Local Graph Embedding Based on Maximum Margin Criterion (LGE/MMC) for Face Recognition

Minghua Wan and Shan Gai
School of Information Engineering, Nanchang Hangkong University
Nanchang 330063, China
E-mail: wmh36@sina.com, gaishan@yahoo.com

Jie Shao
School of computer Science, Shangqiu Institute of Technology
Shangqiu, 476000, China
E-mail: sj012328@163.com

Keywords: locally linear embedding, dimensional reduction, face recognition, maximum margin criterion, local graph embedding

Received: November 14, 2010

Locally linear embedding (LLE) is an efficient dimensional reduction algorithm for nonlinear data, and the low dimensional data can maintain topological relations in the original space after the processing. But this algorithm main application is not very good in the data dimensional reduction, the visualization and learning effects of data classification question and so on. In ordered to solve the above question, this paper proposes an efficient dimensional reduction and data classification method--local graph embedding method based on maximum margin criterion (LGE/MMC) for dimensional reduction, which is applied in face recognition. This goal of algorithm is preserved under nearest neighbour premise, where MMC criterion is used to construct the intrinsic graph and the penalty graph. In the intrinsic graph, the nonlinear structure is discovered in the high dimensional data space by the locally symmetric of linear restructuring, which is caused the similar sample as far as possible to gather in together. At the same time, the different class sample is far away as far as possible in the penalty graph. LGE/MMC seeks to minimize the difference, rather than the ratio, between the locality preserving between-class scatter and locality preserving within-class scatter. The results of face recognition experiments on ORL, YALE and AR face databases demonstrate the effectivity of the proposed method.

Povzetek: Članek opisuje algoritem za zmanjšanje števila dimenzij podatkov, ki se uporablja pri prepoznavanju obrazov.

1 Introduction

Face recognition has been active areas of research because of their potential applications in human–computer interfaces, image and computer vision. Linear dimensionality reduction seeks to find a meaningful low dimensional subspace in a high-dimensional input space. The subspace can provide a compact representation of the input data when the structure of data embedded is linear in the input space. Principal components analysis (PCA) [1] maintains the global Euclidean structure of the data in the high-dimensional space and preserves the total variance by maximizing the trace of the feature covariance matrix. Linear discriminant analysis (LDA) [2] preserves discriminative information between data of different classes and finds the optimal set of projection vectors by maximizing the ratio between the interclass and intraclass scatters.

PCA, LDA, and their variants [5, 6] are not able to reveal the underlying non-linear [3, 4] structure of the face data. Recently, many manifold learning-based algorithms with locality preserving abilities have been presented. Among them, isometric feature mapping (ISOMAP) [7], locally linear embedding (LLE) [8, 9], Laplacian eigenmap (LE) [10, 11] and local tangent space alignment (LTSA) [12] are widely used. He et al. [13, 14] proposed locality preserving projections (LPP), which is a linear subspace learning method derived from Laplacian Eigenmap. LPP can find an embedding space that preserves local information, and it is an unsupervised method. Many modified LPP algorithms have been put forward to consider the discriminant information of recognition task in recent years [15-18].

LLE is another representative local linear manifold learning method. Based on the assumption of the local linearity, LLE first constitutes local coordinates with the least constructed cost and then maps them to a global one. Some supervised versions of LLE [19–22] are introduced to deal with data sets labelled with class information and some other supervised LLE algorithms combined with LDA are becoming popular. Zhang et al. presented a unified framework of LLE and LDA [23,24]. Recently, He et al. [25] proposed another linear dimensionality reduction technique neighbourhood preserving embedding(NPE), which is the linearization of the locally linear embedding(LLE) algorithm and aims at finding a low-dimensional embedding that optimally preserves the local neighbourhood reconstruction relationships on the original data manifold. Some extension methods of NPE [26,27] are introduced to feature extraction. Experiments have proven that LLE and NPE are effective method for visualization.

Some other discriminant manifold learning algorithms, local discriminant embedding (LDE) [28], marginal Fisher analysis (MFA) [29] and neighbourhood preserving discriminant embedding (NPDE) [30] are proposed where their combine the Fisher criterion [31] with manifold criterion. They can be unified under the Fisher graph framework. However, they are different on their objective functions in terms of different graph embedding types and derivations. LDE utilizes LPP to form the intra-class and inter-class graph pair; MFA models the intra-class graph to characterize the intra-class compactness and the inter-class graph to characterize the inter-class separability with binary graph coefficient; and NPDE utilizes NPE to form the intra-class and inter-class graph pair to model the within- and between-neighbourhood scatters.

Above manifold learning algorithms can all be interpreted as the implementations of the linear graph embedding framework (LGE) [32] with different weight matrices or some variations. However, some limitations are exposed when LGE is applied to pattern recognition. One limitation is that some LGE such as LPP, LLE and NPE neglect the class information, which will impair the recognition accuracy. Another limitation lies in that some LGE such as LDE, MFA and DLPP involve inverse matrix of discriminant criterion, which will impair the recognition accuracy. So, in this paper we present local graph embedding method based on maximum margin criterion [33] (LGE/MMC) for dimensional reduction. Therefore, much computational time would be saved for feature extraction which is not necessary to convert the image matrix into high-dimensional image vector and avoids inverse matrix.

The rest of this paper is organized as follows: We review the ideas of linear methods in section 2. In Section 3, we propose the idea of LGE/MMC algorithm in detail. In section 4, we introduce the connections between LLE, NPE and LGE/MMC. Experiments are presented to demonstrate the effectiveness of LGE/MMC on face recognition in section 5. Finally, we give concluding remarks and a discussion of future work in Section 6.

2 Outline of Linear Methods

Let us consider a set of N sample $X=\{x_1,x_2,...,x_N\}$, $x_i \in R^D$ taking values in an n-dimensional image space. Let us also consider a linear transformation mapping the original n-dimensional space into a d-dimensional feature space $Y=\{y_1,y_2,...,y_N\}$, where $y_i \in R^d$ and $n>d$. The new feature vectors $y_i \in R^d$ are defined by the following linear transformation:

$$y_i = U^T x_i, \qquad i=1,\dots,N \tag{1}$$

where $U \in R^{n\times d}$ is a transformation matrix. In this section, we briefly review how the LDA, LPP and UDP algorithms realize subspace learning.

2.1 Linear discriminant analysis (LDA)

LDA [2] is a supervised learning algorithm. Let c denote the total class number and c_i denote the number of training samples in the i-th class. Let x_i^j, denote the j-th sample in i-th class, $\overline{x}$ be the mean of all the training samples, $\overline{x}_i$ be the mean of the i-th class. The between-class and within-class scatter matrices can be evaluated by:

$$S_b = \sum_{i=1}^{c} l_i(\overline{x}_i - \overline{x})(\overline{x}_i - \overline{x})^T \tag{2}$$

$$S_w = \sum_{i=1}^{c}\sum_{j=1}^{c_i}(x_i^j - \overline{x}_i)(x_i^j - \overline{x}_i)^T \tag{3}$$

LDA aims to find an optimal projection U such that the ratios of the between-class scatter to within-class scatter is maximized, i.e.

$$U = \arg\max_U \frac{\left|U^T S_b U\right|}{\left|U^T S_w U\right|} \tag{4}$$

where $\{U_i \mid i=1,2,...,d\}$ is the set of generalized eigenvectors of S_b and S_w corresponding to the d largest generalized eigenvalues $\{\lambda_i \mid i=1,2,...,d\}$, i.e.

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

2.2 Linear preserving projection (LPP)

The similarity matrix S of LPP [13,14] can be Gaussian weight or uniform weight of Euclidean distance using k-neighbourhood or ε-neighbourhood, defined as

$$S_{ij} = \begin{cases} 1, & \left\|x_i - x_j\right\|^2 < \varepsilon \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Hence, the objective function of LPP is defined as：

$$\min \sum_{i,j} \left\|y_i - y_j\right\| S_{ij} \tag{7}$$

where $\|\bullet\|$ means the L_2 norm. After some matrix analysis steps, the minimization problem becomes

$$\arg\min_U U^T XLX^T U$$
$$\text{s.t. } U^T XDX^T U = 1 \tag{8}$$

where $X=[X_1,X_2,\cdots,X_N]$ is the training space of size $n\times N$, and D is a diagonal matrix whose entries are column or row sums of S . $L = D - S$ is the Laplacian matrix.

The optimal d projection vectors that minimizes the objective function can be computed by the minimum eigenvalues solutions to the generalized eigenvalues problem

$$XLX^T U_i = \lambda_i XDX^T U_i \tag{9}$$

2.3 Maximum margin criterion (MMC)

The MMC is based on the difference of between-class scatter matrix and within-class scatter matrix, which is defined as follows:

$$J_s(w) = tr(U^T(S_b - \alpha S_w)U) \tag{10}$$

where the parameter α is a nonnegative constant which balances the relative merits of maximizing the between-class scatter to the minimization of the within-class scatter. The between-class scatter matrix S_b and within-class scatter matrix S_w can be denoted as

$$S_b = \frac{1}{n}\sum_{i=1}^{C} n_i(\mathbf{f}_i - \mathbf{f}_0)(\mathbf{f}_i - \mathbf{f}_0)^T \tag{11}$$

$$S_w = \frac{1}{n}\sum_{i=1}^{C}\sum_{j=1}^{n_i}(\mathbf{x}_j^i - \mathbf{f}_i)(\mathbf{x}_j^i - \mathbf{f}_i)^T \tag{12}$$

where n_i is the number of training samples in class i. In class i, the j[th] training sample is denoted by $\mathbf{x}_j^i$, the mean vectors of training samples in class i is denoted by $\mathbf{f}_i$ and the mean vector of all training samples is $\mathbf{f}_0$. Let $S_t = S_b + S_w$ and S_t denotes the total scatter matrix. As we know, S_b, S_w and S_t are all positive semi-definite.

3 Local Graph Embedding Based on Maximum Margin Criterion

3.1 The idea of LGE/MMC

When only a small number of training samples is available the within-class scatter matrix used by many feature extraction techniques (LDA, LPP etc) is singular, which represents a major obstacle for most techniques as they require an inversion of this singular matrix. Motivated by the idea of MMC, LGE/MMC seeks to minimize the difference, rather than the ratio, between the locality preserving between-class scatter and locality preserving within-class scatter. Then the singularity is avoided. LGE/MMC is theoretically elegant and can derive its discriminant vectors from both the range of the locality preserving between-class scatter and the range space of locality preserving within-class scatter. To gain more discriminative power, it is desirable to minimize the locality preserving between-class scatter and maximize the locality preserving within-class scatter simultaneously.

3.2 Locality preserving within-class scatter

To begin with, we propose to minimize the local scatter compactness of each data point by linear coefficients that reconstruct the data point from other points. The technique of local representation is the same as LLE [8, 9]. LLE regards each data point and its nearest neighbors as the locality. The algorithm can be described in three steps.

The first step of LLE is to select K_c -nearest neighbors of each data points x_i using Euclidean distances.

The second step of LLE is to calculate the reconstructing weight matrix $W = \left[w_{ij}\right]_{N\times N}$, which reconstructs each point x_i from its K_c -nearest neighbours. We can obtain the coefficient matrix W by minimizing the reconstruction error:

$$\min J_L(W) = \sum_{i=1}^{N}\left\| x_i - \sum_{j=1}^{K_c} w_{ij}x_j \right\|^2 \tag{13}$$

where $w_{ij} = 0$ if x_i and x_j are not neighbors, and the rows of W sum to 1: $\sum_{j=1}^{K_c} w_{ij} = 1$.

The reconstruction error can be converted to this form:

$$\begin{aligned}\xi &= \left\| x_i - \sum_{j=1}^{N} w_{ij}x_j \right\|^2 = \left\| \sum_{j=1}^{N} w_{ij}\left(x_i - x_j\right) \right\|^2 \\ &= \sum_{j=1}^{N} w_{ij}\left(x_i - x_j\right)\sum_{j=1}^{N} w_{it}\left(x_i - x_t\right) = \sum_{j=1}^{N}\sum_{t=1}^{N} w_{ij}w_{it}G_{jt}^i\end{aligned} \tag{14}$$

where $G_{jt}^i = \left(x_i - x_j\right)^T\left(x_i - x_t\right)$,called the local Gram matrix. By solving the least-squares problem with the constraint $\sum_{j=1}^{K_c} w_{ij} = 1$, the optimal coefficients are given:

$$w_{ij} = \frac{\sum_{t=1}^{K_c} G_{jt}^{-1}}{\sum_{p=1}^{K_c}\sum_{q=1}^{K_c} G_{pq}^{-1}} \tag{15}$$

After repeating the first step and the second step are performed on all the N data points, we can calculate the reconstruction weights to construct a weight matrix $W = \left[w_{ij}\right]_{N\times N}$.

The third step of LLE is to reconstruct represented y_i by the weight matrix W . To maintain the intrinsic geometrical feature of the data after the embedding process, the reconstruction error function must be minimized:

$$\min J_L(Y)=\sum_{i=1}^{N}\left\| y_i-\sum_{j=1}^{N} w_{ij} y_j \right\|^2 \quad (16)$$

where y_i is the output of x_i, y_j is a neighbor of y_i.

Considering the map in Eq. (1), the objective function reduces to

$$\begin{aligned} J_L(U) &= \sum_{i=1}^{N}\left\| y_i-\sum_{j=1}^{K_c} w_i^j y_j \right\|^2 \\ &= \sum_{i=1}^{N} tr\left\{\left(y_i-\sum_{j=1}^{K_c} w_i^j y_j\right)\left(y_i-\sum_{j=1}^{K_c} w_i^j y_j\right)^T\right\} \\ &= tr\left\{\sum_{i=1}^{N}\left(y_i-\sum_{j=1}^{K_c} w_i^j y_j\right)\left(y_i-\sum_{j=1}^{K_c} w_i^j y_j\right)^T\right\} \quad (17) \\ &= tr\left\{Y\left(I-W^T\right)\left(I-W^T\right)^T Y^T\right\} \\ &= tr\left\{Y\left(I-W\right)^T\left(I-W\right)Y^T\right\} \\ &= tr\left\{U^T X M X^T U\right\} \end{aligned}$$

where $M=\left(I-W\right)^T\left(I-W\right)$.

3.3 Locality preserving between-class scatters

To begin with, for the first aspect of our consideration, we propose to maximize the sum of pair wise squared distances between outputs if they have different labels. So, maximize locality preserving between-class scatter of samples is considered：

$$\max J_G(Y)=\sum_{i=1}^{N}\sum_{j=1}^{N}\left\| y_i-y_j \right\|^2 \quad (18)$$

Considering the map Eq.(1), the objective function reduces to

$$\begin{aligned} J_G(Y) &= \sum_i\sum_j \left\| y_i-y_j \right\|^2 W_{ij}^p \\ &= \sum_i\sum_j \left\| U^T x_i-U^T x_j \right\|^2 W_{ij}^p \\ &= 2U^T X(D^p-W^p)X^T U \\ &= 2U^T X L^p X^T U \quad (19) \end{aligned}$$

The variance of between-class points is deemed as local information. We construct the similarity matrix W_{ij}^p as follows:

$$W_{ij}^p=\begin{cases} 1, \text{ if } x_i \text{ is in the Kp nearest} \\ \quad \text{from different classes of } x_j \\ 0, \text{ otherwise} \end{cases} \quad (20)$$

3.4 Criterion of LGE/MMC

At last, when the locality preserving between-class scatter and the locality preserving within-class scatter have been constructed, an intuitive motivation is to find a common projection that minimizes between-class scatter $J_L(W)$ and maximizes within-class scatter $J_G(U)$ at the same time. Actually, we can obtain such a projection by the following multi-object optimized problem, that is:

$$\begin{cases} \min tr\left\{U^T X M X^T U\right\} \\ \max tr\left\{U^T X L^p X^T U\right\} \end{cases} \quad (21)$$

s.t. $U^T X X^T U = I$

The solution to the constrained multi-object optimized problem is to find a subspace which minimize the locality preserving between-class scatter and maximize the locality preserving within-class scatter simultaneously. Motivated by the idea of MMC, LGE/MMC seeks to minimize the difference, rather than the ratio, between the locality preserving between-class scatter and the locality preserving within-class scatter. So it can be changed into the following constrained problem:

$$\min tr\left\{U^T X\left(M-\mu L^p\right)X^T U\right\}$$

$$\text{s.t. } U^T X X^T U = I \quad (22)$$

where μ is an adjustable parameter to balance between-class scatter and within-class scatter.

Eq. (22) can be solved by Lagrange multiplier method:

$$L(U,\lambda)=\left\{U^T X\left(M-\mu L^p\right)X^T U-\lambda_i\left(U^T X X^T U-I\right)\right\} = 0 \quad (23)$$

where λ_i is the Lagrange multiplier. Thus we get:

$$X\left(M-\mu L^p\right)X^T U=\lambda_i X X^T U \quad (24)$$

where U_i is generalized eigenvector correspondingly to generalized eigenvalue λ_i.

4 Connection between LLE, NPE and LGE/MMC

In this Section, LGE/MMC seems to be formally similar to LLE and NPE. However, LGE/MMC is also obviously different from them. In order to investigate the similarity and the difference, we discuss the connections between LLE, NPE and LGE/MMC.

4.1 Connection between LLE and NPE

LLE and NPE aim to discover the local structure of the data manifold. LLE is defined only on the training samples, and there are no natural maps of the testing sample. Instead, NPE is defined on both the training and test samples. NPE is a linear approximation to LLE.

In NPE, the matrix XMX^T is symmetric and semi-positive definite. In order to remove an arbitrary scaling factor in the projection, we impose a constraint as follows:

$$YY^T = I \Rightarrow U^T XX^T U = I \quad (25)$$

Finally, the minimization problem reduces to finding U :

$$\min_{U^T XX^T U = I} tr\left\{U^T XMX^T U\right\} \quad (26)$$

The transformation matrix U that minimizes the objective function is given by the minimum eigenvalue solution to the following generalized eigenvector problem:

$$XMX^T U_i = \lambda_i XX^T U_i \quad (27)$$

4.1 **Connection between NPE and** LGE/MMC

As a result, LGE/MMC is formulated as the following constrained minimization problem:

$$\min_{U^T XX^T U = I} tr\left\{U^T X\left(M - \mu L^p\right) X^T U\right\} \quad (28)$$

Thus we have:

$$X\hat{M}X^T U_i = \lambda_i XX^T U_i \quad (29)$$

where $\hat{M} = M - \tilde{M}$, $\tilde{M} = \mu L^p$. It is easy to see that NPE is a special case of LGE/MMC (i.e. when $\mu = 0$).

4.2 Connection between LLE, NPE and LGE/MMC

From above discussed, NPE and LGE/MMC yield mappings that are defined not only on the training data points but also on novel testing points. The essence of NPE is the linear approximation to LLE. As we know, the graph construction of LLE and NPE fails to use the global discriminative information. However, we can see from $\hat{M}$ first that LGE/MMC preserves the locality characteristic since M still exists and second that it adds the discriminant information through $\tilde{M}$. From what has been discussed above, it can be concluded that LGE/MMC builds a new graph with different edge weight assignment method, integrating both local information and discriminant information. Thus, by integrating the discriminant into the objective function, LGE/MMC will be more robust than LLE and NPE.

5 Comparisons of Computation complexity and space complexity

In Table 1, we compare the computational and the memory space complexities of the six methods. Here m and n is the number of the rows and the columns of the image matrix. L, M and N are the number of the projection vectors, the testing and the training samples, respectively.

Table 1: The computational and the memory space complexities of the six methods.

Method	Complexity		
	Time (training)	Time (testing)	Memory
PCA	O(m^2n^2L)	O(MNL)	O(m^2n^2)
LDA	O(m^2n^2L)	O(MNL)	O(m^2n^2)
MMC	O(m^2n^2L)	O(MNL)	O(m^2n^2)
LLE	O$(m^2n^2L+ mnN^2)$	O(MNL)	O(m^2n^2)
LLE+LDA	O$(2m^2n^2L+ mnN^2)$	O$(2MNL)$	O$(2m^2n^2)$
LGE/MMC	O$(m^2n^2L+ 2mnN^2)$	O(MNL)	O(m^2n^2)

In Table 1, for the PCA, LDA and MMC, since we need to perform O(MN) tests when using the nearest neighbour rule for classification and for each test it has the time complexity of O(L), the testing time is O(MNL). The memory cost is determined by the size of the matrices of the associated eigen equations, which is O(m^2n^2). The training time complexity depends on both the size of the matrices in the eigen equations and the number of the projection vectors that are required to be computed, which is O(m^2n^2L). For the LLE method, an extra time cost to construct the similarity matrix, i.e., O(mnN^2), will be taken into account. So, LLE+LDA has the time complexity of O$(2m^2n^2L+ mnN^2)$, the testing time is O$(2MNL)$ *and* Memory is O$(2m^2n^2)$. The proposed method LGE/MMC has the time complexity of O$(m^2n^2L+ 2mnN^2)$, the testing time is O(MNL) *and* Memory is O(m^2n^2). So the proposed method is much than other methods in testing time.

6 Experiments and results

To evaluate the proposed LGE/MMC algorithm, we systematically compare it with the PCA [1], LDA [2], LLE [8-9], MMC [33] and LLE+LDA [23-24] algorithm in three face databases: ORL, YALE and AR. When the projection matrix was computed from the training part, all the images including the training part and the test part were projected to feature space. Euclidean distance and nearest neighborhood classifier are used in all the experiments. The experiments were carried out on the same PC (CPU: P4 2.8 GHz, RAM: 1024 MB).

6.1 Database

The ORL face database [34] contains images from 40 individuals, each providing 10 different images where the pose, face expression and sample size vary. The facial expressions and facial details (glasses or no glasses) also vary. The images were taken with a tolerance for some tilting and rotation of the face of up to 20 degrees. Moreover, there is also some variation in this scale of up to about 10 percent. All images normalized to a resolution of 56×46. We test the recognition performances of the six methods: PCA, LDA, LLE, MMC, LLE+LDA and LGE/MMC. 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 $10-l$ images are used for testing. For each l, we independently run 50 times. In the PCA phase of LDA, LLE, MMC, LLE+LDA and LGE/MMC, we keep 95 percent image energy.

The YALE face database [35] contains 165 gray scale images of 15 individuals, each individual has 11 images. The images demonstrate variations in lighting condition, facial expression (normal, happy, sad, sleepy, surprised, and wink). In this experiment, each image in Yale database was manually cropped and resized to 50×40. In the PCA phase of LDA, LLE, MMC, LLE+LDA and LGE/MMC, we keep 95 percent image energy. 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 $11-l$ images are used for testing. For each l, we independently run 50 times.

The AR face database [36] contains over 4,000 color face images of 126 people (70 men and 56 women), including frontal views of faces with different facial expressions, lighting conditions, and occlusions. The pictures of 120 individuals (65 men and 55 women) were taken in two sessions (separated by two weeks) and each section contains 13 colour images. The face portion of each image is manually cropped and then normalized to 50×40 pixels. These images vary as follows: 1. neutral expression 2. smiling 3. angry 4. screaming 5. left light on 6. right light on 7. all sides light on 8. wearing sum glasses 9. wearing sun blasses and left light on 10. wearing sun blasses and right light on. In this experiment, 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 $20-l$ images are used for testing. For each l, we independently run 10 times. In the PCA phase of LDA, LLE, MMC, LLE+LDA and LGE/MMC, the number of principle components is set as 150. The dimension steps are set to be 5 in final low-dimensional subspaces obtained by the seven methods.

Fig.1, Fig.2 and Fig.3 show the sample images from the three databases.

Fig.1 Images of one person on the ORL database

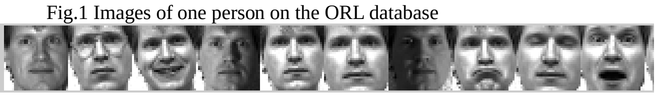


Figure 2: Images of one person on the YALE database.

Figure 3: Images of one subject of the AR database. The first line and the second line images were taken in different time (separated by two weeks).

6.2 Experimental results and analysis

Except PCA and LDA, the local methods involved in the experiments are manifold learning based approaches, where K_c and K_p nearest neighbourhood search are contained. Thus how to select K_c and K_p are an important problem in feature extraction. If the value of K_c and K_p are too small, it is very difficult to preserve the topologic structure in low-dimensional space. On the contrary, if the value of K_c and K_p are too big, it is very difficult to depict the assumption of local linearity in high dimensional space. So it will affect the dimensionality manifold reduction result by the value of K_c and K_p. In the first experiment, we investigate the performance of the LGE/MMC algorithm over the reduced dimensions versus the corresponding varied the value of K_c and K_p. To find how K_c and K_p affect the recognition performance, we change $k_c = l-1$ and K_p are from 1 to 20 with step 1. Fig.4 displays the average recognition rates with varied the value of K_p by carrying out LGE/MMC when only two images per class were randomly selected for training on the YALE face databases. LGE/MMC obtains the best average recognition rate is 94.79% when K_p=4. This indicates that the locality and the globality are with the same importance. In the next experiment, the value of adjustable parameter K_p is taken to be 4.

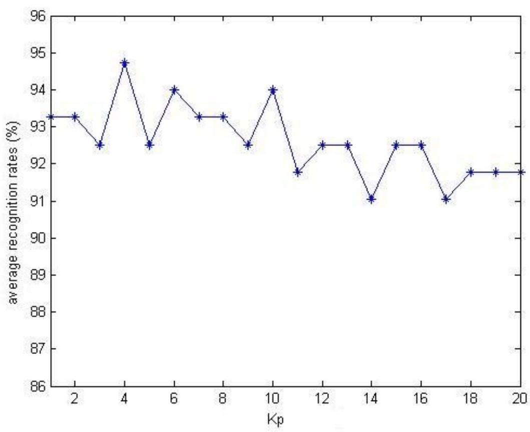


Figure 4: The average recognition rates (%) of LGE/MMC versus the corresponding varied the value of

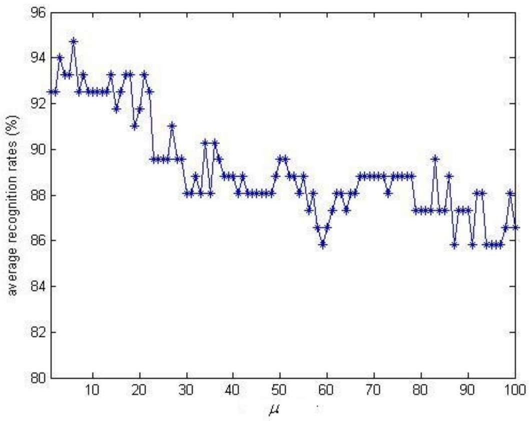


Figure 5: The maximal average recognition rates (%) of LGE/MMC versus the corresponding varied the value of μ when only two images per class were randomly selected for training on the YALE face database.

K_p when only two images per class were randomly selected for training on the YALE face databases.

In the second experiment, we also test the impact of μ on the performance when only two images per class were randomly selected for training on the YALE face database, which can be found in Fig. 5. We varied μ from 1 to 100 with step 1. Fig. 5 displays the maximal average recognition rates with varied parameter μ by carrying out LGE/MMC. From Fig. 5, it can be found that the effectiveness of the LGE/MMC algorithm is sensitive to the value of the parameter μ . LGE/MMC obtains the best average recognition rate is 94.79% when μ =6. This indicates that the locality and the globality are with the same importance. In the next experiment, the value of adjustable parameter μ is taken to be 6.

In the third experiment, we randomly select l images (l varies from 2 to 6) of each individual for training, and the remaining ones are used for testing. We compare the performances of different algorithms. The average recognition rates obtained by different algorithms as well as the corresponding dimensionality of reduced subspace (the numbers in parentheses) on the ORL, YALE and AR face databases are given in the Table 2, Table 3 and Table 4, respectively. Fig.6. is the average recognition rates (%) of LGE/MMC versus the corresponding varied dimensions when only six images per class were randomly selected for training on the ORL, YALE and AR face databases. We change the number of eigenvectors from 2 to 50 with step 2 on the ORL, YALE face databases and the number of eigenvectors from 5 to 150 with step 5 on the AR face database, respectively.

Table 2: The average recognition accuracy(%)of different algorithms on the ORL face database and the corresponding standard deviations and dimensions.

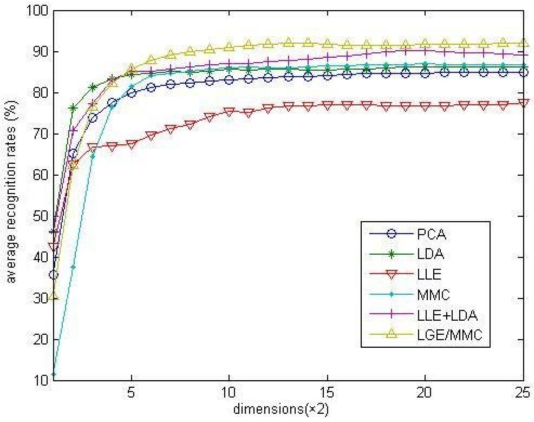
Methods \ l	2	3	4	5	6
PCA	72.25 ± 2.61 (22)	81.62 ± 0.65 (22)	83.98 ± 1.00 (22)	85.41 ± 1.30 (22)	85.41 ± 1.98 (22)
LDA	76.35 ± 1.07 (28)	83.40 ± 1.69 (28)	84.13 ± 2.03 (28)	86.07 ± 1.16 (28)	88.60 ± 0.78 (28)
LLE	69.78 ± 0.82 (32)	72.66 ± 0.73 (30)	76.32 ± 1.14 (36)	78.85 ± 1.15 (20)	88.44 ± 1.55 (28)
MMC	75.01 ± 1.77 (26)	83.96 ± 0.48 (28)	84.37 ± 2.48 (28)	87.47 ± 0.85 (28)	90.49 ± 1.17 (26)
LLE+LDA	73.36 ± 1.24 (18)	85.84. ± 1.23 (20)	89.30 ± 0.83 (22)	88.61 ± 1.79 (20)	94.53 ± 1.82 (10)
LGE/MMC	75.53 ± 1.88 (30)	86.00 ± 1.41 (36)	91.65 ± 0.33 (36)	94.17 ± 1.45 (36)	96.53 ± 0.72 (36)

Table 3: The average recognition accuracy(%) of different algorithms on the YALE face database and the corresponding standard deviations and dimensions.

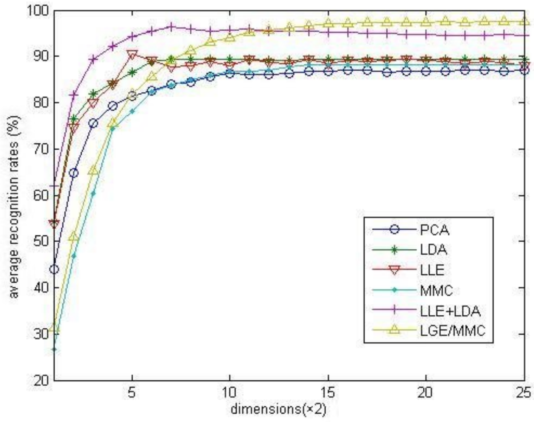
Methods \ l	2	3	4	5	6
PCA	77.29 ± 1.20 (18)	80.20 ± 1.27 (20)	83.99 ± 1.38 (18)	84.72 ± 1.24 (20)	86.21 ± 0.80 (22)
LDA	80.45 ± 1.48 (8)	84.55 ± 1.06 (8)	87.85 ± 0.45 (8)	87.20 ± 1.64 (8)	88.54 ± 0.82 (8)
LLE	83.55 ± 1.38 (14)	83.58 ± 0.59 (6)	85.90 ± 0.75 (6)	88.37 ± 1.63 (8)	88.97 ± 1.56 (8)
MMC	80.58 ± 0.71 (12)	82.84 ± 0.88 (6)	85.87 ± 1.12 (6)	86.83 ± 0.37 (6)	87.79 ± 0.50 (6)
LLE+LDA	88.33 ± 1.21 (22)	92.45 ± 0.75 (18)	92.84 ± 1.01 (12)	95.34 ± 1.21 (10)	95.21 ± 1.24 (10)
LGE/MMC	94.38 ± 0.41 (36)	94.54 ± 0.83 (16)	93.84 ± 1.16 (38)	95.47 ± 0.11 (38)	96.29 ± 1.26 (38)

Table 4: The average recognition accuracy(%) of different algorithms on the AR face database and the corresponding standard deviations and dimensions.

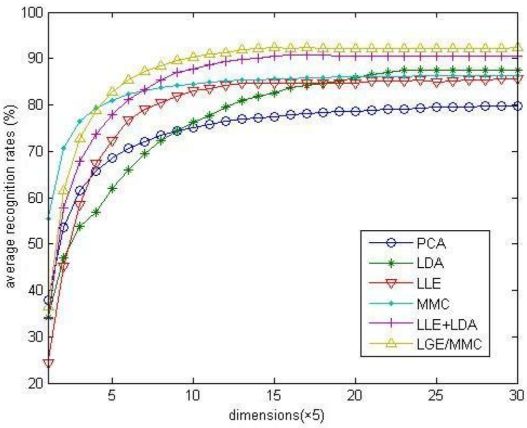
Methods \ l	2	3	4	5	6
PCA	66.68 ± 1.11 (85)	70.21 ± 1.62 (85)	77.60 ± 1.27 (85)	79.03 ± 0.81 (80)	81.54 ± 1.04 (85)
LDA	71.50 ± 0.71 (70)	75.58 ± 0.76 (70)	82.53 ± 0.81 (70)	87.12 ± 0.33 (70)	87.58 ± 0.75 (70)
LLE	70.40 ± 0.89 (85)	74.31 ± 1.16 (75)	83.35 ± 1.49 (70)	84.78 ± 0.76 (80)	86.98 ± 0.40 (85)
MMC	69.39 ± 1.15 (80)	75.71 ± 0.59 (80)	82.98 ± 0.62 (85)	85.27 ± 0.94 (80)	87.86 ± 0.82 (80)
LLE+LDA	69.38 ± 1.20 (90)	79.23 ± 0.89 (80)	88.17 ± 1.36 (80)	89.09 ± 1.49 (85)	92.420 ± 1.08 (80)
LGE/MMC	71.99 ± 0.99 (75)	81.00 ± 0.78 (50)	90.46 ± 0.89 (85)	91.39 ± 0.91 (45)	95.60 ± 0.80 (50)



(a)ORL face database($l = 4$)



(b) Yale face database($l = 6$)



(c)AR face database($l = 5$)

Figure 6: The average recognition rates (%) of LGE/MMC versus the corresponding varied dimensions on the ORL, YALE and AR face databases.

The above experiments showed that the maximal average recognition rates of all methods increases with the increase in training sample size in Table 2, Table 3 and Table 4 respectively. The proposed LGE/MMC algorithm consistently outperforms better than other methods in all experiments in three face databases. From Fig.6 we can find that with the increase number of eigenvectors on three face databases, the average recognition rates also improved.

7 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 local graph embedding method based on maximum margin criterion (LGE/MMC) for dimensional reduction. The results of face recognition experiments on ORL, YALE and AR face databases demonstrate the effectivity of the proposed method. In the future, we will make more tests on other types of data and decide the optimal parameter μ , K_c and K_p . For future work, we will extend LGE/MMC to supervised and semi-supervised cases.

Acknowledgements

This work is partially supported by the National Science Foundation of China under grant no. 60632050, 90820306, 60873151, 60973098, 61162002, 61005008, 61005005, 60963002, the National Science Foundation of Jiangxi Provincial under grant no. 20114BAB201034, China's Aviation Science no. 20115556007 and Youth Foundation of Jiangxi Provincial Department of Education no. GJJ12459.

References

[1] M. Turk, A.P. Pentland (1991) Face recognition using eigenfaces. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, June 1991, pp. 586–591

[2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriengman (1997) Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intelligence 19(7): 711–720

[3] M. Kirby, L. Sirovich (1990) Application of the KL procedure for the characterization of human faces. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(1): 103–108

[4] [4] J.M. Lee (1997) Riemannian Manifolds: An Introduction to Curvature. Springer, Berlin

[5] B. Scholkopf, A. Smola, K.R. Muller (1998) Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput 10 (5): 1299–1319

[6] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, A. Smola, K.-R. Muller (2003) Constructing descriptive and discriminative nonlinear features: Rayleigh coefficients in kernel feature spaces. IEEE Trans. Pattern Anal. Mach. Intell. 25(5): 623–628.

[7] J.B. Tenenbaum, V. de Silva, J.C. (2000) Langford A global geometric framework for nonlinear dimensionality reduction. Science 290(5500): 2319–2323.

[8] S.T. Roweis, L.K. Saul (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500): 2323–2326.

[9] L.K. Saul, S.T. Roweis (2003) Think globally, fit locally: unsupervised learning of low dimensional manifolds. J. Mach. Learn. Res 4: 119–155.

[10] M. Belkin and P. Niyogi, T. G. Dietterich, S. Becker, and Z. Ghahramani (2000) Laplacian eigenmaps and spectral techniques for embedding and clustering. Advances in Neural Information Processing Systems 14: 873-878

[11] M. Belkin, P. Niyogi (2003) Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput 15(6): 1373–1396.

[12] Z. Zhang, H. Zha (2004) Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. SIAM Journal of Scientific Computing 26(1): 313–338

[13] X. He and P. Niyogi (2003) Locality Preserving Projections. In: Proceedings of the 17th Annual Conference on Neural Information Processing Systems, Vancouver and Whistler, Canada, December 2003, pp. 153-160

[14] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang (2005) Face Recognition Using Laplacianfaces. IEEE Trans. Pattern Analysis and Machine Intelligence 27(3): 328-340

[15] H. Hu (2008) Orthogonal neighborhood preserving discriminant analysis for face recognition. Pattern Recognition 41: 2045–2054.

[16] W. Yu, X. Teng, C. Liu (2006) Face recognition using discriminant locality preserving projections. Image Vision Computing 24: 239–248

[17] L. Yang, W. Gong, X. Gu, W. Li, Y. Liang (2008) Null space discriminant locality preserving

projections for face recognition. Neurocomputing 71: 3644–3649

[18] G. F. Lu, Z. Lin, Z. Jin (2010) Face recognition using discriminant locality preserving projections based on maximum margin criterion. Pattern Recognition 43: 3572-2579

[19] D. de Ridder, R.P.W. Duin, Locally linear embedding for classification, Technical Report PH-2002-01, Pattern Recognition Group, Department of Imaging Science and Technology, Delft University of Technology, Delft, The Netherlands, 2002.

[20] D. de Ridder, O. Kouropteva, O. Okun, M. Pietik¨ainen, R.P.W. Duin, Supervised locally linear embedding, artificial neural networks and neural information processing, in: ICANN/ICONIP 2003 Proceedings, Lecture Notes in Computer Science, vol. 2714, Springer, Berlin, 2003, pp. 333–341.

[21] X. Bai, B. Yin, Q. Shi, Y. Sun, Face recognition based on supervised locally linear embedding method, J. Inf. Comput. Sci. (4) (2005) 641–646.

[22] O. Kouropteva, O. Okun, M. Pietikainen, Supervised locally linear embedding algorithm for pattern recognition, IbPRIA 2003, Lecture Notes in Computer Science, vol. 2652, Springer, Berlin, 2003, pp. 386–394.

[23] J. Zhang, H. Shen, Z.-H. Zhou, Unified Locally Linear Embedding and Linear Discriminant Analysis Algorithm for Face Recognition, Lecture Notes in Computer Science, Springer, Berlin, 2004.

[24] J. Zhang, H. Shen, Z.-H. Zhou, Ensemble-based discriminant manifold learning for face recognition, ICNC 2006, Part I, Lecture Notes in Computer Science, vol. 4221, Springer, Berlin, 2006, pp. 29–38.

[25] X.F. He, D. Cai, S.C. Yan, H.J. Zhang (2005) Neighborhood preserving embedding. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), Beijing, china, October 2005, pp. 1208–1213.

[26] X.H. Zeng, S.W. Luo (2007) A supervised subspace learning algorithm: supervised neighborhood preserving embedding. In: Proceedings of 3rd International Conference on Adwanced Data Mining and Applications, Harbin, China, August 2007, pp 81-88

[27] Y. Wang, Y. Wu (2010) Complete neighborhood preserving embedding for face recognition. Pattern Recognition 43:1008-1015

[28] C. Hwann-Tzong, C. Huang-Wei, L. Tyng-Luh, Local discriminant embedding and its variants, in: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 (IEEE Computer Society, 2005).

[29] S.C. Yan, D. Xu, B.Y. Zhang, H.J. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: A general framework for dimensionality reduction, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2007) 40-51.

[30] P.Y. Han, A.T.B. Jin, F.S. Abas, Neighbourhood preserving discriminant embedding in face recognition, J. Vis. Commun. Image Represent. 20 (2009) 532-542.

[31] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (91CH2983-5), (1991).

[32] S. Yan, D. Xu, B. Zhang, and H.-J. Zhang, Graph Embedding: A General Framework for Dimensionality Reduction, IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR'05), vol. 2, pp. 830-837, 2005, San Diego.

[33] H. Li, T. Jiang, K. Zhang, Efficient and robust feature extraction by maximum margin criterion, IEEE Transactions on Neural Networks 17 (1) (2006) 1157–1165.

[34] AT&T Laboratories Cambridge, The ORL database of faces, _http://www.uk.research.att.com/ facedatabase. Html.

[35] Yale database, 1997. Available from: <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.

[36] A. Martinez, R. Benavente, The AR face database, CVC.