Face Recognition Using A Multi-Manifold Discriminant Analysis Method

Wankou Yang, Changyin Sun
School of Automation
Southeast University
Nanjing, China
wankou_yang@yahoo.com.cn cysun@seu.edu.cn

Lei Zhang
Dept. of Computing
The Hong Kong Polytechnic University
Hong Kong, China
cslzhang@comp.polyu.edu.cn

***Abstract*—In this paper, we propose a Multi-Manifold Discriminant Analysis (MMDA) method for face feature extraction and face recognition, which is based on graph embedded learning and under the Fisher discriminant analysis framework. In MMDA, the within-class graph and between-class graph are designed to characterize the within-class compactness and the between-class separability, respectively, seeking for the discriminant matrix that simultaneously maximizing the between-class scatter and minimizing the within-class scatter. In addition, the within-class graph can also represent the sub-manifold information and the between-class graph can also represent the multi-manifold information. The proposed MMDA is examined by using the FERET face database, and the experimental results demonstrate that MMDA works well in feature extraction and lead to good recognition performance.**

Keywords- Multi-Manifold learning; LDA; face recongition

I. Introduction

In the past several decades, many feature extraction methods have been proposed. The most important ones are principle component analysis (PCA) and linear discriminant analysis (LDA) [1]. Since un-supervised learning may not be able to model the underlying structure and characteristics of different classes, discriminant features are often obtained by supervised learning. LDA [1] is the representative approach to learn discriminant subspace. Unfortunately, it cannot be applied directly to small sample size (SSS) problems [2] because the within-class scatter matrix is singular. Face recognition is a typical SSS problem and many works have been proposed to use LDA for face recognition [3-8]. Recent studies have shown that the face images possibly reside on a nonlinear sub-manifold [9-11]. Many manifold-based learning algorithms have been proposed for discovering the intrinsic low-dimensional embedding of the original data. Among the various methods, the most well-known ones are isometric feature mapping (ISOMAP) [9], local linear embedding (LLE) [10], and Laplacian Eigenmap [11]. Experiments have shown that these methods can find perceptually meaningful embedding for facial or digit images and other artificial and real-world data sets. However, how to evaluate the maps remains unclear. He et al. [12] proposed the Locality Preserving Projections (LPP), which is a linear subspace learning method derived from Laplacian Eigenmap. In contrast to most manifold learning algorithms, LPP possesses a remarkable advantage that it can generate an explicit map. This map is linear and can be easily computed, like PCA and LDA. The objective function of LPP is to minimize the local scatter of the projected data. Yang et al. [14] developed an Unsupervised Discriminant Projection (UDP) technique for dimensionality reduction. UDP characterizes the local scatter as well as the nonlocal scatter, seeking for a projection that simultaneously maximizes the nonlocal scatter and minimizing the local scatter.

The basic assumption of manifold learning is that the high-dimensional data can be considered as a set of geometrically related points lying on a smooth low-dimensional manifold. Each object space is usually a sub-manifold. Different object spaces are usually a multi-manifold. LPP and UDP are un-supervised method in nature, and multi-manifold information is partly considered in LPP and UDP. In this paper, we develop a novel method, namely multi-manifold discriminant analysis (MMDA), for feature extraction and pattern recognition. In MMDA, we construct two graphs to characterize the within-class compactness and the between-class separability, and give the criterion function for calculating the projection matrix. The within-class compactness and the between-class separability also can be characterized by the within-class Laplacian matrix and the between-class Laplacian matrix, respectively, which are associated with the within-class matrix, between-class matrix under the Fisher discriminant analysis framework. We seek for the projection matrix by simultaneously maximizing the between-class Laplacian scatter matrix and minimizing the within-class Laplacian matrix.

II. A Mulit-Manifold Discriminant Analysis method (MMDA)

A. Idea

In LDA, the between-class scatter matrix is more determined by the larger distances between class means, so the influence of the distances between class means is ignored in LDA. And each sample in each class has different contribution to within-class scatter matrix, which is also ignored in LDA. To overcome these drawbacks, we propose a multi-manifold discriminant analysis method for face feature extraction and recognition.

The idea of MMDA is to keep the class labeling after embedding or subspace learning. In other words, in the derived low dimensional MMDA subspace, we expect that the points are still close if they are from the same class, and the points from different classes are as far from each other as possible. To this end, we define two types of graphs in MMDA: within-class graph G_w and between-class graph G_b, with N nodes and c nodes respectively.

Denote the sample dataset as $X=[x_1,x_2,\cdots,x_N],\, x_i\in \mathrm{R}^m$ with the class label $l_i\in\{1,2,\cdots,c\}$. For the convenience of discussion, we assume that there is the same number of sample in each class. By using linear projection P, the low-dimensional representation of the sample can be obtained by $y=P^T x$. Thus in the low dimensional space the sample set is $Y=[y_1,y_2,\cdots,y_N],\, y_i\in \mathrm{R}^d,\, d\ll m$.

For the within-class graph G_w, we only consider the points with the same class label. An edge is constructed between nodes x_i and x_j from the same class. The similarity between node x_i and x_j is defined as follows:

$$C_{ij}=\begin{cases}\exp(-\|x_i-x_j\|^2/t) & \text{if class label } l_i=l_j\\ 0 & \text{else}\end{cases} \tag{1}$$

Obviously, for any x_i, x_j and parameter t, $0\le C_{ij}\le 1$ always holds. In addition, the weight function is a strictly monotonically decreasing function with respect to the distance between two points x_i and x_j.

The within-class graph-preserving criterion is

$$\arg\min_P P^T X L_w X^T P \tag{2}$$

where $L_w=D_w-C$ is the Laplacian matrix, D_w is a diagonal matrix with D_{wii} being the column (or row) sum of C, i.e. $D_{wii}=\sum_j C_{ij}$.

The affinity weight matrix C and diagonal matrix D_w can be written as:

$$C=\begin{bmatrix} C_1 & \cdots & 0\\ \vdots & \ddots & 0\\ 0 & \cdots & C_c\end{bmatrix},\quad D_w=\begin{bmatrix} D_{w1} & \cdots & 0\\ \vdots & \ddots & 0\\ 0 & \cdots & D_{wc}\end{bmatrix} \tag{3}$$

$C_1,\cdots,C_c$ represent the affinity weight in each class, $D_{w1},\cdots,D_{wc}$ represent a point's importance in its class. According to matrix D_{wk}, we can get the weighted center of class k:

$$\widetilde{m_k}=\frac{1}{\sum_i D_{kii}}(\sum_i D_{kii}x_{ki}) \tag{4}$$

Then we can get all the class weighted centers $M=[\widetilde{m_1},\widetilde{m_2},\cdots,\widetilde{m_c}]$, which can be more respective than the original mean of the each class.

Each class has its own manifold structure and different classes could reside on different manifolds. For recognition, it would be necessary to distinguish between classes from different manifolds. To achieve an optimal recognition, the recovered embeddings corresponding to different manifolds should be separated as much as possible in the final embedding space.

For the between-class graph G_b, based on the weighted centers of each class, we only consider the point pairs of M. An edge is constructed between nodes $\widetilde{m_i}$ and $\widetilde{m_j}$ with weight being set as

$$B_{ij}=\frac{1}{(\widetilde{m_i}-\widetilde{m_j})} \tag{5}$$

The between-class graph-penalizing criterion is

$$\arg\max_P P^T M L_b M^T P \tag{6}$$

where $L_b=D_b-B$ is the Laplacian matrix, D_b is a diagonal matrix with $D_{bii}=\sum_j B_{ij}$ being the column (or row) sum of B, B_{ij} is the weighted coefficient between nodes $\widetilde{m_i}$ and $\widetilde{m_j}$ and it adjusts the influence of the distance between nodes $\widetilde{m_i}$ and $\widetilde{m_j}$.

According to graph embedding, MMDA should satisfy the following two optimization criteria:

$$\begin{cases}\arg\min_P P^T X L_w X^T P\\ \arg\max_P P^T M L_b M^T P\end{cases} \tag{7}$$

We can further re-write the criteria as follows:

$$P=\arg\max_P \frac{P^T M L_b M^T P}{P^T X L_w X^T P} \tag{8}$$

From the framework of Fisher discrimiant analysis, in the MMDA subspace, the within-class Laplacian scatter can be formulated as:

$$J_w(P)=P^T\alpha_w P \tag{9}$$

where $\alpha_w\propto\sum_{ij}C_{ij}(x_i-x_j)(x_i-x_j)^T\propto XL_wX^T$ is the within-class Laplacian matrix, $L_w=D_w-C$ is the Laplacian matrix, D_w is a diagonal matrix with $D_{wii}=\sum_j C_{ij}$ being the column (or row) sum of C. For non-Gaussian or manifold-value data, we can process them by using local patches because non-Gaussian data can be approximately viewed as locally Gaussian and a curved manifold can be viewed as locally euclidean [35, 36].

The between-class Laplacian scatter can be defined

$$J_b(P)=P^T\alpha_b P \tag{10}$$

where $\alpha_b\propto\sum_{i=1}^{c}B_{ij}(\widetilde{m_i}-\widetilde{m_j})(\widetilde{m_i}-\widetilde{m_j})^T\propto ML_bM^T$ is the between-class Laplacian matrix, $L_b=D_b-B$ is the Laplacian matrix, D_b is a diagonal matrix with $D_{bii}=\sum_j B_{ij}$ being the column (or row) sum of B.

It is obvious that α_w and α_b are nonnegative symmetrical matrix. To maximize the between-class Laplacian scatter and minimize the within-class Laplacian scatter in the MMDA subspace, the objective function can be defined as:

$$J(P)=\arg\max_P\frac{J_b(P)}{J_w(P)}=\frac{P^T\alpha_b P}{P^T\alpha_w P} \tag{11}$$

It can be concluded that our proposed method is a graph embedded learning method and it is under the Fisher discrimiant analysis framework. Therefore, we call the proposed method multi-manifold discriminant analysis (MMDA).

B. The algorithm

The proposed MMDA based feature extraction algorithm can be summarized as follows:

Step1. Use PCA to transform the original image into a low dimensional subspace. Denote by W_{PCA} the transformation matrix of PCA.

Step2. In the PCA subspace, construct the similarity matrix C, within-class Laplacian scatter matrix α_w and the weighted class center M using Eqs. (1-4).

Step3. Construct the between-class Laplacian scatter matrix α_b using Eqs. (5) and (10), then calculate the eigenvectors $P=[p_1, p_2, \cdots, p_d]$ of $(\alpha_w)^{-1}\alpha_b$ corresponding to first d largest nonzero eigenvalues.

Step4. The final projection matrix is $W = P^T * W_{\text{PCA}}^T$.

III. EXPERIMENTS

The FERET face database is a result of the FERET program, which was sponsored by the US Department of Defense through the DARPA Program [14,15]. It has become a standard database for testing and evaluating state-of-the-art face recognition algorithms. The proposed method was tested on a subset of the FERET database. This subset includes 1,400 images of 200 individuals (each individual has seven images). This subset involves variations in facial expression, illumination, and pose. In our experiment, the facial portion of each original images was automatically cropped based on the location of the eyes and the cropped images was resized to 40×40 pixels. Some examples images of one person are shown in Fig.1.

Figure 1. Images of one person in FERET database.

In the experiment, we used the first l ($l = 2,3,4,5,6$) images per class for training and the remaining images for testing. We used PCA (eigenface), LDA (Fisherface), LPP, LPCA, LLD and the proposed MMDA for feature extraction and comparison. In PCA and the PCA stage of LDA, LPP, LLD and MMDA, we preserved nearly 95 percent image energy to select the number of principal components. In LPP, the number of the nearest neighbors was set as $l-1$, and the final dimension is set the same as that in PCA. Finally a nearest neighbor classifier with cosine distance is employed. The final recognition rates are given in Table1, from which we can find that the proposed method has the top recognition rate.

TABLE I. RECOGNITION RATE ON FERET DATABASE

	***l*=2**	***l*=3**	***l*=4**	***l*=5**	***l*=6**
PCA	0.4760	0.4150	0.5017	0.4000	0.3050
LDA	0.6600	0.5925	0.7067	0.7000	0.5900
LPP	0.4040	0.3575	0.4667	0.6125	0.6700
LPCA	0.5270	0.4925	0.6133	0.6875	0.7200
LLD	0.6350	0.5700	0.7033	0.7125	0.7150
Proposed	**0.6720**	**0.5975**	**0.7233**	**0.7250**	**0.7500**

From Table1, we can find that our proposed method (MMDA) outperforms the other methods and LLD has a better recognition rate than LDA. Each class lies on a manifold space and different classes may reside on different manifold spaces. MMDA simultaneously characterizes the sub-manifold information and multi-manifold information, which is important for classification. There are strong pose, expression and illumination variations in FERET face databases. Different training samples have different contributions to the mean of each class, which is explicitly considered in MMDA and useful to characterize the sub-manifold information and multi-manifold information. Compared to LPP, MMDA uses the class label information and does not need to choose the number of nearest neighbors. Compared to LDA, MMDA considers the contribution of each training sample to the class center, which is important to characterize the within-class scatter; MMDA considers the influence of the distances between class centers.

IV. CONCLUSIONS

Many manifold learning based feature extraction methods have been proposed. To model multi-manifolds for classification purpose, it is important to uncover the embeddings corresponding to different manifolds and, at the same time, to make different embedding separated as much as possible in the final embedding space. The proposed MMDA is based on graph embedded learning and is under the Fisher discriminant analysis framework. The experimental results on benchmark FERET face databases show that MMDA outperforms many existing representative subspace learning methods.

ACKNOWLEDGEMENTS

This project is supported by NSFC of China (Grant No: 90820009, 60632050, 60803049, 60875010), Program for New Century Excellent Talents in University Of China (NCET-08-0106) and the Hong Kong RGC General Research Grant (PolyU 5351/08E).

REFERENCES

[1] K. Fukunaga, Introduction to Statistical Pattern Recognition, second ed. Academic Press, Boston, MA, 1990.

[2] S.J. Raudys, A.K. Jain. "Small sample size effects in statistical pattern recognition: recommendations for practitioners," IEEE Trans. Pattern Ana1ysis and Machine Intel1igence, vol.13, no.3, pp.252-264, 1991.

[3] Belhumeur V, Hespanha J, Kriegman D. "Eigenfaces vs Fisherfaces: recognition using class specific linear projection," IEEE Trans. Pattern Ana1ysis and Machine Intel1igence, vol.19, no.7, pp.711-720, 1997.

[4] J. Yang, J. Yang, "Why can LDA be performed in PCA transformed space?" Pattern Recognition. vol.36, no.2, pp. 563-566,2003.

[5] Z. Hong, J. Yang, "Optimal discriminant plane for a small number of samples and design method of classifier on the plane," Pattern Recognition, vol.24, no.4, pp.317-324, 1991.

[6] J.H. Friedman, "Regularized discriminant analysis," Journal of the American Statistical Association, vol.84, pp.165-175, 1989.

[7] T. Hastie and R.Tibshirani, "Penalized discriminant analysis," The Annals of Statistics, vol.23, pp.73-102, 1995.

[8] T. Hastie, R. Tibshirani, A. Buja, "Flexible discriminant analysis by optimal scoring," Journal of the American Statistical Association. vol.89, pp.1255-1270, 1994.

[9] J.B. Tenenbaum, V.de. Silva, and J.C. Langford. "A global geometric framework for nonlinear dimensionality reduction," Science, vol.290, pp.2319-2323, 2000.

[10] S.T. Roweis and L.K. Saul. "Nonlinear dimension reduction by locally linear embedding," Science, vol.290, pp.2323-2326, 2000.

[11] M. Belkin and P. Niyogi. "Laplacian eigenmaps for dimensionality reduction and data representation," Neural Computation. Vol.15, no.6, pp.1373-1396, 2003.

[12] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang. "Face recognition using laplacianfaces," IEEE Trans. Pattern Ana1ysis and Machine Intel1igence, vol.27, no.3, pp.328-340, 2005.

[13] J. Yang, D. Zhang, J. Yang, B. Niu. "Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics," IEEE Trans. Pattern Ana1ysis and Machine Intel1igence, vol.29, no.4, pp.650-664, 2007.

[14] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, The FERET evaluation methodology for face recognition algorithms. IEEE Trans. Pattern Ana1ysis and Machine Intel1igence, 2000, 22(10):1090-1104.

[15] P.J. Phillips, The facial recognition technology (FERET) database, <http://www.itl.nist.gov/iad/humanid/feret/feret_master.html>, 2004