Feature Extraction Based on Class Mean Embedding (CME)

Minghua Wan, Zhihui Lai, Zhong Jin
School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China
wmh36@mail.njust.edu.cn

Abstract—**Recently, local discriminant embedding (LDE) was proposed to manifold learning and pattern classification. In LDE framework, the neighbor and class of data points were 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 investigated 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. Experimental results on ORL and FERET face databases show the effectiveness of the proposed method.**

Keywords-local discriminant embedding (LDE); manifold learning; graph embedding; pattern classification;

I. INTRODUCTION

Face recognition has attracted wide attention of the researchers in biometric authentication.

Recently, Yan et al.[3] proposed a newly general framework which called graph embedding for dimensionality reduction, from which many algorithms, such as Principal Component Analysis (PCA) [1], Linear Discriminant Analysis (LDA) [2], Locality Preserving Projections (LPP) [4], Isometric Feature Mapping (ISOMAP) [5], Local Linear Embedding (LLE) [6], Laplacian Eigenmap [7], Marginal Fisher Analysis (MFA) [3], Local Disriminant Embedding (LDE) [8] can all be reformulated. Using the graph embedding framework as a platform, they developed a novel dimensionality reduction algorithm, and these algorithm all were utilized local neighborhood information to construct a global embedding of the manifold.

Focusing on manifold learning and pattern classification, 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 train 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 be the sampled data points to tend the manifold distribution in the high dimensional space.

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

II. OUTLINE OF LDE[6]

LDE is a supervised subspace learning algorithm. Let us consider a set of m sample $\{x_1, x_2, ..., x_m\}$ 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 $n > d$ .Therefore, class label l_i of $x_i (i = 1, \ldots, m)$ are used in LDE to determine a linear transformation matrix U. The new feature vectors $y_i \in R^d$ are defined by the following linear transformation:

$$y_i = U^T x_i \tag{1}$$

where $U \in R^{m \times d}$ is a transformation matrix. The column vectors of $\mathrm{U}=[u_1, u_2, \ldots, u_d]$ span a d-dimensional subspace.

Its objective is to maximize the function

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

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.1014

IEEE computer society

subject to

$$\sum_{i,j}\left\|U^T x_i - U^T x_j\right\|^2 w_{ij} = 1 \qquad (3)$$

where

$$w'_{ij} = \begin{cases} \exp\left(-\left\|x_i - x_j\right\|^2 / t\right), if\ l_i \neq l_j and\ i \in N_K^+(j) or\ j \in N_K^+(i) \\ 0, else \end{cases} \qquad (4)$$

$$w_{ij} = \begin{cases} \exp\left(-\left\|x_i - x_j\right\|^2 / t\right), if\ l_i = l_j and\ i \in N_K^+(j) or\ j \in N_K^+(i) \\ 0, else \end{cases} \qquad (5)$$

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。

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

$$X\left(D' - W'\right)X^T u = \lambda X\left(D - W\right)X^T u \qquad (6)$$

where the elements of the matrix W' are w'_{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.

III. 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.

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 $X = [x_1 x_2 \cdots x_m x_{m+1} \cdots x_{m+c}] \in R^n$.

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

1. *Construct neighborhood graphs.* Let G and G' denote two (undirected) graphs both over all data points. To construct G, $$G : if \left\{ \begin{array}{c} y_i = y_j \\ (i,j) \in m_i \\ i \in N_K^+(j)\ or\ j \in N_K^+(i) \end{array} \right\}. \qquad (7)$$

For G',

$$G' : if \left\{ \begin{array}{c} y_i \neq y_j \\ (i,j) \in m_i \\ i \in N_K^+(j)\ or\ j \in N_K^+(i) \end{array} \right\} \qquad (8)$$

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^G of G,

$$w_{ij}^G = \begin{cases} \exp\left(-\left\|x_i - x_j\right\|^2 / t\right), i \in N_K^+(j) or\ j \in N_K^+(i) \\ 0, else \end{cases} \qquad (9)$$

The other affinity matrix $W^{G'}$ of G' can be computed in the same way.

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

$$w_{ij}^G = \begin{cases} 1, i \in N_K^+(j) or\ j \in N_K^+(i) \\ 0, \quad else \end{cases} \qquad (10)$$

3. *Complete the embedding.* Find the generalized eigenvectors $u_1, u_2, \cdots, u_l$ that correspond to the l largest eigenvalues in:

$$X\left(D^{G'} - W^{G'}\right)X^T u = \lambda X\left(D^G - W^G\right)X^T u \quad (11)$$

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

$$d_{ii}^G = \sum_j w_{ij}^G, d_{ii}^{G'} = \sum_j w_{ij}^{G'} \qquad (12)$$

The embedding of x_i is accomplished by

$$z_i = U^T x_i \ (i = 1, 2, \ldots, m) \qquad (13)$$

where $U = [u_1, u_2, \cdots, u_l]$

IV. EXPERIMENTS AND RESULTS

To evaluate the proposed CME algorithm, we compare it with the PCA, LDA and LDE algorithm in face databases: ORL and FERET. When the projection matrix U was computed from the training part, all the images including training part and the test part were projected to feature space. Euclidean distance and nearest neighborhood classifier are used in all the experiments.

A. Experiment on the ORL face Database

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 were normalized to a resolution of 56×46. Fig.1

shows sample images of one person from ORL face database.

Figure 1. Images of one person in ORL

Now, we test the recognition performances of the four methods: PCA, LDA , LDE and CME. In the experiments, l images (l varies from2 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 recognition rate of each method and the corresponding dimension are shown 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 percent 90 image energy.

TABLE I. THE MAXIMAL AVERAGE RECOGNITION RATES (%) OF PCA, LDA, LDE AND CME ON THE ORL DATABASE AND THE CORRESPONDING DIMENSIONS (SHOWN IN PARENTHESES).

Training	PCA	LDA	LDE	CME
6	87.39	89.38	98.31	99.18
	(46)	(38)	(36)	(40)
5	86.71	87.23	96.52	98.80
	(46)	(38)	(46)	(30)
4	84.98	86.17	93.66	96.76
	(46)	(38)	(36)	(34)
3	82.27	85.09	89.56	94.39
	(46)	(38)	(36)	(26)
2	74.86	77.42	80.54	89.66
	(46)	(38)	(32)	(24)

Fig.2 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 training. From Fig.2 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, it further shows that the proposed method can extract more discriminative features than the other methods.

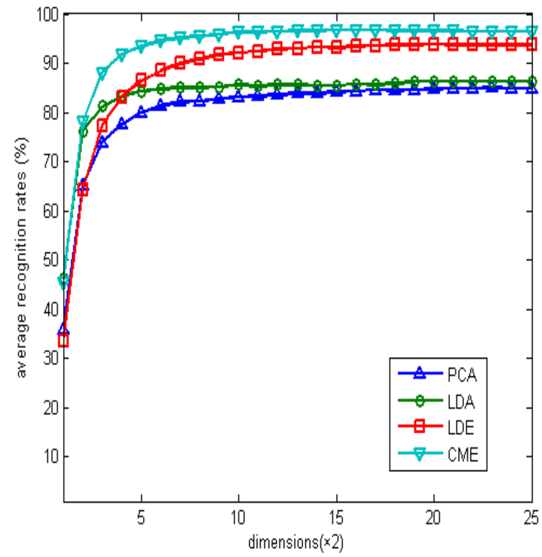

Figure 2. The average recognition rates (%) of PCA, LDA, LDE and CME versus the dimensions when the 4 images per class were randomly selected for training on the ORL face database.

B. Experiments on the FERET face Database

The proposed method is evaluated on a subset of FERET database, which includes 1400 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. 3.

Figure 3. Images of one person in FERET

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 $7-l$ images are used for testing. For each l, we independently rum 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. The maximal recognition rate of each method and the corresponding dimension are shown in Table 2. The recognition rate curves versus the variation of dimensions are shown in Fig.4. 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.

TABLE II. THE MAXIMAL AVERAGE RECOGNITION RATES (%) OF PCA, LDA, LDE AND CME ON THE FERET FACE DATABASE AND THE CORRESPONDING DIMENSIONS(SHOWN IN PARENTHESES).

Training	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)
2	35.49 (150)	45.31 (45)	47.58 (65)	58.18 (145)

Fig. 4 shows the recognition rate achieved by PCA, 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.

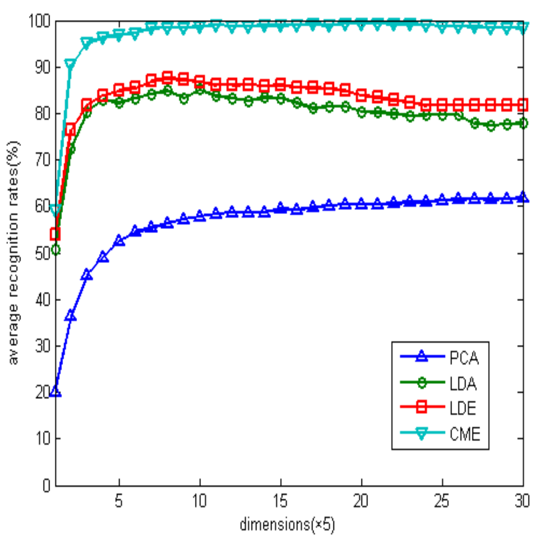

Figure 4. The average recognition rates (%) of PCA, LDA, LDE and CME versus the dimensions when the 6 images per class were randomly selected for training and on the FERET face database.

V. 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), 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 and FERET face databases show the effectiveness of the proposed method.

ACKNOWLEDGMENT

This work is partially supported by the National Science Foundation of China under grant no. 60632050, 90820306, 60873151, 60973098 and Hi-Tech Research and Development Program of China under grant no.2006AA01Z119.

REFERENCES

[1] M. Turk, A.P. Pentland, “Face recognition using eigenfaces”, in: IEEE Conference on Computer Vision and Pattern Recognition, 1991,pp. 586–591.

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

[3] 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.

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

[5] J.B. Tenenbaum, V. deSilva, and J.C. Langford, “A Global Geometric Framework for Nonlinear Dimensionality Reduction”, Science, vol. 290, pp. 2319-2323, 2000.

[6] S.T. Roweis and L.K. Saul, “Nonlinear Dimensionality Reduction by Locally Linear Embedding”, Science, vol. 290, pp. 2323-2326,2000.

[7] M. Belkin and P. Niyogi, “Laplacian Eigenmaps for Dimensionality Reduction and Data Representation”, Neural Computation, vol. 15, no. 6, pp. 1373-1396, 2003.

[8] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, “Face Recognition Using Laplacianfaces”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 328-340, Mar. 2005.