega \tag{20}$$

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

At last, the inter-class scatter matrix can be characterized as following:

$$J_o(\omega) = \sum_i \sum_j \|Y_i - Y_j\|^2 W^o_{ij} = \sum_i \sum_j \|\omega^T x_i - \omega^T x_j\|^2 W^o_{ij} = 2\omega^T X((D^o - W^o) \otimes I_n) X^T \omega = 2\omega^T X(L^o \otimes I_n) X^T \omega \tag{21}$$

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

After the intra-class scatter matrix, the margin scatter matrix and the inter-class scatter matrix have been constructed; an optimization objective function can be devised to maximize the difference between the margin scatter matrix, the inter-class scatter matrix and the intra-class scatter matrix. That is:

$$J(\omega) = \max\ tr[(1-\alpha)(J_p + J_o) - \alpha J_c] \tag{22}$$

where α $(0 < \alpha < 1)$ is an adjustable parameter to balance $J_c(\omega)$ and $J_p(\omega)$.

Then we can easily find that Ω consists of the eigenvectors associated with d top eigenvalues by solving the above generalized eigen-equation.

After the training by 2DMED, the feature matrix of each image X_i and the transformation matrix are obtained. Feature matrices are transformed to 1D image vectors, and then a nearest-neighbor classifier can be used for classification.

Given two images X_1, X_2 represented by 2DMED feature vectors $Y_1 = (y_1^1, y_1^2, \ldots, y_1^d)$ and $Y_2 = (y_2^1, y_2^2, \ldots, y_2^d)$, then the dissimilarity $d(Y_1, Y_2)$ is defined as:

$$d(Y_1, Y_2) = \sum_{k=1}^{d} \|Y_1^k - Y_2^k\|^2 \tag{23}$$

If feature matrices of training images are $Y_1, Y_2, \ldots, Y_N$ (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 $Y \in \pi_l$.

The proposed feature extraction algorithm can be summarized as follows:

Step 1: Constructing the similarity matrix W_{ij}^c, W_{ij}^p and W_{ij}^o using Eq. (16)–(18).
Step 2: Calculating the intra-class scatter matrix $J_c(\omega)$, the margin scatter matrix $J_p(\omega)$ and the inter-class scatter matrix $J_o(\omega)$ using Eqs. (19)–(21), respectively.
Step 3: Extracting the sample feature using Eq. (22).
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. (23).

3.3. Comparisons of computation complexity and space complexity

As it is mentioned before Section 1, the computational complexity of MFA is high. Supposed the image size is $n \times n$, the computational cost of MFA with k-NN penalty graph is:
$O(m^2 n + m^2 \log m + dn^4 + n^4 m + n^2 m^2 + \frac{1}{4}(k(k-1))^2 n + \frac{1}{4}(k(k-1))^2 \log(\frac{1}{2}k(k-1)))$. $O(m^2 n)$ is used to calculate the pair-wise distance between m samples with n-dimensional features and $(m^2 \log m)$ is used for k-nearest neighbors finding for all the m samples. $O\left(\frac{1}{4}(k(k-1))^2 n\right)$ is used for calculating the pairwise distance in k-nearest neighbors samples and $O\left(\frac{1}{4}(k(k-1))^2 \log(\frac{1}{2}k(k-1))\right)$ for finding all the local k-nearest neighbors data pairs. Computing the scatter matrix XL^pX^T needs $O(n^4 m + n^2 m^2)$ in MFA. $O(dn^4)$ is used to compute the first d generalized eigenvectors. For the proposed 2DMED, the computational complexity is $O(m^2 n + m^2 \log m + dn^2 + n^2 m^2 + \frac{1}{2}(k(k-1))^2 n + \frac{1}{2}(k(k-1))^2 \log(\frac{1}{2}k(k-1)))$ in total. Computing the scatter matrix $X(L^p \otimes I_n)X^T$ needs $O(n^2 m^2)$ in 2DMED. $O(dn^2)$ is used to compute the first d generalized eigenvectors of 2DMED. Usually, $O(dn^4 + n^4 m + n^2 m^2) \gg O(dn^2 + n^2 m^2)$. Thus, 2DMED's computational complexity is far less than that of MFA. Particularly, when n is relatively large (when compared with m), MFA is significantly time consuming. Moreover, the space complexity needs $O(n^4)$ for the vector-based MFA. However, the 2DMED algorithm framework can work on the 2D scatter matrix and thus only needs $O(n^2)$. Therefore, 2DMED greatly saves memory cost, which is the same as the other 2D based projection methods such as 2DPCA, 2DLDA and 2DLPP.

4. Experiments

To evaluate the proposed 2DMED algorithm, we compare it with the PCA, LDA, MMC, LPP, MFA, Wavelet + ICA + LDA, 2DPCA, 2DLDA, 2DMSD, 2DLPP, 2DLGEDA and 2DLGE/MMC algorithm in three databases: on the face recognition (on the ORL and Yale face databases) and handwriting digital recognition (on the USPS database). When the projection vectors are computed from the training part, all the images including the training part and the test part are projected to feature space. Euclidean distance and nearest neighborhood classifier are used in all the experiments. The experiments are carried out on a PC (CPU: P4 2.8 GHz, RAM: 1024 MB).

4.1. Data corpora

The ORL database (http://www.uk.research.att.com/facedatabase.html) contains images from 40 individuals, each providing 10 different images. The facial expressions and facial details (glasses or no glasses) also vary. The images are taken with a tolerance for some tilting and rotation of the face of up to 20 degrees. Moreover, there is also some variation in the scale of up to about 10%. All images are normalized to a resolution of 56 × 46. Fig. 2 shows sample images of one person from ORL 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,

Fig. 2. Sample images of one person in the ORL face database.

Fig. 3. Sample images of one person in the Yale database.

each image is manually cropped and resized to 50 × 40 pixels. Fig. 3 shows sample images of one person on the Yale database.

The USPS handwriting digital data includes 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 database.

4.2. Optimization parameters

How to select the weight parameter α is an important problem in feature extraction. To find how the weight parameter α affects the recognition performance, in the first experiment, we change α from 0.1 to 0.9 with step 0.1 on the ORL, Yale face databases and USPS database, respectively. Each database is divided 50% for training, and 25% each for validation and testing [42]. So, the first 4, 5 and 40 samples of each class randomly selected to compose the training set, the second 3, 3 and 30 samples of each class compose the validation set, and the remaining 3, 3 and 30 samples form the test set on the ORL, Yale face databases and USPS database, respectively. And this experiment run 10 times. Figs. 5–7 displays the maximal average recognition rates (%) of 2DMED versus the dimensions when the training set for training and the validation set for testing with varied α on the three databases.

From Figs. 5–7, it can be found that the effectiveness of the 2DMED algorithm is sensitive to the value of the weight parameter α and the best recognition rate obtained with 2DMED is 95.52%, 94.96% and 90.37% when $\alpha = 0.6$, $\alpha = 0.3$ and $\alpha = 0.2$, respectively. This indicates that the proportion of the intra-class compactness matrix, the margin separability matrix and the inter-class separability matrix are different. So, in the later experiment, the value of adjustable weight parameter α is taken to be $\alpha = 0.6$, $\alpha = 0.3$ and $\alpha = 0.4$ on the three databases, respectively.

Based on these parameters, we obtain the maximal average recognition rates of the proposed method on the test set, as listed in Table 3.

4.3. Experimental results and analysis

In the second experiment, we also test the impact of the nearest-neighbor parameter k (K_p and K_o) on the recognition performance. If the value of k is too small, it is very difficult to preserve the topologic structure in the low-dimensional feature

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

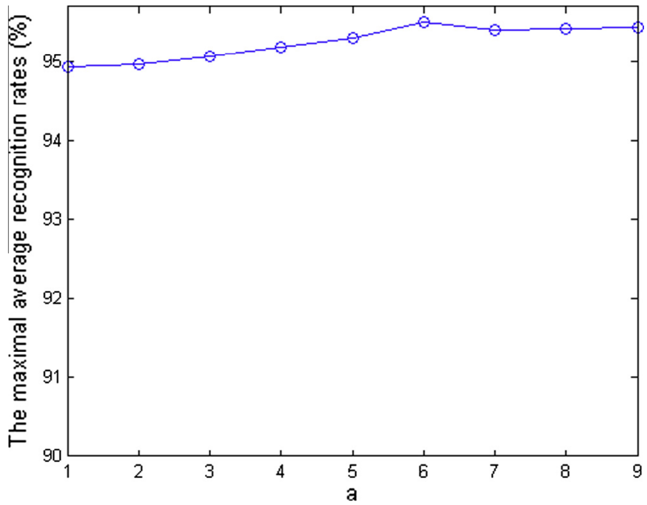

Fig. 5. The maximal average recognition rates (%) of 2DMED versus the dimensions when the training set for training on the ORL face database with varied α.

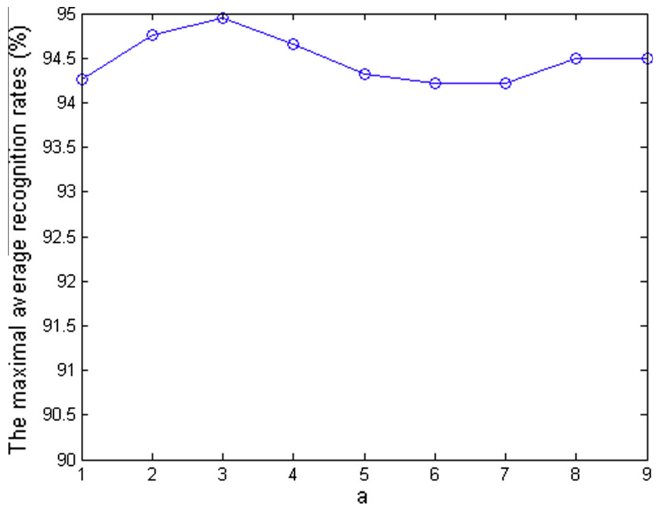

Fig. 6. The maximal average recognition rates (%) of 2DMED versus the dimensions when the training set for training on the Yale face database with varied α.

space. On the contrary, if the value of k is too large, it is very difficult to depict the assumption of local linearity in the high-dimensional feature space.

The K_c-nearest-neighbor parameter in graph embedding method is chosen as $K_c = l - 1$ where they can get best recognition rates and within-class samples are well clustered in the observation space [30]. We change the value of k (K_p and K_o from 2 to 20 with step 2) on the ORL, Yale face databases and the value of k (K_p and K_o from 5 to 50 with step 5) on the USPS database.

Now, we test the recognition performances of each method on three databases, respectively. In these experiments, l images (l varies from 2 to 6 on the ORL, Yale face database, and l varies from 20 to 60 on the USPS database) are randomly selected from the image gallery of each individual to form the training sample set. The remaining images are used for testing. For each l, we independently run 10 times.

For feature extraction, we used PCA, LDA, MMC, LPP, MFA, Wavelet + ICA + LDA, 2DPCA, 2DLDA, 2DMSD, 2DLPP, 2DLGEDA, 2DLGE/MMC and the proposed 2DMED algorithm, respectively. In the PCA phase of each method, we keep 90% image energy. The maximal average recognition rate of each method and the corresponding dimension are given in Tables 4–6 when the l samples per class are randomly selected for training and the remaining images for test, respectively. The maximal average recognition rate is obtained by running 10 times averaging.

Figs. 8–10 showed the average recognition rates (%) of 2DPCA, 2DLDA, 2DLPP, 2DMSD, 2DLGEDA, 2DLGE/MMC and 2DMED versus the dimensions when the 2 images per person are randomly selected for training on the ORL, Yale face

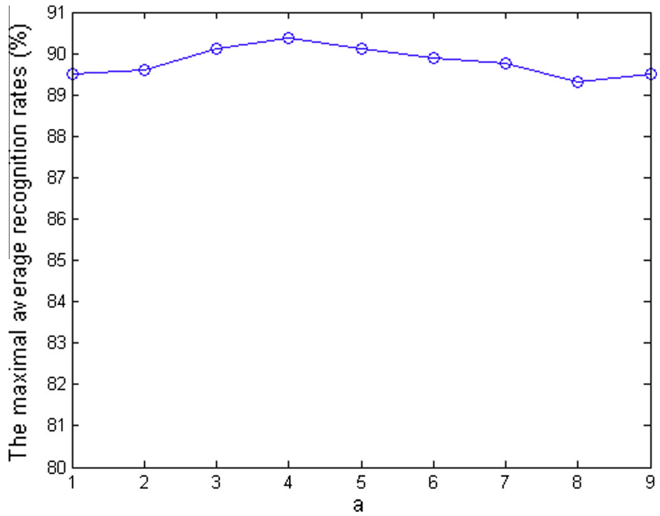

Fig. 7. The maximal average recognition rates (%) of 2DMED versus the dimensions when the training set for training on the USPS database with varied α.

Table 3
The maximal average recognition rates (%) of 2DMED on the test set of three databases.

Method	Database		
	ORL $l = 4$	Yale $l = 5$	USPS $l = 40$
2DMED %(dim)	96.30(56 × 2)	94.96(50 × 22)	89.85(16 × 4)

Table 4
The maximal average recognition rates (%) of each method on the ORL face database.

Method		Number of training images				
		$l = 2$	$l = 3$	$l = 4$	$l = 5$	$l = 6$
PCA	%(dim)	74.86(46)	82.27(46)	84.98(46)	86.71(46)	87.39(46)
LDA	%(dim)	77.42(38)	85.09(38)	86.17(38)	87.23(38)	89.38(38)
MMC	%(dim)	78.78(40)	85.44(40)	86.85(38)	88.32(38)	91.50(36)
LPP	%(dim)	72.05(48)	81.78(46)	87.42(36)	90.82(34)	93.21(32)
MFA	%(dim)	80.54(32)	83.56(36)	88.66(36)	91.52(46)	94.31(36)
Wavelet + ICA + LDA (3-level A sub-band)	%(dim)	83.82(32)	86.94(32)	89.75(32)	92.32(42)	95.80(32)
2DPCA	%(dim)	82.88(56 × 46)	86.79(56 × 44)	88.75(56 × 46)	89.68(56 × 44)	95.63(56 × 44)
2DLDA	%(dim)	86.25(56 × 4)	89.39(56 × 2)	91.87(56 × 4)	92.98(56 × 4)	97.41(56 × 4)
2DMSD	%(dim)	87.43(56 × 8)	88.21(56 × 8)	91.67(56 × 6)	94.00(56 × 6)	97.50(56 × 8)
2DLPP	%(dim)	83.93(56 × 12)	87.78(56 × 14)	88.86(56 × 16)	91.35(56 × 16)	96.83(56 × 12)
2DLGEDA	%(dim)	89.21 (56 × 2)	91.79 (56 × 4)	93.33(56 × 2)	94.00(56 × 4)	97.12(56 × 4)
2DLGE/MMC	%(dim)	90.36 (56 × 10)	92.18 (56 × 14)	93.76(56 × 12)	94.58(56 × 14)	97.60(56 × 14)
2DMED	%(dim)	**92.50** (56 × 10)	**93.21** (56 × 16)	**94.17**(56 × 4)	**95.50**(56 × 12)	**98.75**(56 × 8)

The bold values denote the best results.

database and the 20 images per class are used for training on the USPS database, respectively. From three figures, it is observed that the proposed method outperformed 2DPCA, 2DLDA, 2DMSD, 2DLPP, 2DLGEDA and 2DLGE/MMC methods.

In addition, the average CPU time consumed for training, test and classification, and the maximal average recognition rates of the foregoing each method are given in Tables 7–9. 2DMED achieves its maximal average recognition rate, and it needs less CPU time except 2DLGE/MMC when compared to other methods on three databases, respectively.

4.4. Overall observations and discussions

According to the experiments of three databases, we have several interesting observations:

(1) Different from PCA, LDA, Wavelet + ICA + LDA, MMC, 2DPCA, 2DLDA, and 2DMSD, 2DMED attempt to preserve the local geometric structure, while the others aim to discover the global Euclidean structure. Manifold learning

Table 5
The maximal average recognition rates (%) of each method on the Yale face database.

Method		Number of training images				
		$l = 2$	$l = 3$	$l = 4$	$l = 5$	$l = 6$
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)
MMC	%(dim)	81.29(21)	83.72(14)	86.99(14)	87.20(14)	88.29(14)
LPP	%(dim)	81.45(22)	85.97(24)	88.57(21)	89.00(18)	90.40(21)
MFA	%(dim)	83.87(12)	88.25(16)	89.48(18)	91.89(20)	92.95(24)
Wavelet + ICA + LDA (3-level A sub-band)	%(dim)	85.95(14)	89.75(16)	91.90(16)	92.60(16)	94.15(12)
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)
2DMSD	%(dim)	89.57(50 × 32)	90.42 (50 × 30)	93.60(50 × 32)	93.95(50 × 32)	94.25(50 × 30)
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)
2DLGE/MMC	%(dim)	90.60(50 × 36)	92.08(50 × 36)	94.23(50 × 38)	94.49(50 × 37)	95.08(50 × 37)
2DMED	%(dim)	**91.20**(50 × 25)	**92.68**(50 × 25)	**93.84** (50 × 28)	**95.49**(50 × 25)	**96.80** (50 × 28)

The bold values denote the best results.

Table 6
The maximal average recognition rates (%) of each on the USPS database.

Method		Number of training images				
		$l = 20$	$l = 30$	$l = 40$	$l = 50$	$l = 60$
PCA	%(dim)	80.88(20)	84.56(20)	86.72(29)	87.96(26)	88.90(27)
LDA	%(dim)	82.72 (7)	85.83 (9)	86.80 (8)	88.00 (9)	88.57 (9)
MMC	%(dim)	79.85(30)	83.74(27)	86.43(27)	87.88(27)	89.40(27)
LPP	%(dim)	78.93(25)	82.75(14)	85.70 (29)	86.78(13)	88.82(17)
MFA	%(dim)	80.54(32)	83.56(36)	86.66(36)	88.52(46)	89.31(36)
Wavelet + ICA + LDA (3-level A sub-band)	%(dim)	81.12(36)	85.30(36)	86.80(38)	89.06 (36)	91.70(36)
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)
2DMSD	%(dim)	81.64 (16 × 2)	82.53 (16 × 4)	85.16(16 × 4)	87.95(16 × 5)	88.32(16 × 4)
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)
2DLGE/MMC	%(dim)	82.34(16 × 3)	85.93(16 × 3)	88.45(16 × 3)	89.80(16 × 3)	90.25(16 × 3)
2DMED	%(dim)	**83.30**(16 × 4)	**86.10**(16 × 4)	**89.36**(16 × 4)	**90.38**(16 × 4)	**91.82**(16 × 4)

The bold values denote the best results.

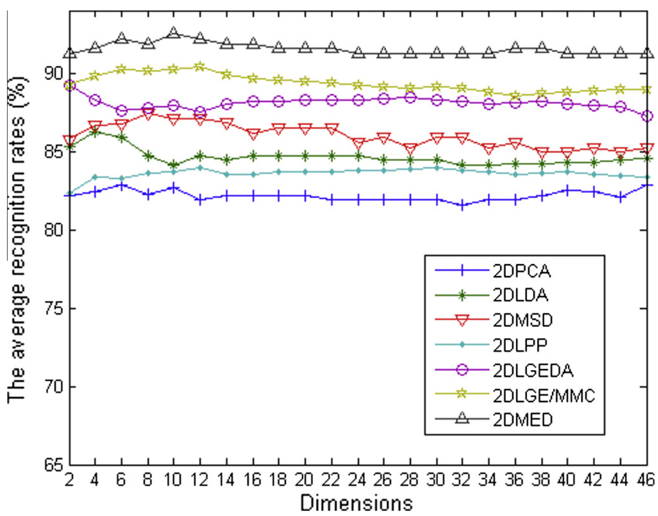

Fig. 8. The average recognition rates (%) of 2DPCA, 2DLDA, 2DLPP, 2DMSD, 2DLGEDA, 2DLGE/MMC and 2DMED versus the dimensions when the 2 images per person were randomly selected for training on the ORL face database. The dimension here is the number of eigenvectors.

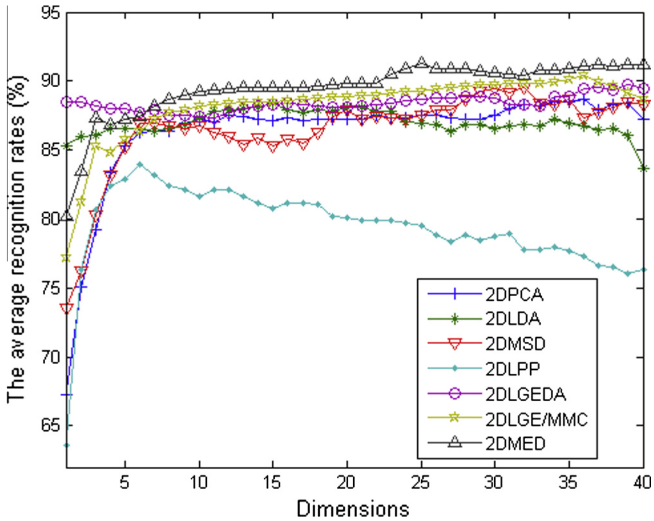

Fig. 9. The average recognition rates (%) of 2DPCA, 2DLDA, 2DLPP, 2DMSD, 2DLGEDA, 2DLGE/MMC and 2DMED versus the dimensions when the 2 images per person were used for training on the Yale face database. The dimension here is the number of eigenvectors.

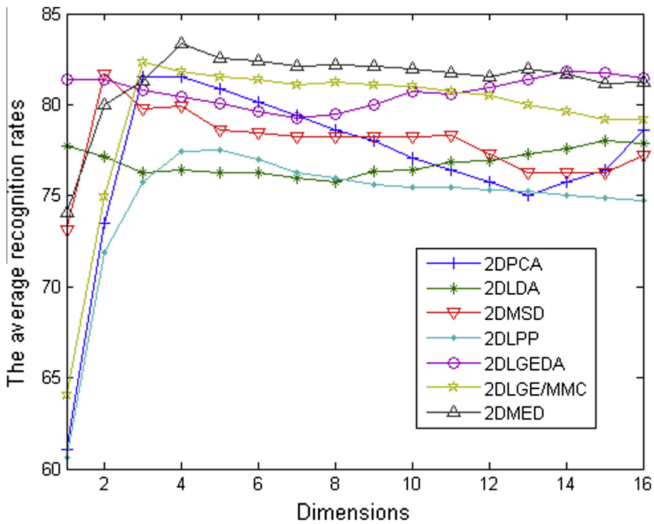

Fig. 10. The average recognition rates (%) of 2DPCA, 2DLDA, 2DLPP, 2DLGEDA, 2DLGE/MMC and 2DMED versus the dimensions when the 20 images per class were used for training on the USPS database. The dimension here is the number of eigenvectors.

algorithms based on local structure are superior to the methods based on global structure. In addition, the maximal average recognition rates of 2DMED is significantly higher than that of other manifold learning algorithms, which are shown in Tables 4–6.

(2) The average recognition rate (%) of 2DMED for a given number of dimensions is always higher than other six methods, which are shown in Figs. 8–10. The proposed method preserves useful discriminant information. In Tables 7–9, compared to other methods, 2DMED needs less CPU time for training, testing, and classification. It is only a little more on CPU time consuming compared to 2DLGE/MMC. 2D methods directly extract image features from 2D image matrices and thus the image matrices do not need to be transformed into vectors. 2DMED and 2DLGE/MMC avoid computing inverse matrices, but 2DMED needs a few CPU time more than 2DLGE/MMC. It is because that 2DMED builds three graphs.

(3) Our 2DMED method obtains the maximal average recognition rate in all the experimental cases, which shows that 2DMED is capable of handling SSS problem. The decent performance of the proposed method also demonstrates that 2DMED is more effective than other methods in extracting and representing facial features for face and USPS

Table 7
The average CPU time (s) consumed for training, test and classification, and the maximal average recognition rates (%) when the 2 images per person were randomly selected for training on the ORL face database.

Methods	Recognition rate (%)	Dim	CPU time (s)
PCA	74.86	(46)	0.281
LDA	77.42	(38)	0.360
LPP	78.78	(40)	0.350
MMC	72.05	(48)	0.316
MFA	80.54	(32)	0.289
Wavelet + ICA + LDA (3-level A sub-band)	83.82	(32)	0.330
2DPCA	82.88	(56 × 46)	0.265
2DLDA	86.25	(56 × 4)	0.258
2DMSD	87.43	(56 × 8)	0.244
2DLPP	83.93	(56 × 12)	0.250
2DLGEDA	89.21	(56 × 2)	0.255
2DLGE/MMC	90.36	(56 × 10)	**0.224**
2DMED	**92.50**	(56 × 12)	**0.248**

The bold values denote the best results.

Table 8
The average CPU time (s) consumed for training, test and classification, and the maximal average recognition rates (%) when the 2 images per person were used for training on the Yale face database.

Methods	Recognition rate (%)	Dim	CPU time (s)
PCA	78.49	(29)	0.155
LDA	81.93	(14)	0.151
LPP	81.45	(22)	0.152
MMC	81.29	(21)	0.140
MFA	83.87	(12)	0.0873
Wavelet + ICA + LDA (3-level A sub-band)	85.95	(14)	0.0930
2DPCA	88.67	(50 × 36)	0.0606
2DLDA	88.37	(50 × 15)	0.0589
2DMSD	89.57	(50 × 32)	0.0542
2DLPP	83.93	(50 × 6)	0.0592
2DLGEDA	89.70	(50 × 39)	0.0587
2DLGE/MMC	90.90	(50 × 36)	**0.0559**
2DMED	**91.20**	(50 × 25)	**0.0565**

The bold values denote the best results.

Table 9
The average CPU time (s) consumed for training, test and classification, and the maximal average recognition rates (%) when the 20 images per class were used for training on the USPS database.

Methods	Recognition rate (%)	Dim	CPU time (s)
PCA	80.88	(20)	1.542
LDA	82.72	(7)	1.504
LPP	78.93	(25)	1.478
MMC	79.85	(30)	1.014
MFA	80.54	(32)	0.723
Wavelet + ICA + LDA (3-level A sub-band)	81.12	(36)	0.816
2DPCA	81.56	(16 × 3)	0.610
2DLDA	78.01	(16 × 15)	0.598
2DMSD	81.75	(16 × 14)	0.505
2DLPP	77.51	(16 × 5)	0.587
2DLGEDA	81.75	(16 × 14)	0.601
2DLGE/MMC	82.34	(16 × 3)	**0.548**
2DMED	**83.30**	(16 × 4)	**0.554**

The bold values denote the best results.

recognition over the variation of lighting, facial expressions, and pose. The reason may be that the proposed 2DMED makes full use of the discriminant information and characteristic in the intra-class compactness graph, the margin separability graph, and the inter-class separability graph, and is able to capture the geometrical structure of data manifold.

5. Conclusions

In this paper, we propose a novel method for feature extraction and recognition, namely the two-dimensional maximum embedding difference (2DMED). It can directly extract the optimal projective vectors from 2D image matrices by simultaneously considering the graph characteristic of intra-class compactness, the margin separability and the inter-class separability combined with difference criterion techniques. Comprehensive comparison and extensive experiments show that 2DMED has the competitive performance against other algorithms such as PCA, LDA, MMC, LPP, MFA, Wavelet + ICA + LDA, 2DPCA, 2DLDA, 2DMSD, 2DLPP, 2DLGEDA and 2DLGE/MMC.

Acknowledgements

This work is partially supported by the National Science Foundation of China under Grant Nos. 61203243, 61202319, 61201439, 61262019, 61362031, the China Postdoctoral Science Foundation under Grant Nos. 2013M530223, 2013M53024, the National Science Foundation of Jiangxi Provincial under Grant Nos. 20114BAB201034, 20122BAB211025, Jiangsu Provincial Postdoctoral Science Foundation under Grant No. 1301095C and Youth Foundation of Jiangxi Provincial Department of Education No. GJJ12459.

References

[1] P.N. Belhumeur, J.P. Hespanha, D.J. Kriengman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[2] Y.F. Guo, J.Y. Yang, Feature extraction method based on the generalised fisher discriminant criterion and facial recognition, Pattern Anal. Appl. 4 (1) (2001) 61–66.
[3] Weh Hu, O. Farooq, S. Datta, Wavelet based sub-space features for face recognition, Proc. of International Congress on Image and Signal Processing, 27–30 May 2008, vol. 3, IEEE Computer Society, Sanya, Hainan, China, 2008, pp. 426–430.
[4] Turk M.A., Pentland A.P., 1991. Face recognition using eigenfaces. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'91), IEEE, pp. 586-591.
[5] J. Yang, D. Zhang, J.Y. Yang, A generalised KL expansion method which can deal with small sample size and high-dimensional problems, Pattern Anal. Appl. 6 (1) (2003) 47–54.
[6] Jie Xu, Jian Yang, Zhihui Lai, K-local hyperplane distance nearest neighbor classifier oriented local discriminant analysis, Inform. Sci. 232 (20) (2013) 11–26.
[7] Omer Boehm, David R. Hardoon, Larry M. Manevitz, Classifying cognitive states of brain activity via one-class neural networks with feature selection by genetic algorithms, Int. J. Mach. Learn. Cybernet. 2 (3) (2011) 125–134.
[8] Alok Sharma, Seiya Imoto, Satoru Miyano, Vandana Sharma, Null space based feature selection method for gene expression data, Int. J. Mach. Learn. Cybernet. 3 (4) (2012) 269–276.
[9] Thomas Rückstieß, Christian Osendorfer, Patrick van der Smagt, Minimizing data consumption with sequential online feature selection, Int. J. Mach. Learn. Cybernet. 4 (3) (2013) 235–243.
[10] Yong Xu, Qi Zhu, Zizhu Fan, David Zhang, Jianxun Mi, Zhihui Lai, Using the idea of the sparse representation to perform coarse-to-fine face recognition, Inform. Sci. 238 (20) (2013) 138–148.
[11] 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. Intell. 26 (1) (2004) 131–137.
[12] 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.
[13] X.Y. Jing, H.S. Wong, D. Zhang, Face recognition based on 2D Fisherface approach, Pattern Recognit. 39 (4) (2006) 707–710.
[14] M. Li, B. Yuan, 2D-LDA: a statistical linear discriminant analysis for image matrix, Pattern Recognit. Lett. 26 (5) (2005) 527–532.
[15] H.L. Xiong, M.N.S. Swanmy, M.O. Ahmad, Two-dimensional FLD for face recognition, Pattern Recognit. 38 (7) (2005) 1121–1124.
[16] J. Yang, D. Zhang, Y. Xu, J.J. Yang, Two-dimensional discriminant transform for face recognition, Pattern Recognit. 38 (7) (2005) 1125–1129.
[17] D.Q. Zhang, Z.H. Zhou, $(2D)^2$PCA: two-directional two-dimensional PCA for efficient face representation and recognition, Neurocomputing 69 (1) (2005) 224–231.
[18] P. Nagabhushan, D.S. Guru, B.H. Shekar, (2D)2FLD: an efficient approach for appearance based object recognition, Neurocomputing 69 (7) (2006) 934–940.
[19] Y.F. Qi, J.S. Zhang, $(2D)^2$PCALDA: an efficient approach for face recognition, Appl. Math. Comput. 213 (1) (2009) 1–7.
[20] Michael E. Tipping, Christopher M. Bishop, Probabilistic principal component analysis, J. Roy. Stat. Soc.: Ser. B (Stat. Methodol.) 61 (3) (1999) 611–622.
[21] T. Yoshioka, R. Morioka, K. Kobayashi, et al, Clustering of gene expression data by mixture of PCA models, in: Artificial Neural Networks—ICANN 2002, Springer, Berlin Heidelberg, 2002, pp. 522–527.
[22] Aapo Hyvärinen, Erkki Oja, Independent component analysis: algorithms and applications, Neural Networks 13 (4) (2000) 411–430.
[23] M. Artac, M. Jogan, A. Leonardis, Incremental PCA for on-line visual learning and recognition, Proceedings of 16th International Conference on Pattern Recognition, vol. 3, IEEE, 2002, pp. 781–784.
[24] Gui-Fu Lu, Yong Wang, Feature extraction using a fast null space based linear discriminant analysis algorithm, Inform. Sci. 193 (2012) 72–80.
[25] Gui-Fu Lu, Wenming Zheng, Complexity-reduced implementations of complete and null-space-based linear discriminant analysis, Neural Networks 46 (10) (2013) 165–171.
[26] M. Kirby, L. Sirovich, Application of the KL procedure for the characterization of human faces, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1) (1990) 103–108.
[27] M. Lee, Riemannian Manifolds: An Introduction to Curvature, Springer, Berlin, 1997.
[28] B. Scholkopf, A. Smola, K.R. Muller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (5) (1998) 1299–1319.
[29] E. Pekalska, B. haasdonk, Kernel discriminant analysis for positive definite and indefinite kernels, IEEE Trans. Pattern Anal. Mach. Intell. 31 (6) (2009) 1017–1031.
[30] X. He, S. Yan, Y. Hu, P. Niyogi, H.-J. Zhang, Face recognition using Laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 328–340.
[31] X. He, 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, 2003.
[32] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, in: Processings of the Annual Conference on Neural Information Processing Systems, December 3–6, 2001, Vancouver, Canada, 2001.
[33] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.

[34] S. Chen, H. Zhao, M. Kong, B. Luo, 2DLPP: a two-dimensional extension of locality preserving projections, Neurocomputing 70 (4–6) (2007) 912–921.
[35] 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.
[36] B. Niu, Q. Yang, S.C.K. Shiu, S.K. Pal, Two-dimensional Laplacianfaces method for face recognition, Pattern Recognit. 41 (10) (2008) 3237–3243.
[37] 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.
[38] Y. Xu, G. Feng, Y.N. Zhao, One improvement to two-dimensional locality preserving projection method for use with face recognition, Neurocomputing 73 (2009) 245–249.
[39] H. Li, T. Jiang, K. Zhang, Efficient and robust feature extraction by maximum margin criterion, IEEE Trans. Neural Networks 17 (1) (2006) 1157–1165.
[40] J.G. Wang, W.K. Yang, Y.S. Lin, et al, Two-dimensional maximum scatter difference discriminant analysis for face recognition, Neurocomputing 72 (1–3) (2008) 352–358.
[41] M. Wan, Z. Lai, Z. Jin, Feature extraction using two-dimensional local graph embedding based on maximum margin criterion, Appl. Math. Comput. 217 (23) (2011) 9659–9668.
[42] T. Hastie, R. Tibshirani, J. Friedman, et al, The elements of statistical learning: data mining, inference, and prediction, Math. Intell. 27 (2) (2005) 83–85.