Machine Vision and Applications (2011) 22:619–627
DOI 10.1007/s00138-011-0332-1

ORIGINAL PAPER

Maximal local interclass embedding with application to face recognition

Zhihui Lai · Cairong Zhao · Yi Chen · Zhong Jin

Received: 22 December 2008 / Revised: 10 February 2011 / Accepted: 11 March 2011 / Published online: 2 April 2011

Abstract Dimensionality reduction of high dimensional data is involved in many problems in information processing. A new dimensionality reduction approach called maximal local interclass embedding (MLIE) is developed in this paper. MLIE can be viewed as a linear approach of a multimanifolds-based learning framework, in which the information of neighborhood is integrated with the local interclass relationships. In MLIE, the local interclass graph and the intrinsic graph are constructed to find a set of projections that maximize the local interclass scatter and the local intraclass compactness simultaneously. This characteristic makes MLIE more powerful than marginal Fisher analysis (MFA). MLIE maintains all the advantages of MFA. Moreover, the computational complexity of MLIE is less than that of MFA. The proposed algorithm is applied to face recognition. Experiments have been performed on the Yale, AR and ORL face image databases. The experimental results show that owing to the locally discriminating property, MLIE consistently outperforms up-to-date MFA, Smooth MFA, neighborhood preserving embedding and locality preserving projection in face recognition.

Keywords Dimensionality reduction · Manifold learning · Graph embedding · Marginal Fisher analysis (MFA)

Z. Lai (✉) · C. Zhao · Y. Chen · Z. Jin
School of Computer Science, Nanjing University of Science and Technology, Nanjing, 210094 Jiansu, People's Republic of China
e-mail: lai_zhi_hui@163.com

C. Zhao
Department of Physics and Electronics, Minjian College, Fuzhou, 350108 Fujian, People's Republic of China

1 Introduction

Techniques for dimensionality reduction have attracted much attention in computer vision and pattern recognition [1]. From the perspective of pattern recognition, dimensionality reduction is an effective way to avoid the "curse of dimensionality" [2,3] and to improve the computational efficiency of pattern matching.

Researchers have developed many useful dimensionality reduction techniques. Principal component analysis (PCA) [4–6] and Linear discriminant analysis (LDA) [5,7–9] have been two of the most popular methods because of their relative simplicity and effectiveness. PCA performs dimensionality reduction by projecting the original n-dimensional data onto the k-dimensional ($k \ll n$) linear subspace spanned by the leading eigenvectors of the data's covariance matrix. Thus, PCA builds a global linear model of the data. LDA searches for the projective axes on which the data point of different classes are far from each other (i.e. between-class scatter is maximized), while constraining the data points in the same class to be as close to each other as possible (i.e. within-class scatter is minimized). Linear models, however, may fail to discover essential data structures that are nonlinear. As a result, a number of nonlinear dimensionality reduction techniques have been developed to address this problem. Kernel-based techniques and manifold learning-based techniques are the two most popular nonlinear dimensionality reduction methods. The basic idea of kernel-based techniques, such as kernel principal component analysis (KPCA) [10], kernel Fisher discriminant (KFD) [11] and kernel ridge regression [12], is to implicitly map the observed patterns into potentially much higher dimensional feature vectors by a nonlinear mapping determined by a kernel. In contrast with kernel-based techniques, the motivation of manifold learning is straightforward as it seeks to directly find the intrinsic low-dimensional nonlinear data structures hidden in the observation space. Among the most well-known manifold learning methods are isometric feature mapping (ISOMAP) [13], local linear embedding (LLE) [14], and Laplacian Eigenmaps [15]. Although such embeddings are good for representation, they are defined only on the training data and it is still unclear how to evaluate the embeddings on the test data points.

To facilitate the nearest-neighbor searches for new test data, recently, He et al. [16,17] proposed locality preserving projection (LPP), which attempts to preserve the data's localities or similarities in the low-dimensional Euclidean space. In LPP algorithm, a local neighborhood similarity/affine matrix, i.e. a graph, is constructed and used to model the patterns' distributions on the manifold. This idea was further extended and a novel algorithm called marginal fisher analysis (MFA) was proposed in [18]. By using the Laplacian penalty to incorporate the prior information that neighboring pixels are correlated, smooth MFA (SMFA) was developed in [19] to learn a smooth subspace for feature extraction. However, in MFA-family algorithms, the penalty graph characterizing the interclass separability cannot reflect the neighbor relationships between the different classes in that only the k_p nearest data pairs of the different classes are used and other neighbor relationships of the different classes are discarded. As a result, the penalty graph focuses more on the local neighborhood of the outliers than elsewhere. Thus, the projections of MFA are sensitive to outliers. Another disadvantage of MFA algorithm is that the k_p nearest sample pairs are found by the global searching among all the data pairs in k_p nearest neighbor samples. Thus, the computational complexity of MFA is high and MFA algorithm is relatively time consuming. In order to obtain a high recognition rate with less computational time and to preserve all the advantages of MFA, we construct the local interclass graph and a new algorithm called maximal local interclass embedding (MLIE) is proposed in this paper.

The remainder of this paper is organized as follows. Section 2 outlines LDA and MFA. Section 3 develops the idea of MLIE and the relevant theory and algorithm. Section 4 describes the related experiments. Our conclusions and future work are given in Sect. 5.

2 Outline of LDA and MFA

2.1 LDA

LDA seeks to find a projection axis such that the Fisher criterion (i.e. the ratio of the between-class scatter to the within-class scatter) is maximized. Specifically, given a set of M training samples (pattern vectors) $x_1, x_2, \ldots, x_M$ in R^n. The between-class and within-class scatter matrices S_B and S_W are defined by

$$S_B = \frac{1}{M}\sum_{i=1}^{c} l_i\,(m_i - m_o)\,(m_i - m_o)^T, \tag{1}$$

$$S_W = \frac{1}{M}\sum_{i=1}^{c}\sum_{j=1}^{l_i}\left(x_{ij} - m_i\right)\left(x_{ij} - m_i\right)^T \tag{2}$$

where x_{ij} denotes the j-th training sample in i-th class, c is the number of classes, l_i is the number of training samples in i-th class, and m_i is the sample mean of i-th class, m_o is the global mean of all samples.

It is easy to show that S_B and S_W are both nonnegative definite matrices. The Fisher criterion is defined by

$$\max\ J_{\mathrm{LDA}}(\omega) = \max\ \frac{\omega^T S_B \omega}{\omega^T S_W \omega} \tag{3}$$

The optimal projections of LDA can be obtained by solving the following generalized Eigen-equation

$$S_B\omega = \lambda S_W \omega \tag{4}$$

Let $w_1, w_2, \ldots, w_d$ be the generalized Eigenvectors of Eq. (4) corresponding to the first d largest generalized Eigenvalues. Then $W_{\mathrm{LDA}} = [\omega_1, \omega_2, \ldots, \omega_d]$ is the optimal transformation matrix of LDA.

2.2 MFA

Marginal Fisher analysis [12] designs the intrinsic graph and the penalty graph to characterize the intraclass compactness and interclass separability, respectively. The intrinsic graph describes the data points' adjacency relationships of intraclass, and each sample is connected to its k_w-nearest neighbors in the same class. The penalty graph describes the interclass marginal points' adjacency relationships and the marginal point pairs in k_p-nearest neighbors of different classes are connected.

By the following graph embedding formulation, interclass separability is characterized by the penalty graph with the term

$$\begin{aligned} S_p &= \frac{1}{2}\sum_i \sum_{(i,j)\in P_{k_p}(c_i)\ \text{or}\ (i,j)\in P_{k_p}(c_i)} \left\|\omega^T x_i - \omega^T x_j\right\|^2 \\ &= \omega^T X\left(D^p - W^p\right)X^T\omega, \end{aligned} \tag{5}$$

where

$$W_{ij}^p = \begin{cases} 1, & \text{if } (i,j)\in P_{k_p}(c_i) \text{ or } (i,j)\in P_{k_p}(c_i) \\ 0, & \text{else.} \end{cases} \tag{6}$$

and D^p is a diagonal matrix with diagonal elements $d_{ii}^p = \sum_j W_{ij}^p$, $X = [x_1, x_2, \ldots, x_M]$, and $P_{k_p}(c)$ is a set of data pairs that are the k_p nearest pairs among the set $\{(i,j), i \in \pi_c, j \notin \pi_c\}$, where π_c denotes the index set belonging to the c-th class. Note that constructing the penalty graph needs two steps: the first step is to calculate the pairwise distance between all samples and search for k_p -nearest neighbors; the second step is to calculate the pairwise distance among k_p samples and search for k_p -nearest neighbors data pairs since there are $k_p\,(k_p-1)/2$ data pairs, and then connect them. The computational complexity will be further analyzed in Sect. 3.4.

Intraclass compactness is characterized from the intrinsic graph by the term

$$S_C = \frac{1}{2}\sum_i \sum_{i\in N^+_{k_w}(j) \text{ or } j\in N^+_{k_w}(i)} \left\|\omega^T x_i - \omega^T x_j\right\|^2 = \omega^T X\,(D-W)\,X^T\omega, \tag{7}$$

where

$$W_{ij} = \begin{cases} 1, & \text{if } i \in N^+_{k_w}(j) \text{ or } j \in N^+_{k_w}(i) \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

and D is a diagonal matrix with diagonal elements $d_{ii} = \sum_j W_{ij}$, and $N^+_{k_w}(i)$ indicates the index set of the k_w nearest neighbors of the sample x_i in the same class.

With the intrinsic graph and penalty graph, MFA algorithm aims to find the optimal projection

$$\omega^*_{\text{MFA}} = \arg\min_{\omega} \frac{\omega^T X(D-W)X^T\omega}{\omega^T X(D^p - W^p)X^T\omega}. \tag{9}$$

The optimal projections can be obtained by solving the corresponding generalized Eigen-equation:

$$X\,(D-W)\,X^T\omega = \lambda X\left(D^p - W^p\right)X^T\omega \tag{10}$$

The optimal projections of MFA are the d generalized eigenvectors corresponding to the first d smallest generalized eigenvalues of Eq. (10).

3 Maximal local interclass embedding

3.1 Basic ideas

In graph-embedding algorithm, it is crucial to construct the graphs that can reflect the relationships of the samples correctly. In MFA algorithm, the intraclass compactness is characterized by the intrinsic graph and the interclass separability is characterized by the penalty graph. In our opinion, the intrinsic graph is constructed correctly to characterize the intraclass compactness. However, the penalty graph cannot completely reflect the neighbor relationship between classes because only the k_p nearest point pairs of different classes are used but other (or remaining) neighbor relationships between samples in different classes are discarded. Discarding these neighbor relationships may yield a bad result.

We randomly generate ten data points with an outlier in one class and compute the projections of MFA and MLIE using Matlab. The results are shown in Fig. 1. Figure 1 illustrates the drawback of penalty graph of MFA and the advantage of the local interclass graph in the proposed method when $k_p = k_s = 2$, where k_s is the number of nearest neighbors in the local interclass graph in our proposed method. As it is illustrated in Fig. 1a, when there is an outlier, the penalty graph only focuses on the local neighborhood of the outlier (or $k_p = 2$ nearest data pairs between different classes) and MFA yields a bad projection axis on which the samples in different classes are highly overlapped. It happens that some samples in the same class may enter or be close to the other classes when they are projected to the subspace learned by MFA. In other words, some samples' neighbor relationships are thrown into confusion in the subspace of MAF, which may greatly degrade the recognition rate. As a result, MFA is sensitive to outliers. Another disadvantage of MFA algorithm is that the k_p nearest sample pairs are found by a global searching among all the sample point pairs. Thus, MFA algorithm is relatively time consuming.

In order to address these disadvantages and to obtain a high recognition rate, a local interclass graph, which characterizes the local interclass relationships, is constructed directly. Then, a novel algorithm, called MLIE, is proposed.

3.2 Local interclass graph

By following the graph embedding formulation, interclass separability is characterized by a local interclass graph with the term

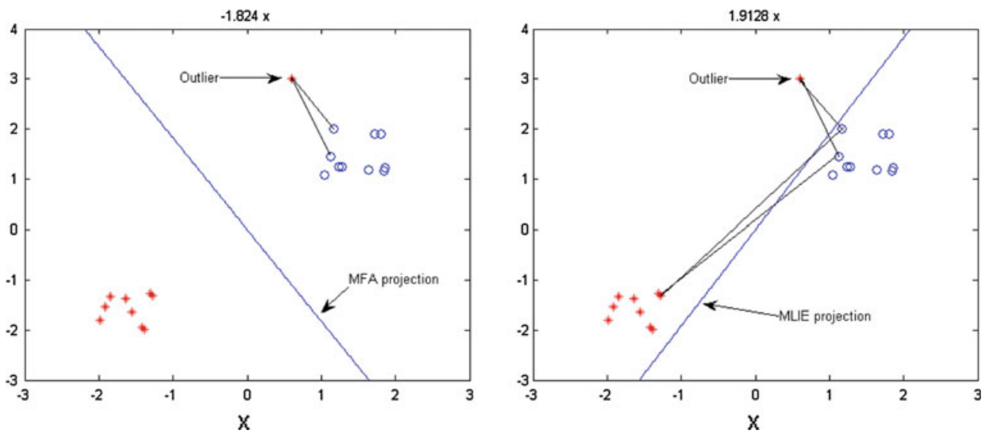
$$\begin{aligned} S_s &= \frac{1}{2}\sum_i\sum_j \left\|\omega^T x_i - \omega^T x_j\right\|^2 W^s_{ij} \\ &= \frac{1}{2}\sum_i\sum_j W^s_{ij}\left(\omega^T x_i - \omega^T x_j\right)\left(\omega^T x_i - \omega^T x_j\right)^T \\ &= \frac{1}{2}\omega^T\left(\sum_i\sum_j W^s_{ij}\,(x_i - x_j)\,(x_i - x_j)^T\right)\omega \\ &= \omega^T\left(\sum_i\sum_j D^s_{ii} x_i x_i^T - \sum_i\sum_j W^s_{ij}\, x_i x_j{}^T\right)\omega \\ &= \omega^T X\left(D^s - W^s\right)X^T\omega \end{aligned} \tag{11}$$

where

$$W^s_{ij} = \begin{cases} 1, & \text{if } i \in N^-_{k_s}(x_j) \text{ or } j \in N^-_{k_s}(x_i) \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

$N^-_{k_s}(x_i)$ indicates the index set of the k_s nearest neighbors of x_i in different classes and D^s is a diagonal matrix with diagonal elements $d^s_{ii} = \sum_j W^s_{ij}$.

Figure 1b shows the connected edges in local interclass graph when $k_s = 2$, which also shows the difference between interclass graph and penalty graph. From the form of S_s,



(a) **(b)**

Fig. 1 The graphs and the computed projections by MFA (**a**) and MLIE (**b**) when there is an outlier. **a** Penalty graph ($k_p = 2$). **b** Local interclass graph ($k_s = 2$)

we can see that it just characterizes the sum distance or the sum scatter of interclass in each local neighborhood between the classes. It can also be seen that the sample points of the local interclass graph are connected without further local nearest neighbor searching (i.e. without second step in constructing penalty graph in MFA), which is essentially different from the penalty graph in MFA algorithm. Figure 1b illustrates the projection axis computed by the proposed MLIE in the following Sect. 3.3 when $k_s = 2$. The local interclass graph not only considers the local neighborhood of the nearest data pairs, but also takes the other relationships between classes into account. As it is shown in the Fig. 1b, maximizing the local interclass graph in MLIE will yield a nice projection which is better than the one of MFA. The connected points between different classes will be mapped far apart as possible as it can in the low-dimensional space. Figure 1 also shows the robustness of the local interclass graph against penalty graph when there are outliers.

3.3 MLIE algorithm

In summary of the preceding description, the following steps provide the MLIE algorithm:

Step 1 Perform PCA transform of the data. Calculate the d_{PCA} eigenvectors $\alpha_1, \alpha_2, \ldots, \alpha_{d_{PCA}}$ corresponding to d_{PCA} largest positive eigenvalues. Let W_{PCA} denote the transformation matrix of PCA. For convenience, we still use x_i and X to denote the i-th sample and the data matrix in the PCA subspace, respectively.

Step 2 Construct the intraclass graph and the interclass graph. In the intraclass compactness graph, for each sample x_i, set the adjacency matrix W_{ij} as defined in Eq. (8). In the interclass graph, set the adjacency matrix W^s_{ij} as described in Eq. (12).

Step 3 MLIE Criterion. After embedding, we want to keep neighboring points close if they have the same label, whereas prevent data points of other classes from entering the neighborhood. With these two aspects in consideration, we arrive at the following constrained optimization problem:

$$\begin{aligned} \text{Maximize } J(\omega) &= \frac{1}{2}\sum_{ij}\left\|\omega^T x_i - \omega^T x_j\right\|^2 W^s_{ij} \\ &= \omega^T X \left(D^s - W^s\right) X^T \omega \\ \text{subject to } &\frac{1}{2}\sum_{ij}\left\|\omega^T x_i - \omega^T x_j\right\|^2 W_{ij} \\ &= \omega^T X\,(D - W)\,X^T \omega = 1, \end{aligned} \tag{13}$$

The optimization formulation essentially integrates the labels and neighbor information through the two adjacency matrices W and W^S computed in Step 2 in MLIE algorithm. Following the graph embedding framework [18], MLIE has the maximized criterion

$$\omega^* = \arg\max_{\omega} \frac{\omega^T X\,(D^s - W^s)\,X^T\omega}{\omega^T X\,(D - W)\,X^T\omega}, \tag{14}$$

The criterion of Eq. (14) is formally similar to the classical Fisher criterion since both of them are formed by Rayleigh quotients. Therefore, we can obtain its optimal solutions by solving a generalized Eigen-equation:

$$X\left(D^s - W^s\right)X^T\omega = \lambda X\,(D - W)\,X^T\omega \tag{15}$$

where λ is the generalized eigenvalue and ω is the generalized eigenvector, correspondingly. Suppose $\omega_1, \omega_2, \ldots, \omega_d$ are the generalized eigenvectors associated to the first d largest generalized eigenvalues of Eq. (15). Let $W = [\omega_1, \omega_2, \ldots, \omega_d]$ be the optimal projection matrix.

Step 4 Output the final linear projection matrix as

$$w = W_{\text{PCA}} W. \tag{16}$$

Fig. 2 Sample images of one person from Yale face database

Once we have learned the projection matrix w using the MLIE algorithm, the samples can be projected on this subspace and a suitable classifier can be used for classification.

3.4 Computational complexity of MFA and MLIE

As it is mentioned in Sect. 1, the computational complexity of MFA is high. The computational cost of MFA with k-NN penalty graph is $O(m^2n+m^2\log m+qdmk+\frac{1}{4}(k(k-1))^2n+\frac{1}{4}(k(k-1))^2\log(\frac{1}{2}k(k-1)))$. The cost with local interclass graph in MLIE is $O(m^2n+m^2\log m+qdmk)$. $O(m^2n)$ is used to calculate the pairwise 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. $O(\frac{1}{4}(k(k-1))^2n)$ is used for calculating the pairwise distance in k-nearest neighbors samples and $O(\frac{1}{4}(k(k-1))^2\log(\frac{1}{2}k(k-1)))$ for finding all the local k-nearest neighbors data pairs. The k-NN graph matrix is sparse and the Lanczos algorithm [20] can be used to efficiently compute the first d eigenvectors of the Eigen-problem in Eqs. (10) and (15) within $O(qdmk)$, where q is the number of iterations in Lanczos. Thus, MLIE's computational complexity is less than that of MFA. Particularly, when k and n is relatively large (when compared with m), MFA is significantly time consuming. This point will also be proved in Table 3 by experiments.

4 Application to face recognition: experiments and analysis

In this section, in order to evaluate the proposed MLIE algorithm, we compare it with the PCA (Eigenfaces), LDA (Fisherfaces), LPP (Laplacianfaces), neighborhood preserving embedding (NPE, the linearization of LLE) [21], MFA and Smooth MFA (SMFA) algorithm on three face databases: Yale, AR and ORL. The Yale database was used to examine the performance of MLIE when both facial expressions and illumination are varied. The AR database was employed to test the performance of the algorithm under conditions where there was a variation over time, in facial expressions, and in lighting conditions. The ORL database was used to evaluate the performance of the algorithm under conditions where the face images were taken with a tolerance for some tilting and rotation. Nearest neighbor classifier with Euclidean distance was used in all the experiments.

Table 1 The maximal recognition rates (percent) of the seven methods on the Yale database and the corresponding dimensions

Methods	PCA	LDA	LPP	NPE	MFA	SMFA	MLIE
Recognition rates (%)	92.00	93.33	93.33	93.33	94.67	96.00	98.667
Dimensions	30	14	32	16	16	17	14

4.1 Experiments using the yale database

The Yale face database (http://www.cvc.yale.edu/projects/yalefaces/yalefaces.html) contains 165 images of 15 individuals (11 different images per individual) under various facial expressions and lighting conditions. In our experiments, each image was manually cropped and resized to 50×40 pixels. Figure 2 shows sample images of one person.

We focus on the case that there are outliers in training set and test set. The experiment was performed using the first six images (i.e., center-light, with glasses, happy, left-light, without glasses, and normal) per class for training, and the remaining five images (i.e., right-light, sad, sleepy, surprised, and winking) for test. In the experiments, the left-light, right-light and surprised images can be viewed as outliers. For feature extraction, we used, respectively, PCA, LDA, LPP, NPE, MFA, SMFA and the proposed MLIE. Note that LDA, LPP, NPE, MFA and MLIE all involve a PCA phase. In this phase, we keep 89% image energy. The maximal recognition rate of each method and the corresponding dimension are given in Table 1. As it is shown in Table 1, the top recognition rate of MLIE is significantly higher than those of the other methods.

Why can MLIE significantly outperform the other algorithms? In our opinion, the reason is that the interclass graph is more robust to reflect the neighborhood and the local interclass relationships and provides more important discriminant information than the penalty graph. Moreover, the local interclass graph not only considers the local neighborhood of the nearest data point pairs, but also takes the local interclass relationships into account. Thus, MLIE is superior to

Fig. 3 Samples of the cropped images of one person on the AR face database

Table 2 The maximal recognition rates (percent) of the seven methods on the AR-7 database and the corresponding dimensions (shown in parentheses) when the first 3, 4, 5 samples per class are used for training and the remaining for test

Training sample number	PCA	LDA	LPP	NPE	MFA	SMFA	MLIE
5	75.00 (120)	95.50 (115)	96.25 (145)	95.83 (130)	97.08 (170)	97.50 (200)	98.33 (180)
4	66.67 (120)	91.33 (115)	92.22 (140)	87.22 (150)	88.06 (135)	93.89 (185)	95.83 (135)
3	73.54 (95)	94.17 (100)	93.96 (125)	94.17 (125)	93.75 (100)	94.37 (150)	96.04 (90)

MFA and SMFA. Since the locality is not taken into account and the outlier images may cause errors in estimating within-class scatter, between-class scatter and global covariance, the classical methods, i.e. PCA and LDA, obtain low recognition rates in face recognition. Furthermore, preserving the locality can not maximize the separability of different classes. As a result, the recognition rates of LPP and NPE are lower than that of MLIE.

4.2 Experiments using the AR database

The AR face database [22] 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 session contains 13 color images. The face portion of each image is manually cropped and then normalized to 50×40 pixels. 20 face images per person are used in the experiments. The sample images of one person used in the experiments are shown in Fig. 3. These images vary in the following order: neutral expression, smiling, angry, screaming, left light on, right light on, all sides light on, wearing sun glasses, wearing sun glasses and left light on, wearing sun glasses and right light on. The first seven images in the first session are selected to form the subset, called AR-7, which is used in these experiments in current subsection. The 20 images per person in both sessions are selected to form the subset, called AR-20, which will be used for random experiments in Sect. 4.4.

On the AR-7 face database, the first l ($l = 3$, 4 and 5) images per individual are selected to form the training sample set. The remaining $7 - l$ images per individual are used for test. For each l, PCA, LDA, LPP, MFA, SMFA and MLIE are, respectively, used for face recognition. In the PCA phase of LDA, LPP, MFA and MLIE, the energy is set to be above 95%. Finally, a nearest-neighbor classifier is employed for classification. The maximal recognition rates and the dimensions are shown in Table 2. From Table 2, we can see that MLIE significantly outperforms MFA, SMFA and LDA. As a supervised method, MLIE is more robust than MFA, SMFA and LDA when there are different facial expressions and lighting conditions, irrespective of the variations in training sample size and dimensions.

Moreover, it should be noted that the recognition rates of MFA is greatly degraded when the training number is 4 (please see Table 2). The reason may be that when there are no lighting samples in training set the local neighbor information cannot reflect or predict the neighbor relationships with the lighting samples in test set in a certain sense. However, MLIE is superior to MFA for characterizing the interclass relationships. This also shows that the interclass graph is more robust and reasonable to reflect the neighboring relationships and provides more important discriminant information than the penalty graph in MFA can do. With the interclass graph, a new test point can be more reliably predicted by the nearest neighbor criterion, owing to the locally maximized interclass discriminating property. Thus, the experimental results are consistent with the ones in Sect. 4.1 and also show that MLIE is more robust than MFA when there are lighting variations.

Table 3 Comparisons of CPU time (S) for feature extraction using AR-7 database (CPU: Pentium IV 3.2 GHz, RAM: 512 Mb)

#Training samples/class	3	4	5
MFA	2.812	4.14	8.25
MLIE	1.782	4.00	8.172

Table 4 The maximal recognition rates (percent) of the seven methods on the ORL database and the corresponding dimensions when the first five samples per class are used for training and the remaining for test

Methods	PCA	LDA	LPP	NPE	MFA	SMFA	MLIE
Recognition rates (%)	89.00	92.50	93.50	93.50	94.00	95.00	96.00
Dimensions	34	33	46	42	50	48	42

Another consistent observation from Tables 1 and 2 is that SMFA is also superior to MFA.

The MLIE method is also superior to MFA in terms of computational efficiency for feature extraction. The comparisons of CPU Time corresponding to the top recognition rate for feature extraction using AR-7 database are show in Table 3. Table 3 shows that MLIE takes less time in feature extraction. The main reason is that the k_p nearest sample pairs are found by a global searching among all the sample pairs in MFA algorithm. However, MLIE avoids these steps and constructs the interclass graph directly.

4.3 Experiments using the ORL database

The ORL database (http://www.uk.research.att.com/ face-database.html) is used to evaluate the performance of MLIE under conditions where the pose and the face expression vary. The ORL face database contains 400 images from 40 individuals with 10 different images per individual. 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°. Moreover, there is also some variation in the scale of up to about 10%. All images normalized to a resolution of 56×46. Sample images of one person are shown in Fig. 4. The first five images per individual are selected to form the training sample set. The remaining five images per individual are used for test. Note that LDA, LPP, NPE, MFA, and MLIE all involve a PCA phase. The experimental results are shown in Table 4. From this experiment, we find that the top recognition rates of MLIE, MFA and SMFA are significantly higher than PCA and LDA on ORL face database. The reason is that the local neighbor relationships can provide important discriminant information. The results also show that MLIE is superior to the other methods because of the robustness of the local interclass graph.

4.4 Random experiments

We performed random experiments on the three databases mentioned above, e.g. Yale, AR-20 and ORL. A random subset with T $(T = 2, 3, 4, 5)$ training samples per individual was selected to form the training set and the rest of the database was considered to be the test set in each database. 50 independent runs were performed on the Yale and the ORL face databases, and 10 independent runs were performed on the AR-20 face database. For ease of representation, Gm/Pn means m images per person are randomly selected for training and the remaining n images are for test. Table 5 summarizes the results of these experiments. All the recognition rates were empirically optimized by choosing the best parameters within a wide range for each method. We compared the recognition rates of PCA, LDA, LPP, MFA, SMFA and MLIE in low-dimensional feature space using the nearest neighbor classifier. In order to avoid singularities encountered in those methods that require solving generalized eigenvalue problems, we use PCA to pre-process the data and keep above 95% image energy. The experimental results are shown in Table 5.

From Table 5, we can see that MLIE achieves the best performances in experiments. Since the constructed graphs are different and the penalty graph is sensitive to outliers, SMFA is slightly less effective than MLIE. Moreover, SMFA is shown to be superior to MFA due to its smoothness, which is consistent with the conclusion presented in [19].

5 Conclusions and future work

This paper develops a new approach for dimensionality reduction of high dimensional data, which is called MLIE. The projection of MLIE can be viewed as a linear approximation of the nonlinear map that uncovers and separates

Fig. 4 Sample images of one person from ORL face database

Table 5 The average recognition rates (percent) and the corresponding dimensions (shown in parentheses) on the three face databases

Database	Method	PCA	LDA	LPP	NPE	MFA	SMFA	MLIE
Yale	G2/P9	78.49 (29)	81.93 (14)	81.45 (22)	81.62 (30)	82.10 (28)	82.62 (28)	83.01 (29)
	G3/P8	81.47 (40)	85.61 (14)	85.97 (24)	86.23 (39)	86.47 (35)	87.07 (38)	87.88 (40)
	G4/P7	85.26 (37)	88.30 (14)	88.57 (21)	89.01 (33)	89.67 (28)	89.86 (39)	90.34 (40)
	G5/P6	85.96 (40)	88.84 (14)	89.00 (18)	90.24 (35)	91.47 (38)	91.89 (39)	92.27 (39)
AR-20	G2/P18	67.79 (150)	72.21 (119)	72.45 (140)	71.39 (150)	72.56 (150)	73.09 (145)	73.91 (150)
	G3/P17	71.83 (150)	76.34 (119)	76.39 (130)	75.52 (120)	77.83 (150)	78.47 (140)	81.64 (150)
	G4/P16	78.87 (150)	83.84 (119)	84.39 (130)	84.14 (135)	86.48 (145)	87.46 (140)	88.83 (145)
	G5/P15	79.84 (150)	87.45 (119)	88.68 (110)	88.31 (145)	90.34 (135)	90.64 (145)	91.28 (150)
ORL	G2/P8	74.91 (50)	77.40 (39)	72.05 (48)	72.46 (49)	77.72 (50)	80.79 (47)	81.46 (50)
	G3/P7	82.23 (46)	85.09 (39)	81.78 (46)	81.23 (46)	85.32 (47)	88.04 (50)	89.65 (45)
	G4/P6	84.53 (34)	86.17 (39)	87.42 (36)	86.74 (49)	90.59 (48)	93.06 (49)	93.69 (48)
	G5/P5	86.71 (46)	87.23 (35)	90.82 (34)	90.56 (49)	93.69 (44)	96.08 (48)	96.35 (35)

embeddings corresponding to different manifolds in the final embedding space. MLIE seeks to find a projection that not only maximizes the compactness of intraclass but also maximizes the local interclass scatter at the same time. This characteristic makes MLIE more powerful to find the intrinsic low-dimensional nonlinear data structures for classification tasks than LDA and MFA. The proposed MLIE is also superior to MFA in terms of computational efficiency and robust for feature extraction. Our experimental results on three popular face image databases demonstrate that MLIE is an effective method for face recognition. Although this paper only focuses on linear dimensionality reduction techniques, it should be figured out that the graph embedding framework with different graphs can be generalized to kernel methods [10,11]. In the future, we plan to study the variant extensions of the graph embedding framework using kernel technique such as kernel ridge regression [12].

Acknowledgments This work is partially supported by the National Science Foundation of China under grant No. 60632050, 60705006, 60873151, 60973098, 61005005 and Hi-Tech Research and Development Program of China under grant No.2006AA01Z119.

References

1. Batur, A., Hayes, M.: Linear Subspaces for Illumination Robust Face Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01), vol. 2, pp. 296–301. Kauai Marriott, Hawaii, 8–14 Dec 2001
2. Yang, J., Zhang, D., Yang, J.: Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics. IEEE Trans. Pattern Anal. Mach. Intell. **29**(4), 650–664 (2007)
3. Jain, A.K., Duin, R.P.W., Mao, J.: Statistical pattern recognition: a review. IEEE Trans. Pattern Anal. Mach.Intell. **22**(1), 4–37 (2000)
4. Joliffe, I.: Principal Component Analysis. Springer, Berlin (1986)
5. Martinez, A.M., Kak, A.C.: PCA versus LDA. IEEE Trans. Pattern Anal. Mach. Intell. **23**(2), 228–233 (2001)
6. Turk, M., Pentland, A.: Face Recognition Using Eigenfaces. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'91), pp. 586–591. Maui, Hawaii, 3–6 June 1991
7. Fukunnaga, K.: Introduction to Statistical Pattern Recognition. 2nd edn. Academic Press, San Diego (1991)
8. Ye, J., Janardan, R., Park, C., Park, H.: An optimization criterion for generalized discriminant analysis on under-sampled problems. IEEE Trans. Pattern Anal. Mach. Intell. **26**(8), 982–994 (2004)
9. Yu, H., Yang, J.: A direct LDA algorithm for high dimensional data-with application to face recognition. Pattern Recognit. **34**(10), 2067–2070 (2001)
10. Schölkopf, B., Smola, A., Muller, K.R.: Nonlinear component analysis as a Kernel Eigenvalue problem. Neural Comput. **10**(5), 1299–1319 (1998)
11. Baudat, G., Anouar, F.: Generalized discriminant analysis using a Kernel approach. Neural Comput. **12**(10), 2385–2404 (2000)
12. An, S., Liu, W., Venkatesh, S.: Face recognition using Kernel ridge regression. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'07). Minneopolis, Minnesota, 17–22 June 2007
13. Tenenbaum, J.B., deSilva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**, 2319–2323 (2000)
14. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**, 2323–2326 (2000)
15. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Proceedings of the 15th Annual Conference on Neural Information Processing Systems, Vancouver, Canada, 3–6 Dec 2001
16. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.: Face recognition using laplacianfaces. IEEE Trans. Pattern Anal. Mach. Intell. **27**(3), 328–340 (2005)
17. He, X., Niyogi, P.: Locality preserving projections. In: Proceedings of the 17th Annual Conference on Neural Information Processing Systems, Vancouver and Whistler, Canada, 8–13 Dec 2003
18. Yan, S., Xu, D., Zhang, B., Zhang, H.-J.: Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans. Pattern Anal. Mach. Intell. **29**(1), 40–51 (2007)
19. Cai, D., He, X., Hu, Y., Han, J., Huang, T.: Learning a Spatially Smooth Subspace for Face Recognition. IEEE Computer

Society Conference on Computer Vision and Pattern Recognition (CVPR'07). Minneapolis, Minnesota, 17–22 June 2007
20. Golub, G.H., Loan, C.F.V.: Matrix Computations. 3rd edn. Johns Hopkins University Press, Baltimore (1996)
21. He, X., Cai, D., Yan, S., Zhang, H.: Neighborhood preserving embedding. In: Proceedings of the 10th IEEE International Conference on Computer Vision, pp. 1208–1213. Beijing, China, 17–20 October 2005
22. Martinez, A.M., Benavente, R.: The AR Face Database. CVC Technical Report #24, June 1998

Author Biography

Zhihui Lai received his BS degree in mathematics from the South China Normal University and MS degree from the Ji'nan University, China, in 2002 and 2007, respectively. He is currently pursuing his PhD degree in the School of Computer Science from the Nanjing University of Science and Technology (NUST). His research interests include face recognition, image processing and content-based image retrieval, pattern recognition, compressive sense, human vision modelization and applications in the fields of intelligent robot research. E-mail: lai_zhi_hui@163.com

Cairong Zhao received his BS degree in electronics and information scientific technology from the Jilin University, China, in 2003 and his MS degree in electronic circuit and system from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, in 2006, respectively. He is currently pursuing his PhD degree in computer science from the Nanjing University of Science and Technology. His research interests include face recognition, building recognition and vision attention. E-mail:cairong.zhao@yahoo.com

Yi Chen received his BS degree in the School of Computer Science and Technology, Nanjing University of Science and Technology (NUST), PR China, in 2003. Currently, he is pursuing his PhD degree in pattern recognition and intelligent systems from the Nanjing University of Science and Technology. His current research interests include pattern recognition, computer vision and image processing. E-mail: cystory@qq.com

Zhong Jin received his BS degree in mathematics, MS degree in applied mathematics and PhD degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology (NUST), China, in 1982, 1984 and 1999, respectively. He is a professor in the Department of Computer Science, NUST. He had a stay of 15 months as research assistant at the Department of Computer Science and Engineering, Chinese University of Hong Kong from 2000 to 2001. He visited the Laboratoire HEUDIASYC, Universite de Technologie de Compiegne, France, from October 2001 to July 2002. He was at the Centre de Visio per Computador, Universitat Autonoma de Barcelona, Spain, as the Ramon y Cajal Program Research Fellow from September 2005 to October 2005. His current interests are in the areas of pattern recognition, computer vision, face recognition, facial expression analysis and content-based image retrieval. E-mail: zhongjin@mail.njust.edu.cn