Artif Intell Rev
DOI 10.1007/s10462-011-9214-1

Class mean embedding for face recognition

Minghua Wan · Guowei Yang · Wei Huang · Zhong Jin

Abstract Recently, local discriminant embedding (LDE) was proposed as a means of addressing manifold learning and pattern classification. In the LDE framework, the neighbor and class of data points are used to construct the graph embedding for classification problems. From a high dimensional to a low dimensional subspace, data points of the same class maintain their intrinsic neighbor relations, whereas neighboring data points of different classes no longer stick to one another. But, neighboring data points of different classes are not deemphasized efficiently by LDE and it may degrade the performance of classification. In this paper, we investigate its extension, called class mean embedding (CME), using class mean of data points to enhance its discriminant power in their mapping into a low dimensional space. After joined class mean data points, (1) CME may cause each class of data points to be more compact in the high dimension space; (2) CME may increase the quantity of data points, and solves the small sample size (SSS) problem; (3) CME may preserve well the local geometry of the data manifolds in the embedding space. Experimental results on ORL, Yale, AR, and FERET face databases show the effectiveness of the proposed method.

Keywords Local discriminant embedding (LDE) · Manifold learning · Graph embedding · Pattern classification · Small sample size (SSS)

M. Wan (✉) · G. Yang
School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China
e-mail: wmh36@sina.com

M. Wan · Z. Jin
School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China

W. Huang
Department of Math and Information Technology, Hanshan Normal University, Chaozhou 521041, China

Published online: 21 April 2011

1 Introduction

Face recognition has attracted wide attention of the researchers in biometric authentication. In face recognition, 2D face images are usually transformed into 1D vector through column by column or row by row concatenation. The resulting image vectors of faces usually lead to a high dimensional image vector space, where it is difficult to evaluate the covariance matrix accurately due to its large size and the relatively small number of training samples. Two of the most fundamental dimensionality reduction methods are principal component analysis (PCA) (Turk and Pentland 1991) and linear discriminant analysis (LDA) (Belhumeur et al. 1997.)

PCA is a classical feature extraction and data repesentation technique widely used in the areas of pattern recognition and computer vision. PCA aims to find a linear mapping, which preserves total variance by maximizing the trace of feature variance. The optimal mapping is the leading eigenvectors of the data's total variance matrix associated with the leading eigenvalues. Thus, PCA preserve the total variance by maximizing the trace of feature variance, but PCA cannot preserve local information due to pursuing maximal variance.

LDA is used to find the optimal set of projection vectors that maximize the determinant of the between-class scatter matrix and at the same time minimize the determinant of the within-class scatter matrix (Belhumeur et al. 1997). But, the dimension of vectors is high and the number of observations is small, usually tens or hundreds of samples. An intrinsic limitation of traditional LDA is that it fails to work when the within-class scatter matrix becomes singular. This is known as the small sample size (SSS), the undersampled or singularity problem. Furthermore, class discrimination in LDA is based upon interclass and intraclass scatters, which is optimal only in cases where the data of each class is approximately Gaussian distributed, a property that cannot always be satisfied in real-world applications.

Recently, Yan et al. (2007) proposed a newly general framework called graph embedding for dimensionality reduction, from which many algorithms, such as PCA, LDA, LPP (He and Niyogi 2003), ISOMAP (Tenenbaum et al. 2000), LLE (Roweis and Saul 2000), Laplacian Eigenmap (Belkin and Niyogi 2003), MFA (Yan et al. 2007), LDE (Chen et al. 2005) can all be reformulated. Using the graph embedding framework as a platform, they developed a novel dimensionality reduction algorithm, They all utilize local neighborhood information to construct a global embedding of the manifold. Locality Preserving Projections (LPP) has been proposed for dimensionality reduction that can preserve local relationships within the data set that lies on a lower dimensional manifold. Other nonlinear methods, such as isometric feature mapping (ISOMAP), local linear embedding (LLE), and Laplacian Eigenmap, have been proposed to find the intrinsic low-dimensional nonlinear data structures hidden in observation space. However, Current manifold learning algorithms might be unsuitable for pattern recognition tasks in that they concentrate on representing the high-dimensional data with low-dimensional data instead of classification or that they only considered the locality and couldn't give a clear nonlinear map when applied to a new sample, such as ISOMAP and LLE. More recently, Local discriminant embedding (LDE) and marginal fisher analysis (MFA) were proposed to overcome the drawbacks of LPP. LDE and MFA were developed by different researchers, but the underlying ideas of which are almost the same: the neighbor and class relations of data are utilized to construct the face space (subspace of the image space). Compared with LDA, MFA and LDE do not depend on the assumption that the data of each class is Gaussian distributed.

Focusing on manifold learning and pattern classification, LDE achieves good discrimination performance by integrating the information of neighbor and class relations between data points. LDE incorporates the class information into the construction of embedding and derives the embedding for nearest-neighbor classification in a low-dimensional space, which learns the embedding for the submanifold of each class by solving an optimization problem. Nevertheless, distant points are not deemphasized efficiently by LDE and it may degrade the performance of classification. In this paper, we investigate its extension, called class mean embedding (CME), using class mean of data points to enhance its discriminant power in their mapping into a low dimensional space. The crux of our approach is to learn CME by taking account of the respective submanifold of each class. While maintaining the original neighbor relations for neighboring data points of the same class is important, it is also crucial to differentiate and to keep away neighboring data points of different classes after the CME. With that, the class of a new test point can be more reliably predicted by the nearest neighbor criterion, owing to the locally discriminating nature. The structure of CME can be characterized by three key ideas: After joined class mean data points, (1) CME may cause each class of data points to be more compact in the high dimension space; (2) CME may make data points quantity to increase, and solves the small sample size (SSS) problem; (3) CME may preserve well the local geometry of the data manifolds in the embedding space.

The rest of the paper is structured as follows: In Sect. 2 we introduce PCA, LDA and LDE. In Sect. 3, we propose the idea and describe CME in detail. In Sect. 4, experiments on ORL, Yale, AR, and FERET face database are presented to demonstrate the effectiveness of CME. Finally, we give concluding remarks and a discussion of future work in Sect. 5.

2 Outline of PCA, LDA and LDE

Let us consider a set of m sample $\{x_1, x_2, \ldots, x_m\}$ taking values in an n-dimensional image space, and assume that each image belongs to one of c classes. Let us also consider a linear transformation mapping the original n-dimensional space into an d-dimensional feature space, where $n > d$. The new feature vectors $y_k \in R^d$ are defined by the following linear transformation:

$$y_k = U^T x_k, \quad k = 1, \ldots, m \tag{1}$$

where $U \in R^{n\times d}$ is a transformation matrix.

2.1 Principal component analysis (PCA)

A fundamental unsupervised dimensionality reduction method is principal component analysis (PCA). Let S_T be the total scatter matrix:

$$S_T = \sum_{i=1}^{N} (x_i - \overline{x})(x_i - \overline{x})^T \tag{2}$$

where $\bar{x}$ is the mean of all the training samples. The PCA transformation matrix is defined as

$$U_{opt} = \arg\max_u \left[tr\left(U^T S_T U \right) \right] = [u_1,\ u_2, \ldots, u_d] \tag{3}$$

where $u_i (i = 1, 2, \ldots, d)$ is the eigenvector corresponding to the largest eigenvalue of S_T.

2.2 Linear discriminant analysis (LDA)

LDA is a supervised learning algorithm. Let l denote the total class number and l_i denote the number of training samples in the ith class. Let x_i^j, denote the jth sample in ith class, $\overline{x}$ be the mean of all the training samples, $\overline{x}_i$ be the mean of the ith class. The between-class and within-class scatter matrices can be evaluated by:

$$S_b = \sum_{i=1}^{l} l_i (\overline{x}_i - \overline{x})(\overline{x}_i - \overline{x})^T \tag{4}$$

$$S_w = \sum_{i=1}^{l} \sum_{j=1}^{l_i} \left(x_i^j - \overline{x}_i\right)\left(x_i^j - \overline{x}_i\right)^T \tag{5}$$

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

$$U_{opt} = \arg\max_U \frac{\left|U^T S_b U\right|}{\left|U^T S_w^T U\right|} = [u_1,\ u_2,\ \ldots\ , u_d] \tag{6}$$

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

$$S_b u = \lambda_i S_w u, \quad i = 1, 2, \ldots, d. \tag{7}$$

Note that there are at most $c - 1$ non-zero generalized eigenvalues.

2.3 Local discriminant embedding (LDE)

LDE is a supervised subspace learning algorithm. Therefore, class label l_i of $x_i\,(i = 1, \ldots, m)$ are used in LDE to determine a linear transformation matrix U such that:

$$y_i = U^T x_i \tag{8}$$

The column vectors of $U = [u_1, u_2, \ldots, u_d]$ span a d -imensional subspace. The aim of LDE is, in the low subspace, to keep neighboring points close if they have the same class label, whereas to prevent points of other classes from entering the neighborhood.

Its objective is to maximize the function

$$J_{LDE}\,(U) = \sum_{i,j} \left\|U^T x_i - U^T x_j\right\|^2 w'_{ij} \tag{9}$$

subject to

$$\sum_{i,j} \left\|U^T x_i - U^T x_j\right\|^2 w_{ij} = 1 \tag{10}$$

where

$$w'_{ij} = \begin{cases} \exp\left(-\left\|x_i - x_j\right\|^2/t\right), & \text{if } l_i \neq l_j \text{ and } i \in N_K^+(j) \text{ or } j \in N_K^+(i) \\ 0, & \text{else} \end{cases} \tag{11}$$

$$w_{ij} = \begin{cases} \exp\left(-\left\|x_i - x_j\right\|^2/t\right), & if\ l_i = l_j \text{ and } i \in N_K^+(j) \text{ or } j \in N_K^+(i) \\ 0, & \text{else} \end{cases} \tag{12}$$

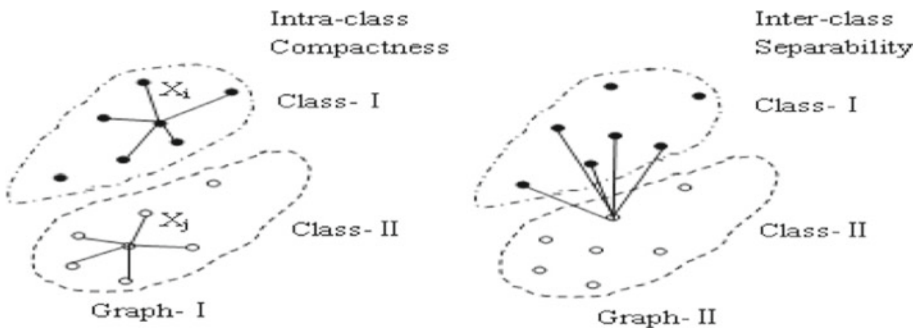

Fig. 1 The adjacency relationships of the intrinsic and penalty graphs for the local disriminant embedding algorithm (LDE)

where $N_K^+(j)$ or $N_K^+(i)$indicates the index set of the K nearest neighbors of the sample X_i or X_j.

The optimization can be reduced to the following generalized eigenvalue problem:

$$X\ (D' - W')\ X^T u = \lambda X\ (D - W)\ X^T u \tag{13}$$

where the elements of the matrix W' arew'_{ij}, the elements of the matrix W are w_{ij}. The elements of diagonal matrices D and D' are defined as $d_{ii} = \sum_j w_{ij}$ and $d'_{ii} = \sum_j w_{ij}{}'$ respectively.

3 The proposed class mean embedding (CME)

Suppose there are c known pattern classes, $w_1, w_2, \ldots, w_c$, where m is the total number of training samples, and m_i is the number of training samples in class i. In class i, the jth training sample is denoted by x_i^j, the mean vector of training samples in class i is denoted by $\overline{m}_i$.

First, we describe the steps of the CME algorithm, and then justify them in detail. Recall that the data points $\{x_i\}_{i=1}^{m+c}$ are in $\Re^n$, and each x_i is labeled by some class label y_i. We also write the data matrix as $X = \left[x_1 x_2 \cdots x_n \cdots x_{m+c}\right] \in R^n$. Figures 1 and 2 show the adjacency relationships of the intrinsic and penalty graphs, which respectively represent the local discriminant embedding algorithm and the class mean embedding algorithm. Studying the comparison embedding of LDE graphs and CME graphs, we discover CME graphs to be joined class mean based LDE graphs. In CME graphs, two each kind of data points revolves around the respective class mean, which causes between each kind of data points the distance more compact in the high dimension space and solves the small sample size problems. After joined a kind of class mean, with each kind of data points increases, and the data points is more compact in the high dimensional space, which causes the sampled data to tend the manifold distribution in the high dimensional space.

Then, the proposed CME can be realized by the following three steps.

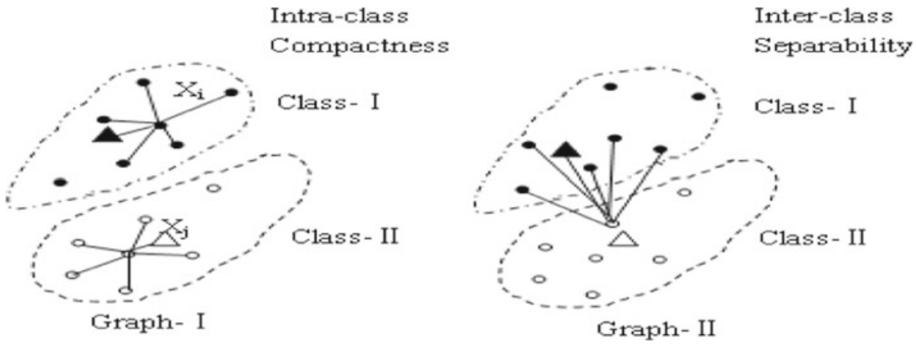

Fig. 2 The adjacency relationships of the intrinsic and penalty graphs for the class mean embedding algorithm (CME)

1. *Construct neighborhood graphs.* Let G and G' denote two (undirected) graphs both over all data points. To construct

$$G, G : \text{if} \left\{ \begin{array}{l} y_i = y_j \\ (i, j) \in m_i \\ i \in N_K^+(j) \text{ or } j \in N_K^+(i) \end{array} \right\} \tag{14}$$

For G',

$$G' : \text{if} \left\{ \begin{array}{l} y_i \neq y_j \\ (i, j) \in m_i \\ i \in N_K^+(j) \text{ or } j \in N_K^+(i) \end{array} \right\} \tag{15}$$

where $N_K^+(j)$ or $N_K^+(i)$ indicates the index set of the K nearest neighbors of the sample X_i in the same class,

2. *Compute affinity weights.* Specify the affinity matrix W of G,

$$w_{ij}^G = \begin{cases} \exp\left(- \left\|x_i - x_j\right\|^2 / t\right), & i \in N_K^+(j) \text{ or } j \in N_K^+(i) \\ 0, & \text{else} \end{cases} \tag{16}$$

The other affinity matrix $W^{G'}$ of G' can be computed in the same way. Of note, the affinity weights defined in (16) are derived from the heat kernel.

Another possible choice called "simple-minded" is also suggested in:

$$w_{ij}^G = \begin{cases} 1, & i \in N_K^+(j) \text{ or } j \in N_K^+(i) \\ 0, & \text{else} \end{cases} \tag{17}$$

3. *Complete the embedding.* Find the generalized eigenvectors $u_1, u_2, \ldots, u_d$ that correspond to the d largest eigenvalues in:

$$X\ (D^{G'} - W^{G'})\ X^T u = \lambda X\ (D^G - W^G)\ X^T u \tag{18}$$

where D^G and $D^{G'}$ are diagonal matrices with diagonal elements

$$\begin{cases} d_{ii}^G = \sum_j\ w_{ij}^G \\ d_{ii}^{G'} = \sum_j\ w_{ij}^{G'} \end{cases} \tag{19}$$

Fig. 3 Sample images of one person in the ORL face database

The embedding of x_i is accomplished by

$$z_i = U^T x_i, \quad (i = 1, 2, \ldots, m) \tag{20}$$

where $U = [u_1, u_2, \ldots, u_d]$

Once we have learned the projection matrix U using the CME algorithm, nearest neighbor classifications become straightforward.

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

Given two images x_1, x_2 represented by FCME 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:

$$d(y_1, y_2) = \sum_{k=1}^{d} \left\| y_1^k - y_2^k \right\| \tag{21}$$

If the feature matrices of training images are $y_1, y_2, \ldots, y_m$, and each image is assigned to a class C_i. Then for a given test image y, if $d(y, y_l) = \min_j d(y, y_j)$ and $y_l \in C_i$, the resulting decision is $y \in C_i$.

4 Experiments and results

To evaluate the proposed CME algorithm, we systematically compare it with the PCA, LDA and LDE algorithm in real-work face databases: ORL, Yale, AR, and FERET. The ORL database was used to evaluate the performance of CME under conditions where the pose and sample size are varied. The Yale database was used to examine the system performance when both facial expressions and illumination are varied. The AR database was employed to test the performance of the system under conditions where there is a variation over time, in facial expressions, and in lighting conditions. The FERET face databases involves variations in facial expression, illumination and pose. 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.

4.1 Experiment on the ORL database

The ORL database is used to evaluate the performance of CME under conditions where the pose, face expression and sample size vary. The ORL face database contains images from 40 individuals, each providing 10 different images. 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 the scale of up to about 10 percent. All images normalized to a resolution of 56×46. Figure 3 shows sample images of one person from ORL face database.

Table 1 The maximal average recognition rates (%) of PCA, LDA, LDE, CME on the ORL database and the corresponding dimensions (shown in parentheses) when the 3, 4, 5, 6 samples per class are randomly selected for training and the remaining for testing

Training sample number	PCA	LDA	LDE	CME
6	87.39 (46)	89.38 (38)	98.31 (36)	99.18 (40)
5	86.71 (46)	87.23 (38)	96.52 (46)	98.80 (30)
4	84.98 (46)	86.17 (38)	93.66 (36)	96.76 (34)
3	82.27 (46)	85.09 (38)	89.56 (36)	94.39 (26)

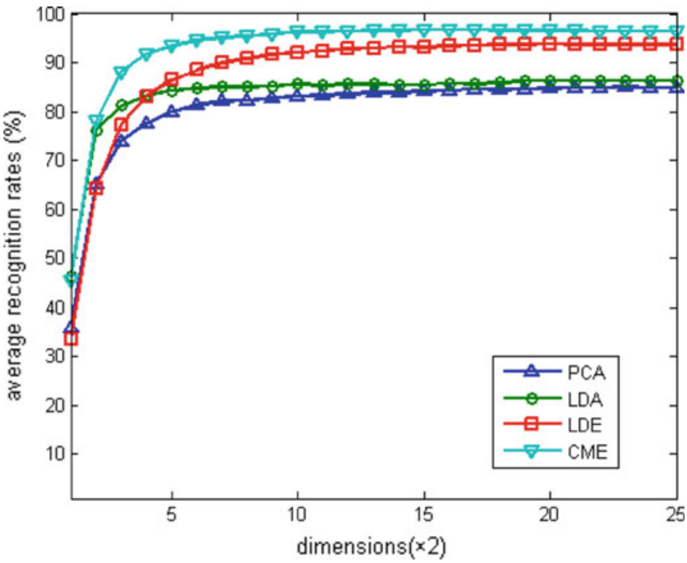

Fig. 4 The average recognition rates (%) of PCA, LDA, LDE and CME versus the dimensions when the 4 images per class were randomly selected for training on the ORL face database

Now, we test the recognition performances of the four methods: PCA, LDA, LDE and CME. In the experiments, l images (l varies from 3 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 the system 50 times. The maximal average recognition rates (percent) of PCA, LDA, LDE, CME on the ORL database and the corresponding dimensions (shown in parentheses) when the 3, 4, 5, 6 samples per class are used for training and the remaining for testing. The results are show in Table 1. From this experiment, we find that CME can achieve higher recognition rate on ORL face database when the training sample number is small. Table 1 presents the top average recognition accuracy of the methods, which corresponds to different number of images per person used for training. The performance of CME is better than PCA, LDA and LDE. The values in parentheses denote the number of eigenvectors for the best recognition accuracy. In the PCA phase of LDA, LDE and CME, we keep 90 percent image energy.

Figure 4 showed the variation of accuracy along different number of eigenvectors used and the recognition accuracy when the four images per class are randomly selected for train-

Fig. 5 Sample images of one person in the Yale database

Table 2 The maximal average recognition rates (%) of PCA, LDA, LDE, CME on the Yale database and the corresponding dimensions when the 3, 4, 5, 6 samples per class are randomly selected for training and the remaining for testing

Training sample number	PCA	LDA	LDE	CME
6	87.01 (46)	89.36 (14)	93.95 (24)	97.76 (24)
5	85.96 (40)	88.84 (14)	92.89 (20)	96.69 (20)
4	85.37 (36)	88.30 (14)	90.48 (18)	95.68 (18)
3	81.47 (40)	85.61 (14)	87.25 (16)	94.37 (16)

ing. From Fig. 4 we can see that CME performs always better than the other three methods. The figures also demonstrate that the performance of the proposed method outperforms the other methods under the same conditions, and it further shows that the proposed method can extract more discriminative features than the other methods.

4.2 Experiment on the Yale database

The Yale face database contains 165 images of 15 individuals (each person providing 11 different images) under various facial expressions and lighting conditions. In our experiments, each image was manually cropped and resized to 100×80 pixels. For computational effectiveness, we down sample 100×80 to 50×40 in this experiment. In the PCA phase of LDA, LDE and CME, we keep 90 percent image energy. Figure 5 shows sample images of one person.

For feature extraction, we used, respectively, PCA (eigenface), LDA (Fisherface), LDE and the proposed FLDE. In the experiments, l images (l varies from 3 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 the system 50 times. The maximal average recognition rate of each method and the corresponding dimension are given in Table 2 when the 3, 4, 5, 6 samples per class are randomly selected for training and the remaining for testing. Table 2 presents the top average recognition accuracy of the four methods, which corresponds to different number of images per person used for training. The values in parentheses denote the number of eigenvectors for the best recognition accuracy.

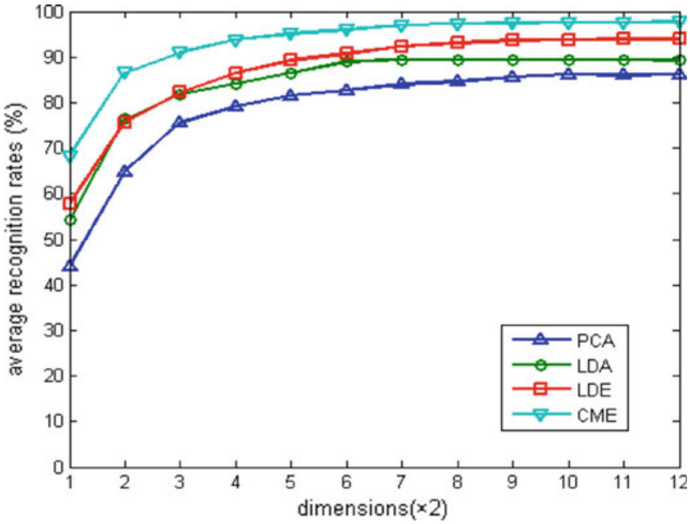

Fig. 6 The average recognition rates (%) of PCA, LDA, LDE and CME versus the dimensions when the 6 images per class were randomly selected for training on the Yale face database

Fig. 7 Sample images of one subject of the AR database. The first line and the second line images were taken in different time (separated by 2 weeks)

Several methods are tested: PCA, LDA, LDE and the proposed CME. Figure 6 showed average recognition rates (%) of PCA, LDA, LDE, CME versus the dimensions when the 5 images per person were used for training on the Yale face database. From Fig. 6 we can see that CME performs always better than the other three methods.

4.3 Experiment on the AR face database

The AR face database 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 2 weeks) and each section contains 13 color images. The face portion of each image is manually cropped and then normalized to 50 × 40 pixels. The sample images of one person are shown in Fig. 7.

In this experiment, l images (l varies from 3 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 the system 10 times. In the PCA phase of

Table 3 The maximal average recognition rates (%) of PCA, LDA, LDE, CME on the subset 1 of AR database and the corresponding dimensions (shown in parentheses) when the 3, 4, 5, 6 samples per class are randomly selected for training and the remaining for testing

Training sample number	PCA	LDA	LDE	CME
6	82.58 (150)	88.33 (115)	96.07 (80)	96.96 (130)
5	79.84 (150)	87.45 (115)	91.86 (85)	95.34 (100)
4	78.87 (150)	83.84 (115)	90.47 (85)	90.78 (100)
3	71.83 (150)	76.34 (115)	82.09 (85)	87.83 (130)

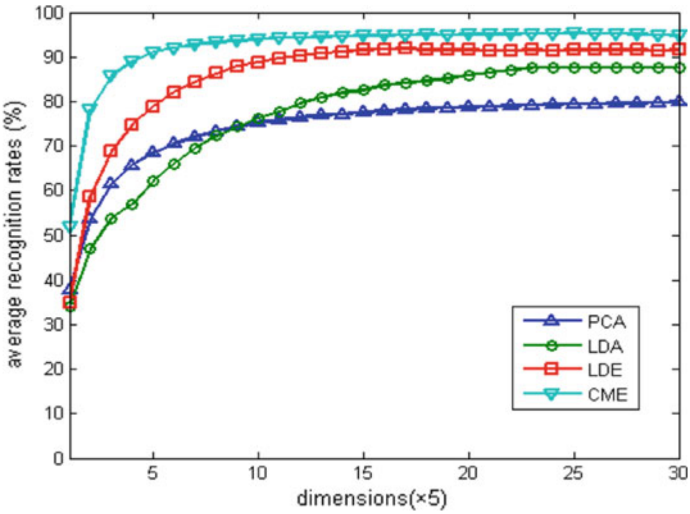

Fig. 8 The average recognition rates (%) of PCA, LDA, LDE and CME versus the dimensions when the 5 images per class were randomly selected for training on the subset 1 of AR face database

LDA, LDE and CME, 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 5 methods. The maximal recognition rate of each method and the corresponding dimension are shown in Table 3.

From the Fig. 8, it is observed that the proposed method outperformed PCA method, LDA and LDE methods comprehensively. It is observed from the graph that the proposed method's performance is far better than other three methods.

4.4 Experiments on the FERET face database

The proposed method is evaluated on a subset of FERET database, which includes 1,400 images of 200 distinct subjects, each subject has seven images. The subset involves variations in facial expression, illumination and pose. In our experiment, the facial portion of each original image is cropped automatically based on the location of eyes and resized to 40 × 40 pixels. Some facial portion images of one person are shown in Fig. 9

In this experiment, l images (l varies from 3 to 6) are randomly selected from the image gallery of each individual to form the training sample set. The remaining $7 - l$ images are

Fig. 9 Samples of the cropped images from FERET database

Table 4 The maximal average recognition rates (percent) of PCA LDA, LDE and CME on the FERET face database and the corresponding dimensions (shown in parentheses) when 3, 4, 5, 6 samples per class are selected for training and the remaining for testing

Training sample number	PCA	LDA	LDE	CME
6	61.80 (150)	85.95 (45)	87.45 (40)	99.40 (105)
5	60.38 (150)	77.60 (40)	79.78 (45)	88.85 (45)
4	56.75 (150)	72.98 (45)	74.92 (40)	76.18 (110)
3	47.73 (145)	62.31 (40)	63.56 (50)	69.56 (100)

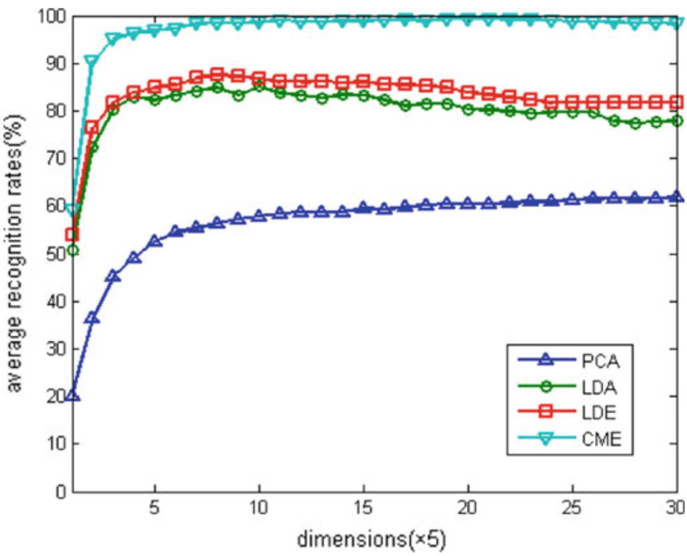

Fig. 10 The average recognition rates (%) of PCA, LDA, LDE and CME versus the dimensions when the 6 images per class were randomly selected for training and the remaining 1 images per class for testing on the FERET face database

used for testing. For each l, we independently rum the system 10 times, and PCA, LDA, LDE and CME are, respectively, used for feature extraction. The number of principal components is set to 150. The dimension step is set to be 5. After feature extraction, a nearest-neighbor classifier is employed for classification. The maximal recognition rate of each method and the corresponding dimension are shown in Table 4. The recognition rate curves versus the variation of dimensions are shown in Fig. 10. Once more, we can see that CME significantly outperforms other methods when there are Different facial expressions and lighting conditions, irrespective of the variation in training sample size and dimensions. Therefore, CME can obtain useful information for discrimination based on modeling the embedding process for the analysis of high-dimensional data set.

Figure 10 shows the recognition rate achieved by PC A, LDA, LDE and CME methods respectively. It is observed that from the graph CME method has obtained good recognition rate even for less number of training samples and less number of principal components.

5 Conclusion

In pattern recognition, feature extraction techniques are widely employed to reduce the dimensionality of data and to enhance the discriminatory information. In this paper, we develop a supervised discriminant technique, called class mean embedding (CME), and using class mean of data points to enhance its discriminant power in their mapping into a low dimensional space. Based on the class information, our approach achieves good accuracy by realigning the submanifolds and rectifying the neighbor relations in the embedding space. Experimental results on ORL, Yale, AR, and FERET face databases show the effectiveness of the proposed method.

Acknowledgments This work is partially supported by the National Science Foundation of China under grant no. 60632050, 90820306, 60873151, 60973098, 61005008 and 61005005.

References

Belhumeur PN, Hespanha JP, Kriengman DJ (1997) Eigenfaces vs Fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720

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

Chen H, Chang H, Liu T (2005) Local discriminant embedding and itsvariants. CVPR

He X, Niyogi P (2003) Locality preserving projections. In: Proc. 16th Conference neural information processing systems

Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290:2323–2326

Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290:2319–2323

Turk M, Pentland AP (1991) Face recognition using eigenfaces. In: IEEE conference on computer vision and pattern recognition, pp 586–591

Yan S, Xu D, Zhang B, Zhang H, Yang Q, Lin S (2007) Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans on Pattern Anal and Mach Intell (T-PAMI) pp 40–51