LETTER

Sparsity Preserving Embedding with Manifold Learning and Discriminant Analysis

Qian LIU[†,††], ***Student Member***, **Chao LAN**[†], **Xiao Yuan JING**[†,†††a)], **Shi Qiang GAO**[†], **David ZHANG**[††††], ***and*** **Jing Yu YANG**[†††††], ***Nonmembers***

SUMMARY In the past few years, discriminant analysis and manifold learning have been widely used in feature extraction. Recently, the sparse representation technique has advanced the development of pattern recognition. In this paper, we combine both discriminant analysis and manifold learning with sparse representation technique and propose a novel feature extraction approach named sparsity preserving embedding with manifold learning and discriminant analysis. It seeks an embedded space, where not only the sparse reconstructive relations among original samples are preserved, but also the manifold and discriminant information of both original sample set and the corresponding reconstructed sample set is maintained. Experimental results on the public AR and FERET face databases show that our approach outperforms relevant methods in recognition performance.

key words: *sparsity preserving embedding, manifold learning, discriminant analysis, feature extraction*

1. Introduction

In the literature of image recognition, feature extraction plays an important role and has been extensively studied. Typical feature extraction methods include principle component analysis (PCA) [1] and linear discriminant analysis (LDA) [2]. PCA seeks a projective space where the data variety is maximally preserved; LDA takes the class information into consideration and looks for a linear embedded space where the separability of inter-class samples is maximized and the separability of intra-class samples is minimized.

Manifold learning, with its successful applications in feature extraction, has attracted broad research interests. It tends to preserve the manifold structure of a given data set in a low-dimensional embedded subspace. Typical manifold learning methods include locally linear embedding (LLE) [3], Laplacian eigenmaps [4], locality preserving projection (LPP) [5] and neighborhood preserving embedding (NPE) [6]. Specially, LPP seeks a linear embedded space where local neighbor relations of samples can be preserved, while NPE looks for an embedded space where reconstructive relations of samples by their k nearest neighbors can be preserved. All above manifold learning methods are unsupervised, which do not consider the class label information while training. To take advantage of the class separability, some supervised manifold learning methods, which incorporate discriminant analysis, have been presented, such as local discriminant embedding (LDE) [7], locally discriminating projection (LDP) [8] and marginal fisher analysis (MFA) [9]. Particularly, LDE maintains the intrinsic neighbor relations of intra-class samples by setting affinity weights. MFA constructs an intra-class compactness graph and an inter-class separability graph based on the available class label information.

Recently, the sparse representation technique has advanced the development of pattern recognition. It shows that one sample can be sparsely recovered by the others. Based on this idea, sparsity preserving projections (SPP) [10] is developed for feature extraction, which aims at preserving the sparse reconstructive relations among samples in a low-dimensional subspace by minimizing the distance between sparsely reconstructed samples and their corresponding original samples.

Enlightened by above works, in this paper, we combine both manifold learning and discriminant analysis with sparse representation technique and propose a novel feature extraction approach named sparsity preserving embedding with manifold learning and discriminant analysis, which aims at both preserving the sparse reconstructive relations among original samples and maintaining the manifold and discriminant information of original samples and sparsely reconstructed samples. This aim is computationally achieved in two ways. On the one hand, we minimize the distance between sparsely reconstructed samples and their corresponding original samples as SPP does; on the other hand, providing that inter-class samples lie on different sub-manifolds while intra-class samples lie on the same sub-manifold, for each sparsely reconstructed sample, we minimizes its distance from the intra-class original samples and simultaneously maximizes its distance from the inter-class original samples. Experiments on the AR [11] and FERET [12] face databases validate the effectiveness of our approach.

Manuscript received March 29, 2011.
Manuscript revised August 3, 2011.

[†]The authors are with Nanjing University of Posts and Telecommunications, Nanjing, 210003, China.
[††]The author is also with Nanjing University of Information Science and Technology, Nanjing, 210044, China.
[†††]The author is also with State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210093, China and State Key Laboratory of Software Engineering, Wuhan University, Wuhan, 430079, China.
[††††]The author is with Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.
[†††††]The author is with Nanjing University of Science and Technology, Nanjing, 210094, China.
a) E-mail: jingxy_2000@yahoo.com
DOI: 10.1587/transinf.E95.D.271

2. Sparsity Preserving Embedding with Manifold Learning and Discriminant Analysis

SPP merely preserves the sparse reconstructive relations among original samples, but does not study the manifold structure of the data set. In this section, we investigate sparse reconstructive relations, manifold structure and discriminant information in dimension reduction simultaneously.

Suppose that $X = [x_1, x_2, \ldots, x_N]$ is the original sample set of size N, where x_i is the i^{th} sample. According to the sparse representation technique, x_i can be linearly recovered by the rest $N-1$ samples as

$$x_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + 0 \cdot x_i + \cdots + a_{iN}x_N, \tag{1}$$

where a_{ij} indicates the sparse reconstructive coefficient associated with the j^{th} sample for x_i. Note that $a_{ii} = 0$, indicating that x_i does not contribute to the sparse reconstruction of itself. Let $a_i = [a_{i1}, a_{i2}, \cdots, a_{iN}]^T$, and it is expected to have as few nonzero entries as possible in order to be sparse. The sparse representation technique calculates a_i by

$$\min \|a_i\|_0, \quad s.t.\ x_i = Xa_i\ . \tag{2}$$

However, solving Formula (2) has proved an NP-hard problem, and in our approach, we use an approximate L1-norm optimization problem [10] to calculate $\{a_i\}_{i=1}^N$, which is

$$\min \|a_i\|_1, s.t.\ \|x_i - Xa_i\|_2 < \varepsilon, E^T a_i = 1\ , \tag{3}$$

where ε is used to control the reconstructive error, $E \in R^N$ is a vector of all ones. Let $A = [a_1, a_2, \cdots, a_N]$. The overall sparsely reconstructed sample set X' can be calculated by

$$X' = XA\ . \tag{4}$$

Let y_i denote the class label of x_i. We define the inter-class weight matrix $H_b = \left[H_{ij}^b\right]_{N\times N}$ and intra-class weight matrix $H_w = \left[H_{ij}^w\right]_{N\times N}$ as follows:

$$H_{ij}^b = \begin{cases} \exp(-\left\|x_i - x_j\right\|^2/t), & \text{if } y_i \neq y_j \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

and

$$H_{ij}^w = \begin{cases} \exp(-\left\|x_i - x_j\right\|^2/t), & \text{if } y_i = y_j \\ & \text{and } i \neq j\ , \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where t is a parameter and it is usually set as the variance of the overall sample set. Then, we define the distance S_b between reconstructed samples and their inter-class original samples, and the distance S_w between reconstructed samples and their intra-class original samples as follows:

$$S_b = \sum_{i=1}^{N}\sum_{j=1}^{N} \left\|x_j - Xa_i\right\|^2 H_{ij}^b \tag{7}$$

and

$$S_w = \sum_{i=1}^{N}\sum_{j=1}^{N} \left\|x_j - Xa_i\right\|^2 H_{ij}^w\ . \tag{8}$$

The sparse reconstructive relations among original samples are evaluated by the total reconstructive errors defined below:

$$E = \sum_{i=1}^{N} \|x_i - Xa_i\|^2\ . \tag{9}$$

Equations (8) and (9) can be written in a unified form, i.e.,

$$S = \sum_{i=1}^{N}\sum_{j=1}^{N} \left\|x_j - Xa_i\right\|^2 H_{ij}\ , \tag{10}$$

where

$$H_{ij} = \begin{cases} \exp(-\left\|x_i - x_j\right\|^2/t), & \text{if } y_i = y_j \\ 0, & \text{otherwise}\ . \end{cases} \tag{11}$$

Based on Formulas (7) and (10), we build our objective function that maximizes S_b and simultaneously minimizes S in the feature space as follows:

$$\max_v \frac{v^T S_b v}{v^T S v}\ , \tag{12}$$

where v is the projective vector. S_b can be rewritten in a matrix form as

$$\begin{aligned} S_b &= \sum_{i=1}^{N}\sum_{j=1}^{N} \left\|x_j - Xa_i\right\|^2 H_{ij}^b \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} (x_j - Xa_i)(x_j - Xa_i)^T H_{ij}^b \\ &= \sum_{j=1}^{N} x_j(\sum_{i=1}^{N} H_{ij}^b)x_j^T - XH_b(XA)^T\ , \\ &\quad -(XA)H_bX^T + \sum_{i=1}^{N}(Xa_i)(\sum_{j=1}^{N} H_{ij}^b)(Xa_i)^T \\ &= X(D_b - H_bA^T - AH_b + AD_bA^T)X^T \end{aligned} \tag{13}$$

and S can be rewritten in a matrix form as

$$\begin{aligned} S &= \sum_{i=1}^{N}\sum_{j=1}^{N} \left\|x_j - Xa_i\right\|^2 H_{ij} \\ &= X(D - HA^T - AH + ADA^T)X^T\ , \end{aligned} \tag{14}$$

where $H = [H_{ij}]_{N\times N}$, $D = diag\{D_{ii}\}_{N\times N}$ and $D_b = diag\{D_{ii}^b\}_{N\times N}$ are two diagonal matrices, $D_{ii} = \sum_{j=1}^{N} H_{ij}$ and $D_{ii}^b = \sum_{j=1}^{N} H_{ij}^b$. By using the Lagrange multiplier method, the optimal solution is the eigenvector of matrix $S^{-1}S_b$ associated with the largest eigenvalue.

The algorithm of our approach is summarized in Fig. 1. For simplicity, we assume that sparse reconstructive coefficient matrix A has been obtained.

(1) Calculate weight matrices H and H_b by Formulas (11) and (5).
(2) Compute diagonal matrices D and D_b.
(3) Compute matrices S and S_b by Formulas (14) and (13).
(4) Calculate l eigenvectors $v_1, v_2, ..., v_l$ of matrix $S^{-1}S_b$ associated with the largest l eigenvalues, and let $V = [v_1, v_2, ..., v_l]$.
(5) Transform original samples into the embedded space by $Y = V^T X$, where Y represents the extracted features.

Fig. 1 Algorithm of our approach.

Fig. 2 Demo images of one subject in AR database.

Fig. 3 Demo images of one subject in FERET database.

3. Experiments

3.1 Introduction of Databases

The AR face database [11] contains 119 individuals, each 26 images with cropped size 60×60. All image samples of one subject are shown in Fig. 2. In order to effectively evaluate the impact of different variations to the recognition results, we in turn choose the following 2-10 representative images of every subject as training samples: (1), (14), (2), (5), (8), (11), (17), (19), (23) and (25), and use the remainders as testing samples.

The FERET database [12] includes 14126 facial images from 1199 individuals, which were captured under various illuminations, facial expressions and pose angles. We adopt the CSU Face Identification Evaluation System [13] to preprocess the full-frontal facial images and test the recognition performance of our approach. This system follows the FERET test procedure for semi-automatic face recognition algorithms [12] with slight modifications. All preprocessed images of one subject are shown in Fig. 3. Each image is cropped to 130×150. In our experiments, we choose the standard training subset for training, and use the gallery and dup1 probe sets for testing.

3.2 Recognition Performance Evaluation

We compare the recognition performance of our approach with several related methods, including discriminant analysis method LDA [2], unsupervised manifold learning methods LPP [5] and NPE [6], supervised manifold learning methods LDE [7] and MFA [9], and sparsity preserving method SPP [10]. For the related manifold learning methods, the number of nearest neighbors (k) is determined by the values that can yield the best recognition results. In all compared methods, we first perform PCA on the data to reduce dimension and avoid the singularity problem of the inverse matrix, and lastly use the nearest neighbor classifier to do classification. The number of PCA dimensions in all compared methods is $N - C$, where N is the number of all training samples, and C is the class number of the sample set. For all compared methods, the number of feature dimensions is determined by the values that can yield the best recognition results.

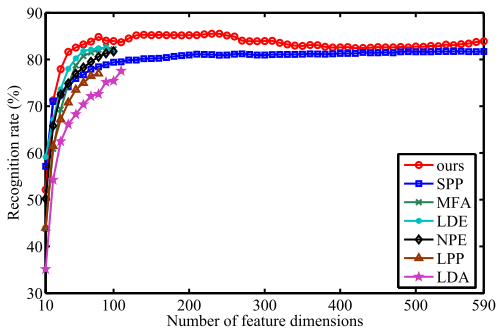

Fig. 4 Recognition rates versus different feature dimension numbers on AR database.

Table 1 Feature dimension numbers with corresponding recognition rates and average recognition rates of all compared methods on AR database.

Method	Feature dimension number (Recognition rate (%))	Average recognition rate (%)
LDA	110 (77.56)	78.62
LPP	80 (77.01)	78.63
NPE	100 (81.73)	80.47
LDE	80 (82.42)	80.18
MFA	90 (82.70)	80.43
SPP	560 (81.76)	80.49
Ours	**230 (85.50)**	**83.74**

Figure 4 shows the recognition rates of all compared methods versus different feature dimension numbers on the AR database, where the number of training samples per class is fixed to 6. Table 1 shows the final chosen feature dimension numbers. Figure 5 shows the recognition rates of all compared methods versus different training sample numbers per class on the AR face database. According to Fig. 5, the average recognition rates are shown in Table 1. Compared with LDA, LPP, NPE, LDE, MFA and SPP, our approach boosts the average recognition rates at least by 3.25% (= 83.74%-80.49%) on the AR database. Figure 6 shows the standard cumulative match curves of all compared methods on the FERET database. From Fig. 5, Table 1 and Fig. 6, we can see that our approach generally achieves the highest recognition results.

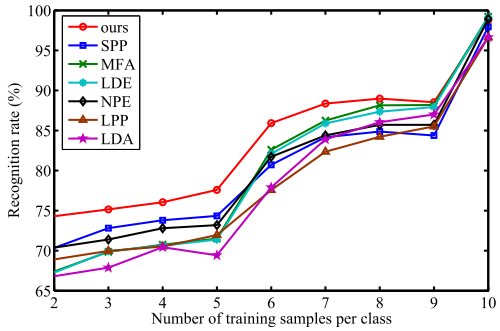

Fig. 5 Recognition rates versus different training sample numbers per class on AR database.

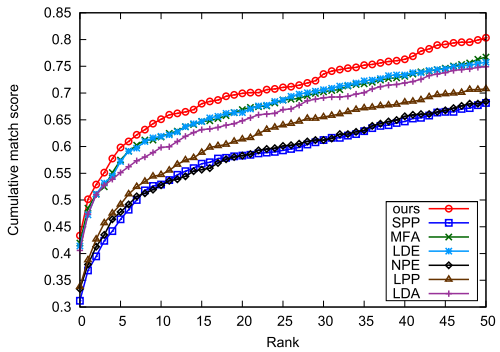

Fig. 6 Standard cumulative match curves on FERET database.

4. Conclusions

In this paper, we propose a novel feature extraction method named sparsity preserving embedding with manifold learning and discriminant analysis, which aims at both preserving the sparse reconstructive relations among original samples and maintaining the manifold and discriminant information of original samples and sparsely reconstructed samples. Computationally, our approach reconstructs each sample by a sparse linear representation of the other samples and obtains the sparsely reconstructed sample set. Then, it minimizes the distance between sparsely reconstructed samples and their corresponding original samples in the embedded space. In order to maintaining the manifold and discriminant information, for each sparsely reconstructed sample, our approach minimizes its distance from the intra-class original samples and simultaneously maximizes its distance from the inter-class original samples. Experimental results on the AR and FERET face databases show that, compared with LDA, LPP, NPE, MFA, LDE and SPP, our approach achieves the best recognition performance.

Acknowledgements

The work described in this paper was fully supported by National Natural Science Foundation of China under Project No.61073113 and No.60772059, New Century Excellent Talents of Education Ministry under Project No.NCET-09-0162, Doctoral Foundation of Education Ministry under Project No.20093223110001, Qing-Lan Engineering Academic Leader of Jiangsu Province and Foundation of Jiangsu Province Universities under Project No.09KJB510011.

References

[1] M. Turk and A. Pentland, "Eigenfaces for recognition," Cognitive Neuroscience, vol.3, no.1, pp.71–86, 1991.
[2] X.Y. Jing, D. Zhang, and Y.Y. Tang, "An improved LDA approach," IEEE Trans. Syst. Man Cybern. B, Cybern., vol.34, no.5, pp.1942–1951, 2004.
[3] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, vol.290, pp.2323–2326, 2000.
[4] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," Proc. Conf. Advances in Neural Information Processing Systems, pp.585–591, 2001.
[5] X. He and P. Niyogi, "Locality preserving projections," Proc. Conf. Advances in Neural Information Processing Systems, 2003.
[6] X. He, D. Cai, S. Yan, and H. Zhang, "Neighborhood preserving embedding," Int. Conf. Computer Vision, pp.1208–1213, 2005.
[7] H.T. Chen, H.W. Chang, and T.L. Liu, "Local discriminant embedding and its variants," Proc. Conf. Computer Vision and Pattern Recognition, vol.2, pp.846–853, 2005.
[8] H.T. Zhao, S.Y. Sun, Z.L. Jing, and J.Y. Yang, "Local structure based supervised feature extraction," Pattern Recognit., vol.39, pp.1546–1550, 2006.
[9] S.C. Yan, D. Xu, B.Y. Zhang, H.J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.1, pp.40–51, 2007.
[10] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," Pattern Recognit., vol.43, pp.331–341, 2010.
[11] A.M. Martinez and R. Benavente, "The AR face database," CVC Technical Report 24, 1998.
[12] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.10, pp.1090–1104, 2000.
[13] J.R. Beveridge, D. Bolme, B.A. Draper, and M. Teixeira, "The CSU face identification evaluation system: Its purpose, features and structure," Machine Vision and Applications, vol.16, no.2, pp.128–138, 2005.