A Multi-Manifold Discriminant Analysis Method for Image Feature Extraction

Wankou Yang[a], Changyin Sun[a], Lei Zhang[b,1]

[a]School of Automation, Southeast University, Nanjing 210096, China
[b]Biometrics Research Centre, Dept. of Computing, the Hong Kong Polytechnic University, Hong Kong

Abstract: In this paper, we propose a Multi-Manifold Discriminant Analysis (MMDA) method for image feature extraction and pattern recognition based on graph embedded learning and under the Fisher discirminant analysis framework. In MMDA, the within-class graph and between-class graph are respectively designed to characterize the within-class compactness and the between-class separability, seeking for the discriminant matrix to simultaneously maximize the between-class scatter and minimize the within-class scatter. In addition, in MMDA, the within-class graph can represent the sub-manifold information, while the between-class graph can represent the multi-manifold information. The proposed MMDA is extensively examined by using the FERET, AR and ORL face databases, and the PolyU finger-knuckle-print databases. The experimental results demonstrate that MMDA is effective in feature extraction, leading to promising image recognition performance.

Index terms: Multi-manifold learning, LDA, feature extraction, image recognition

---

[1] Corresponding author. Email: cslzhang@comp.polyu.edu.hk.

1. Introduction

Dimensionality reduction is to construct a meaningful low-dimensional representation of high dimensional data. Since there are large volumes of high-dimensional data in numerous real-world applications, dimensionality reduction is a fundamental problem in many scientific fields such as visualization, computer vision and pattern recognition. With respect to pattern recognition, dimensionality reduction is an effective approach to reveal the distinctive features from the original data for pattern matching [1].

Linear discriminant analysis (LDA) is one representative approach to learning discriminant subspaces. Unfortunately, it cannot be applied directly to small sample size (SSS) problems [2] because the within-class scatter matrix is singular. Many image recognition problems such as face recognition are typical SSS problems, and hence numerous works [4-16] have been proposed to adapt LDA for face recognition, such as the Fisherface [3, 4]. In Fisherface, PCA is first used to reduce the dimension of the original features space before applying LDA. By throwing away the smallest projections in the PCA step, however, some useful discriminatory information may also be lost. More discussions about PCA and LDA can be found in [5].

Recent studies have shown that the face images possibly reside on a nonlinear sub-manifold [15-17]. Many manifold learning algorithms have been proposed for discovering the intrinsic low-dimensional embedding of the original data, among which the most well-known ones are isometric feature mapping (ISOMAP) [15], local linear embedding (LLE) [16], and Laplacian Eigenmap [17]. Experiments have shown that these methods can find perceptually meaningful embedding for artificial and real-world datasets such as facial or digit images. However, how to evaluate the maps remains unclear. He et al. [18] proposed the Locality Preserving Projections (LPP), which is a linear subspace learning method derived from Laplacian Eigenmap. The objective function of LPP is to minimize the local scatter of the projected data. In contrast to most manifold

learning algorithms, LPP possesses a remarkable advantage that it can generate an explicit map, which is linear and can be easily computed like PCA and LDA. Yang et al. [19] 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. Xu et al. [20] proposed some solutions to overcome SSS problem when LPP is applied in face recognition. Inspired by LPP, Zhao et al. proposed LPCA and LLD for feature extraction and recognition [21, 22]. Hu et al. proposed a 2DLPP method [23] for palmprint recognition, and in [24] Feng et al. proposed a framework for matrix-based feature extraction, based on which a method called 2-dimensional discriminant embedding analysis (2DDEA) was developed.

Neither LPP nor UDP uses the class label information and they are un-supervised methods in nature. In [25] and [26], Yan et al. proposed the Marginal Fisher Analysis (MFA) and Chen et al. proposed the Local Discriminant Embedding (LDE), respectively, for feature extraction and recognition. The two methods are very similar in formulation, and both of them combine locality and class label information to represent the intra-class compactness and interclass separability. MFA and LDE take advantage of the partial structural information of classes and neighborhoods of samples; however, it is difficult to decide the number of nearest neighbors of each sample and the number of shortest pairs from different classes in MFA and LDE. In addition, the region covariance matrix lies on the connected Riemannian manifold, instead of the subspace. B. Li et al. proposed a namely constrained maximum variance mapping (CMVM) method [27], which uses the local scatter and between-class scatter to characterize the sub-manifold and multi-manifold information, respectively. The projection matrix is calculated under the Fisher framework. However, CMVM ignores the class label information when characterizing the sub-manifold using local scatter, and ignores the influences of different class distances when charactering the multi-manifold using the between-class scatter. H. Wang et al. proposed a called locality-preserved maximum information

projection (LPMIP) method [28], which uses local scatter and nonlocal scatter to characterize the sub-manifold and multi-manifold information, respectively. LPMIP can be seen as a scatter difference version of UDP, and it is an unsupervised method.

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 form a multi-manifold. Actually, the multi-manifold information is partly considered in LPP and UDP. In this paper, we develop a novel method, namely multi-manifold discriminant analysis (MMDA), to extend the LDA formulation and make new contribution to manifold learning. Yang et al. [19] and Wang et al. [28] used the non-local scatter, which involves the between-class scatter, to represent the multi-manifold information. In [28], Li et al. used the between-class scatter matrix to represent the multi-manifold information. These methods achieved interesting results of multi-manifold learning. Inspired by these works, we propose to use the between-class graph to represent the multi-manifold information and use the within-class graph to represent the sub-manifold information. So in MMDA, we construct two graphs to characterize the within-class compactness and the between-class separability, and define the criterion function to calculate the projection matrix. The within-class compactness and the between-class separability can also be characterized by the within-class Laplacian matrix and the between-class Laplacian matrix, 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.

It is worthwhile to highlight several aspects of the proposed method here. First, MMDA does not overemphasize the effect of large class distance, which degenerates the effect of LDA; second, MMDA explicitly considers the sub-manifold information and multi-manifold information, which is directly related to classification and recognition; third, MMDA is a linear supervised method and

4

the available class labels are used in it; fourth, MMDA is based on graph embedded learning and belongs to the Fisher discriminant analysis framework.

The organization of this paper is as follows. In Section 2, we review briefly the LDA and LPP. In Section 3, we propose the idea and describe the new method in detail. In Section 4, experiments on face and finger-knuckle-print image databases are presented to demonstrate the effectiveness of the new method. Conclusions are made in Section 5.

2. Related Work

2.1 LDA

LDA seeks for a projection matrix such that the Fisher criterion (i.e. the ratio of the between-class scatter to the within-class scatter) is maximized after the projection. Suppose there are C known pattern classes, $\omega_1, \omega_2, \cdots, \omega_C$. The between-class scatter matrix S_b, within-class scatter matrix S_w and the total scatter matrix S_t can be denoted as:

$$S_b = \frac{1}{N} \sum_{i=1}^{C} N_i (m_i - m_o)(m_i - m_o)^T \tag{1}$$

$$S_w = \frac{1}{N} \sum_{i=1}^{C} \sum_{x_k \in \omega_i} (x_k - m_i)(x_k - m_i)^T \tag{2}$$

$$S_t = \frac{1}{N} \sum_{i=1}^{N} (x_k - m_o)(x_k - m_o)^T \tag{3}$$

where N is the total number of training samples, and N_i is the number of training samples in class i. The jth training sample in class i is denoted by x_i^j, the mean vector of training samples in class i is denoted by m_i and the mean vector of all training samples is m_o.

If S_w is nonsingular, the optimal projection W_{opt} is chosen as the matrix with orthonormal columns which maximize the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples, i.e.

$$W_{opt} = \arg\max_{W} \frac{W^T S_b W}{W^T S_w W} \tag{4}$$

where $\{w_i \mid i = 1, 2, \cdots m\}$ is the set of generalized eigenvectors of $(S_w)^{-1}S_b$ corresponding to the m largest generalized eigenvalues.

2.2 LPP

Locality preserving projections (LPP) [18] is a well-known method for image feature extraction and dimension reduction. The objective of LPP is to preserve the local structure of the image space by explicitly considering the manifold structure. The following objective function is used:

$$\min \sum_{ij} (y_i - y_j) S_{ij} = \min W^T S_L W \tag{5}$$

where $y_i = W^T x_i$ is the low-dimensional representation of original data vector x_i and the matrix S is a similarity matrix, which can include the Gaussian weights or uniform weights of Euclidean distances using k-neighborhood or ε-neighborhood, and $S_L = XLX^T$.

By imposing a constraint $Y^T DY = W^T XDX^T W = W^T S_D W = 1$, where $X = [x_1, x_2, \cdots, x_N]$, and D is a diagonal matrix whose entries are column (or row) sums of S, $D_{ii} = \sum_j S_{ij}$, $S_D = XDX^T$, then the minimization problem becomes

$$\arg\min_{W^T S_D W = 1} W^T S_L W \tag{6}$$

The optimal projection matrix W is given by the minimum eigenvalue solution to the generalized eigenvalue problem:

$$S_L W = \lambda S_D W \tag{7}$$

S_D is often singular in the case of SSS. He et al. [18] proposed to implement LPP in the PCA subspace, which is similar to the Fisherface framework.

3. The Proposed MMDA

3.1 Motivation

Before going into the details of our proposed method, let's make more analysis about LDA and LPP. In LDA, the between-class scatter matrix S_b and within-class scatter matrix S_w can be rewritten as:

$$S_b \propto \sum_{ij}(m_i - m_j)(m_i - m_j)^T \tag{8}$$

$$S_w \propto \sum_{k=1}^{c} \sum_{x_{ki}, x_{kj} \in \omega_k} (x_{ki} - x_{kj})(x_{ki} - x_{kj})^T \tag{9}$$

If the distance between m_i and m_j is large, S_b will be more determinate by class i and class j, so LDA will put more emphasis on class i and class j and the neighboring classes may merge together in the LDA subspace. If the distance between m_i and m_j is small (even $m_i = m_j$), S_b will have no discriminative information or it is difficult to discriminate class i from class j according to S_b. In each class, each sample has different contribution to S_w, or each class has its individual manifold structure, which is also ignored in LDA.

As for LPP, the matrix D provides a natural measure on the data points. The bigger the value D_{ii} (corresponding to x_i) is, the more important x_i is. LPP characterizes its global quantity according to the distance between each point and the origin according to

$$W^T X D X^T W = \sum_i D_{ii}(W^T x_i) \tag{10}$$

A reasonable method is to characterize the global quantity according to the distance between each point and the mean weight of all the samples, as pointed out in [26]. There are

$$\tilde{m} = \frac{1}{\sum_i D_{ii}}(\sum_i D_{ii} x_i) \tag{11}$$

$$Y^T D Y = W^T X D X^T = \sum_i D_{ii}(W^T(x_i - \tilde{m}))^2 \tag{12}$$

We see that LPP does not explicitly consider the relations between different clusters, and LPP is an unsupervised linear method and the class label information is ignored. Although CMVM and LPMIP are multi-manifold-based methods, the influence between class means is ignored in CMVM, and the class label information is ignored in LPMIM.

To overcome the shortcomings of LDA and LPP as discussed above, next we propose a novel method for feature extraction and recognition, namely multi-manifold discriminant analysis (MMDA), which is a graph driven manifold learning method under the Fisher framework.

3.2 Fundamentals

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 samples 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 $Y = [y_1, y_2, \cdots, y_N], y_i \in \mathrm{R}^d, d \ll m$.

The motivation of MMDA is to keep the class labeling after graph embedding or subspace learning. In other words, in the 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.

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:

$$C_{ij} = \begin{cases} \exp(-\|x_i - x_j\|^2 / t) & \text{if class label } l_i = l_j \\ 0 & \text{else} \end{cases} \tag{13}$$

The above similarity function is widely used in manifold learning. Obviously, for any x_i, x_j and parameter t, $0 \leq C_{ij} \leq 1$ always holds. In addition, the weight function is a strictly monotonically decreasing function with respect to the distance between points x_i and x_j.

The within-class graph-preserving criterion is defined as

$$\arg \min_{P} P^T X L_w X^T P \tag{14}$$

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:

$$C = \begin{bmatrix} C_1 & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & \cdots & C_c \end{bmatrix}, \quad D_w = \begin{bmatrix} D_{w1} & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & \cdots & D_{wc} \end{bmatrix} \tag{15}$$

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

$$\tilde{m}_k = \frac{1}{\sum_i D_{kii}} (\sum_i D_{kii} x_{ki}) \tag{16}$$

Then we can get all the class weighted centers $M = [\tilde{m}_1, \tilde{m}_2, \cdots, \tilde{m}_c]$, 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 $\tilde{m}_i$ and $\tilde{m}_j$ with weight

$$B_{ij} = \exp(-\left\| \tilde{m}_i - \tilde{m}_j \right\|^2 / t) \tag{17}$$

The purpose of the weight in Eq. (17) is to directly enhance the contributions of classes that have smaller distances.

The between-class graph-penalizing criterion is defined as

$$\arg \max_{P} P^T M L_b M^T P \tag{18}$$

where $L_b = D_b - B$ is the Laplacian matrix and 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 $\tilde{m}_i$ and $\tilde{m}_j$ and it adjusts the influence of the distance between nodes $\tilde{m}_i$ and $\tilde{m}_j$.

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

$$\begin{cases} \arg \min_{P} P^T X L_w X^T P \\ \arg \max_{P} P^T M L_b M^T P \end{cases} \tag{19}$$

We can further re-write the criteria as follows:

$$P = \arg \max_{P} \frac{P^T M L_b M^T P}{P^T X L_w X^T P} \tag{20}$$

According to the framework of Fisher discrimiant analysis, with MMDA the within-class Laplacian scatter can be formulated as:

$$J_w(P) = P^T \alpha_w P \tag{21}$$

where $\alpha_w \propto \sum_{ij} C_{ij} (x_i - x_j)(x_i - x_j)^T \propto X L_w X^T$ 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 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 [29, 30].

The between-class Laplacian scatter can be defined

$$J_b(P) = P^T \alpha_b P \qquad (22)$$

where $\alpha_b \propto \sum_{i=1}^{c} B_{ij}(\tilde{m}_i - \tilde{m}_j)(\tilde{m}_i - \tilde{m}_j)^T \propto ML_bM^T$ 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 matrices. 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:

$$J(P) = \arg\max_P \frac{J_b(P)}{J_w(P)} = \frac{P^T \alpha_b P}{P^T \alpha_w P} \qquad (23)$$

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).

The projection matrix P consists of the generalized eigenvectors corresponding to the largest

eigenvalues of $\alpha_b P = \lambda \alpha_w P$. However, often the within-class Laplacian matrix α_w is singular

because the training sample size is smaller than the dimension of the image vector space. To

address this issue, we first use PCA to reduce the dimension of the original image vectors so that

α_w is nonsingular in the PCA subspace.

3.3 The algorithm

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

Step1. Use PCA to transform the original image into a lower 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

11

matrix α_w and the weighted class center M using Eqs. (13-16).

Step3. Construct the between-class Laplacian scatter matrix α_b using Eqs. (17) and (22), and then calculate the eigenvectors $P = [p_1, p_2, \cdots, p_d]$ of $(\alpha_w)^{-1}\alpha_b$ corresponding to the first d largest nonzero eigenvalues.

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

3.4 Connections with LDA, LPP, CMVM, LPMIP, 2DDEA

It is worthwhile to pointed out that the within-class Laplacian matrix $\alpha_w \propto XL_w X^T$ is the same as the within-class scatter matrix of LDA, if the parameter t in Eq. (13) is set as $t = +\infty$. When $t = +\infty$ the similarity between node x_i and x_j becomes:

$$C_{ij} = \begin{cases} 1 & \text{if class label } l_i = l_j \\ 0 & \text{else} \end{cases} \tag{24}$$

Then the within-class Laplacian matrix is equivalent to the within-class scatter matrix S_w in LDA:

$$\alpha_w \propto \sum_{ij} C_{ij}(x_i - x_j)(x_i - x_j)^T \propto XL_w X^T \propto \sum_k \sum_{ij} (x_{ki} - x_{kj})(x_{ki} - x_{kj})^T \propto S_w \tag{25}$$

The class centers are the means of each class and $M = [m_1, m_2, \cdots, m_c]$. The between-class Laplacian matrix is equivalent to the between-class scatter matrix S_b in LDA when the parameter t in Eq. (17) is set as $t = +\infty$:

$$\alpha_b \propto \sum_{i=1}^{c} B_{ij}(\tilde{m}_i - \tilde{m}_j)(\tilde{m}_i - \tilde{m}_j)^T \propto ML_b M^T \propto \sum_{i=1}^{c} (m_i - m_j)(m_i - m_j)^T \propto S_b \tag{26}$$

From the above analysis we can see that LDA is a special case of MMDA. Sub-manifold information in each class is considered in MMDA by the within-class Laplacian matrix, which is ignored in LDA. In MMDA, the influence of the distances between class means is considered in MMDA by the between-class Laplacian matrix, which is ignored in LDA.

LPP and LPMIP are unsupervised subspace learning methods. The objective of LPP is to minimize the local scatter of the projected data. In some case, this criterion cannot guarantee to yield a good projection matrix for classification purpose. The objective of LPMIP is to minimize the local scatter of the projected data and maximize nonlocal scatter of the projected data. The class label information is ignored in LPP and LPMIP, which is important for recognition and classification. In MMDA, the within-class Laplacian matrix represents the local scatter of the projected data or the sub-manifold information; the between-class Laplacian matrix represents the between-class scatter or the multi-manifold information. By simultaneously maximizing the between-class scatter and minimizing the within-class scatter, MMDA seeks for the optimal projection matrix for classification and recognition.

The objective functions of LPP, CMVM and LPMIP are respectively equivalent to

$$\arg\min J(W) = \frac{W^T S_L W}{W^T S_D W} \Leftrightarrow \arg\max J(W) = \frac{W^T S_D W}{W^T S_L W} \qquad (27)$$

$$\arg\max J_{CMVM}(W) = \frac{W^T S_b W}{W^T S_L W} \qquad (28)$$

$$\arg\max J_{LPMPIP}(W) = \alpha W^T S_N W - (1-\alpha) W^T S_L W \qquad (29)$$

In LPP, CMVM and LPMIP, when the dataset has an ideal clustering, i.e. each local neighborhood contains exactly the same number of training samples belonging to the same class, the local scatters of LPP, CMVM and LPMIP are equivalent to the within-class Laplacian scatter of MMDA. In LPP, the purpose of constraint $W^T S_D W = 1$ is to remove the scaling in the embedding and the purpose of the matrix D is to provide a natural measure of the vertices of the adjacency graph, and maximizing $W^T S_D W$ does not make sense with respect to discrimination. CMVM uses the between-class scatter matrix S_b to characterize the multi-manifold information which is used in

LDA, and has the same shortcomings as LDA. In LPMIP, the nonlocal scatter is used to characterize the multi-manifold information and the class label information is ignored. In contrast, the between-class Laplacian matrix in MMDA has a more transparent link to classification and recognition. Its physical meaning is very clear: if samples belong to different classes, they become as far from each other as possible.

2DDEA is a general theoretical framework for matrix-based feature extraction algorithms from the point of view of graph embedding. MMDA is also based on graph embedding learning and under Fisher discriminant analysis framework. Sub-manifold information of each class is all considered in our proposed method and the methods in 2DDEA. Multi-manifold information of different classes is not explicitly considered in the methods in 2DDEA, but explicitly considered in our method. Since our method is based on vectors, the image structural information is ignored.

4 Experiments

Three face image databases, namely, the Yale database, the ORL database, the FERET database, and the PolyU finger-knuckle-print (FKP) database are used to evaluate the proposed MMDA approach in comparison with the following algorithms: PCA (Eigenface) [4], LDA (Fisherface) [4], LPP ($t = +\infty$) [18], LPCA [21], LLD [22], CMVM [27] and LPMIP [28], ICA [31] The parameter in Eq. (13) is important for the recognition rate and it depends on the dataset. Currently there is not a good rule to choose the parameter t in Eq. (13) and Eq. (17). In the experiments, we set the parameter $t=r\times N$ (where N is the image size) and then choose r (r is usually set as *0.03*). The experiments were implemented on a Laptop with P8600 CPU and 2G RAM under the MATLAB (Version 7.01) programming environment.

4.1 Experiments on the FERET Database

The FERET face image database is a result of the FERET program, which was sponsored by the US Department of Defense through the DARPA Program [32, 33]. It has become a standard database for testing and evaluating state-of-the-art face recognition algorithms.

The proposed algorithm 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 has variations of facial expression, illumination, and pose. In our experiment, the facial portion of each original image was automatically cropped based on the location of eyes and the cropped images were resized to 40 by 40 pixels. Some example images of one person are shown in Fig. 1.

Figure 1. Images of one person in FERET.

In the experiment, we used the first l (l=3,4,5,6) images per class for training and the remaining images for testing. In PCA and the PCA stage of LDA, LPP, LLD, CMVM, LPMIP and MMDA, we preserved nearly 95% 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 highest recognition rate.

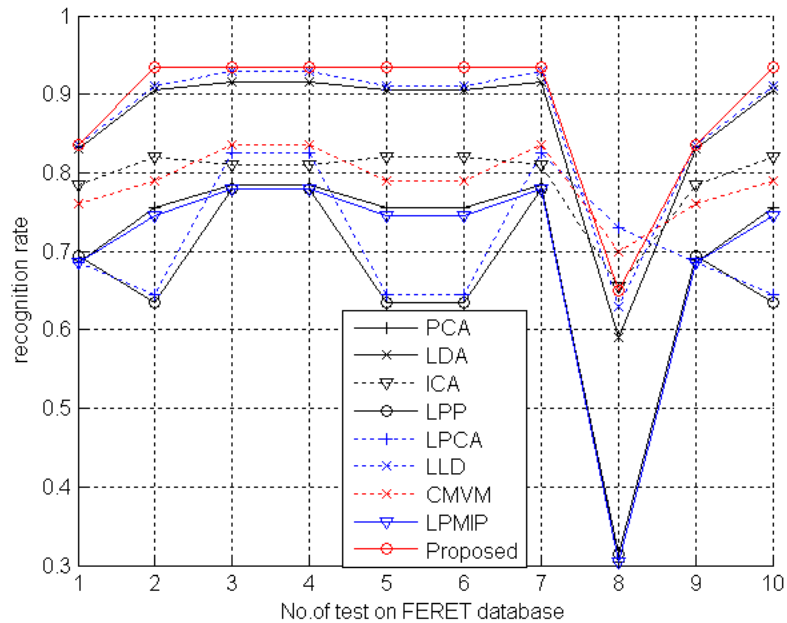**Table 1 Recognition Rates on the FERET database by different methods**

	PCA	LDA	ICA	LPP	LPCA	LLD	CMVM	LPMIP	**Proposed**
$l = 3$	0.4150	0.5925	0.5275	0.3575	0.4925	0.5700	0.4088	0.4150	**0.6000**
$l = 4$	0.5017	0.7067	0.6733	0.4667	0.6133	0.7033	0.4950	0.5017	**0.7217**
$l = 5$	0.4000	0.7000	0.6800	0.6125	0.6875	0.7125	0.4075	0.4000	**0.7250**
$l = 6$	0.3050	0.5900	0.5700	0.6700	0.7200	0.7150	0.3300	0.3050	**0.7750**

Table 2 Average Recognition Results on the FERET database by different methods

	PCA	LDA	ICA	LPP	LPCA	LLD	CMVM	LPMIP	**Proposed**
Mean	0.7050	0.8615	0.7935	0.6585	0.7155	0.8600	0.7885	0.6995	**0.8865**
Std	0.1453	0.1010	0.0505	0.1363	0.0802	0.0987	0.0421	0.1429	**0.0929**

Table 3 Average training time and testing time on the FERET database(s)

	PCA	LDA	ICA	LPP	LPCA	LLD	CMVM	LPMIP	**Proposed**
Training	16.1688	16.1861	14.7705	80.8437	13.9079	80.1875	79.0062	129.7406	**83.1375**
Testing	4.1874	4.0407	4.2405	4.1812	4.0827	4.1343	4.0672	4.2282	**4.2268**

Figure 2. The recognition rates of PCA, LDA, LPP, LPCA, LLD, CMVM, LPMIP and the proposed MMDA versus different training sets on the FERET database.

In the second experiment, six images per person were randomly chosen for training, and the other five for testing. Thus, the training set has 1200 images and the test set has 200 images. We run the system 10 times. All possible dimensions of the low-dimensional representation were evaluated, and

the best average recognition rates and the corresponding standard deviations (std) across 10 runs are reported in Table 2. The curve of recognition rate versus the 10 different training sets is shown in Fig. 2. We can find that MMDA has good performance. The average training time and testing time are shown in Table 3.

4.2 Experiments on the ORL database

The ORL (http://www.cam-orl.co.uk) database contains 40 persons, each having 10 different images. The images of the same person are taken at different times, under slightly varying lighting conditions and with various facial expressions. Some people are captured with or without glasses. The heads in images are slightly titled or rotated. The images in the database are manually cropped and rescaled to 112×92. Fig. 3 shows ten images of one person in ORL.

Figure 3. Ten images of one person in ORL

Table 4 Recognition Rate on the ORL database

	PCA	LDA	ICA	LPP	LPCA	LLD	CMVM	LPMIP	**Proposed**
$l = 3$	0.8429	0.8821	0.7714	0.8571	0.8857	0.9000	0.8964	0.8571	**0.9071**
$l = 4$	0.8750	0.9125	0.8375	0.8708	0.9167	0.9208	0.8958	0.8833	**0.9500**
$l = 5$	0.9050	0.9300	0.8700	0.8750	0.9200	0.9300	0.9000	0.9150	**0.9500**
$l = 6$	0.9563	0.9563	0.9250	0.9063	0.9375	0.9563	0.9625	0.9500	**0.9812**

In the experiment, we used the first l (l=3,4,5,6) images per class for training and the remaining images for testing. In PCA and the PCA stage of LDA, LPP, LLD, CMVM, LPMIP and MMDA, we kept nearly 95% 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 the same as that in PCA. Finally a

nearest neighbor classifier with cosine distance is employed for classification. The final recognition

rates are given in Table 4. Again, the proposed method achieves the highest recognition rate.

Table 5 Average Recognition Results on the ORL database

	$l = 3$	$l = 4$	$l = 5$	$l = 6$
PCA	0.8918(0.0189)	0.9375(0.0224)	0.9615(0.0118)	0.9750(0.0093)
LDA	0.9061(0.0211)	0.9512(0.0195)	0.9710(0.0120)	0.9769(0.0078)
ICA	0.8789(0.0245)	0.9308(0.0250)	0.9510(0.0173)	0.9681(0.0116)
LPP	0.8789(0.0235)	0.9300(0.0195)	0.9515(0.0183)	0.9681(0.0133)
LPCA	0.9032(0.0251)	0.9521(0.0177)	0.9660(0.0094)	0.9774(0.0069)
LLD	0.9129(0.0237)	0.9504(0.0182)	0.9770(0.0092)	0.9819(0.0055)
CMVM	0.8525(0.0273))	0.9063(0.0245)	0.9325(0.0151)	0.9519(0.0164)
LPMIP	0.8864(0.0301)	0.9221(0.0265)	0.9440(0.0115)	0.9631(0.0130)
Proposed	**0.9164(0.0271)**	**0.9550(0.0164)**	**0.9760(0.0081)**	**0.9831 (0.0084)**

Table 6 Average training time and testing time on the ORL database(s)

	$l = 3$	$l = 4$	$l = 5$	$l = 6$
	Training-Testing	Training-Testing	Training-Testing	Training-Testing
PCA	0.0828-0.5890	0.1497-0.7018	0.1969-0.6780	0.3029-0.6908
LDA	0.1157-0.7109	0.1795-0.8219	0.2889-0.8625	0.4405-0.8533
ICA	2.2094-0.6110	4.1390-0.6921	6.8407-0.7156	11.8514-0.6876
LPP	1.2627-0.6093	2.2172-0.6828	3.3921-0.7299	4.7842-0.6751
LPCA	4.1390-0.5876	4.9718-0.6735	6.5532-0.7452	7.9501-0.6859
LLD	1.2704-0.5718	2.2188-0.6501	3.3404-0.6564	4.8065-0.6328
CMVM	1.2001-0.5764	2.1031-0.6423	3.2906-0.6656	4.7313-0.6358
LPMIP	2.0358-0.6110	3.5842-0.6612	5.3484-0.6642	7.5921-0.6346
Proposed	**1.3078-0.5846**	**2.1795-0.6391**	**3.5768-0.7063**	**5.3938-0.6842**

To further evaluate the performance of PCA, LDA, LPP, LPCA, LLD, CMVM, LPMIP and the

proposed MMDA, in each test, l (l=3,4,5,6) images per class were randomly chosen for training,

while the remaining eight images were used for testing. We run the system 10 times. All possible

dimensions of the final low-dimensional representation were evaluated, and the best average results

are reported in Table 5. We can find that the proposed method works consistently well. The average

training time and testing time are shown in Table 6.

4.3 Experiments on the AR database

The AR face [34, 35] 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 section contains 13 color images. Twenty face images (each session contains 10) of these 120 individuals were selected and used in our experiments. These images have variations of neutral expression, smiling, angry, screaming, left light on, right light on, all sides light on, wearing sun glasses, wearing sun glasses with left light on, and wearing sun glasses with right light on. The face portion of each image was manually cropped. The example images of one person are shown in Fig. 4.
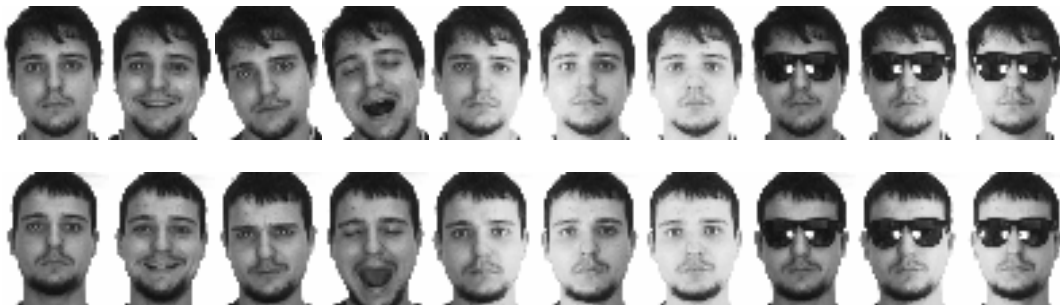
Figure 4. Samples of the cropped images of one person in AR database.

Table 7 Recognition rates on the AR database by different methods.

	PCA	LDA	ICA	LPP	LPCA	LLD	CMVM	LPMIP	**Proposed**
Recognition Rate	0.6367	0.6750	0.6350	0.5767	0.6617	0.6892	0.6869	0.6512	**0.7107**
Dim	356	119	356	356	356	119	119	119	**119**

In the experiment, the first 10 images per class were used for training and the remaining images for testing. In the PCA phase, we selected the first 356 principal components. In the second stage of

LDA, LLD, CMVM, LPMIP and MMDA, the dimension was reduced to c-1=119, where c is the class number. The final recognition rate of each method and the corresponding dimension are given in Table 7.

4.4 Experiments on the PolyU FKP database

In the last experiments, we used the PolyU finger-knuckle-print (FKP) database [36-38] to evaluate the performance of the proposed method. The PolyU FKP database was collected from 165 volunteers, including 125 males and 40 females. Among them, 143 subjects are 20-30 years old and the others are 31-50 years old. The images were collected in two separated sessions. In each session, the subject was asked to provide 6 images for each of the left index finger, the left middle finger, the right index finger and the right middle finger. In total, the database contains 7,920 images from 660 different fingers. The size of the ROI (region of interest) is 110×200. Fig. 5 shows twelve sample images of one left index finger. In the experiments, we choose a subset of the PolyU FKP database, the images of the left index finger, to evaluate the performance of the proposed method. In the experiments, we resize the image size to 55×110. Please note that in this paper the FKP database is used to illustrate the effectiveness of the proposed MMDA method LDA-based methods and manifold learning methods. Our goal is not to compare the FKP recognition accuracy of MMDA with our previous methods proposed in [36-37] because they use totally different categories of techniques.
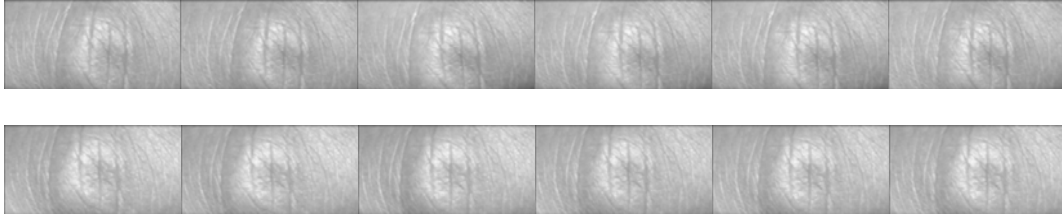
Figure 5. Samples of the cropped images in the PolyU FKP database

Table 8 Recognition rates on the PolyU FKP database by different methods.

	PCA	LDA	ICA	LPP	LPCA	LLD	CMVM	LPMIP	**Proposed**
Recognition rate	0.3374	0.6303	0.5202	0.4556	0.6444	0.7091	0.6859	0.6687	**0.7303**
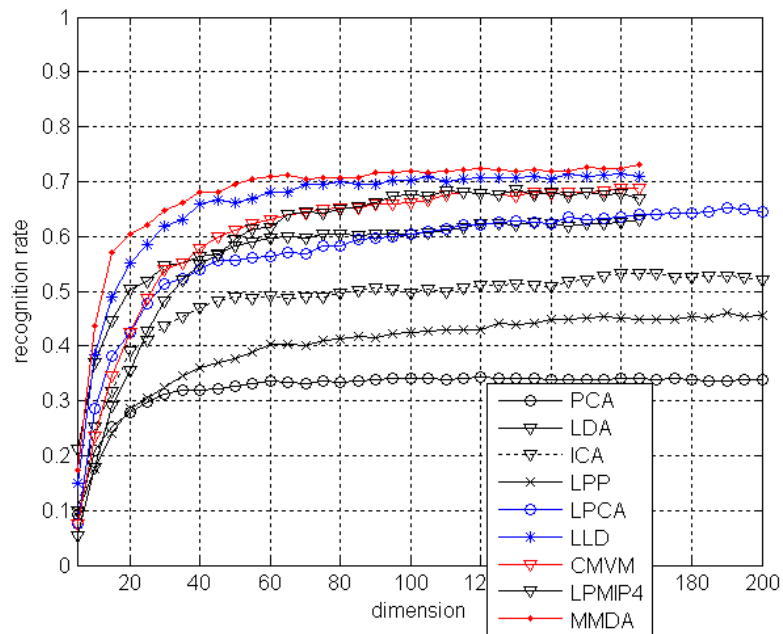
Figure 6. The recognition rates versus different dimension on the PolyU FKP database.

According to the protocol of this database, the images captured in the first session were used for training and the images captured in the second session for testing. Thus, for each class, there are six training samples and six testing samples. PCA, LDA, LPP, LPCA, LLD, CMVM, LPMIP and the proposed MMDA are used for FKP feature extraction. In the PCA phase of LDA, LPP, LLD and MMDA, the number of principal components was set as 200. In LPP, the number of the nearest

neighbors is set as 5. After feature extraction, a nearest neighbor classifier with cosine distance is employed for classification. The maximal recognition rate of each method and the corresponding dimension are listed in Table 8. As on the face databases, we see that MMDA achieves the highest recognition rate. Fig. 6 shows the recognition rate curve versus the variation of the dimensions and the proposed method consistently outperforms over the other methods.

4.4 Discussion

From the above results, we can conclude that MMDA has good performance and surpasses the other competing methods. 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. Both CMVM and LPMIP are multi-manifold methods. However, CMVM only uses the between-class scatter to characterize the multi-manifold information and it ignores the distribution information between sub-manifolds. LPMIP uses nonlocal scatter to characterize the multi-manifold information and but it ignores the class label information. Different training samples have different contributions to the mean of each class, which is explicitly considered in the proposed MMDA to characterize the sub-manifold information and multi-manifold information. Compared with LPP, CMVM and LPMIP, MMDA uses the class label information and does not need to choose the number of nearest neighbors. In LPP, CMVM and LPMIP, it is difficult to choose the suitable number of nearest neighbors in real applications, and class label information is not used. Compared to LDA, MMDA considers the contribution of each training sample to the class center, which is important to characterize the within-class scatter. In addition, MMDA considers the influence of the distances between class centers, which is not considered in LDA and CMVM. Experimental results on three benchmark face

databases and PolyU FKP database verify the effectiveness of MMDA, indicating that MMDA can obtain a robust subspace that approximates the intrinsic geometric structure of the image manifold.

5 Conclusions

Recently, many manifold learning based image feature extraction methods have been proposed. To model multi-manifolds for classification, 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. To this end, we proposed the multi-manifold discriminant analysis (MMDA) for feature extraction and recognition. MMDA is based on graph embedded learning and is under the Fisher discriminant analysis framework. In MMDA, two graphs are constructed to characterize the within-class compactness and the between-class separability, which are corresponding to the within-class scatter matrix and between-class scatter matrix in Fisher LDA. The within-class graph characterizes the sub-manifold information of each class and the between-class graph characterizes the multi-manifold information of different classes. The experimental results on benchmark face databases (FERET, AR, ORL) and the PolyU FKP database showed that MMDA outperforms many representative and state-of-the-art subspace learning methods.

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 Transactions on Pattern Ana1ysis and Machine Intel1igence 13 (3) (1991) 252-264.

[3] D. L. Swets, J. Weng, Using discriminant eigenfeatures for image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (8) (1996) 831-836.

[4] V. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs Fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Ana1ysis and Machine Intel1igence 19 (7) (1997) 711-720.

[5] J. Yang, J. Yang, Why can LDA be performed in PCA transformed space? Pattern Recognition 36 (2) (2003) 563-566.

[6] Z. Hong, J. Yang, Optimal discriminant plane for a small number of samples and design method of classifier on the plane, Pattern Recognition 24 (4) (1991) 317-324.

[7] J.H. Friedman, Regularized discriminant analysis, Journal of the American Statistical Association 84 (1989) 165-175.

[8] T. Hastie and R.Tibshirani, Penalized discriminant analysis, The Annals of Statistics, 23 (1995) 73-102.

[9] T. Hastie, R. Tibshirani, A. Buja, Flexible discriminant analysis by optimal scoring, Journal of the American Statistical Association 89 (1994) 1255-1270.

[10] L.F. Chen, H.Y.M. Liao, M.T. Ko, G. J. Yu. A new LDA-based face recognition system which can solve the small sample size problem, Pattern Recognition 33 (1) (2000) 1713-1726.

[11] H. Yu, J. Yang, A direct LDA algorithm for high dimensional data-with application to face recognition, Pattern Recognition 34 (10) (2001) 2067-2070.

[12] M. Loog, Approximate pairwise accuracy criterion for multiclass linear dimension reduction: generalization of the Fisher criterion, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (7) (2001) 762-766.

[13] R. Lotlikar and R. Kothari. Fractional-step dimensionality reduction, IEEE Transactions on Pattern Recognition and Machine Intelligence 22 (6) (2000) 623-627.

[14] Q. Gao, L. Zhang and D. Zhang. Face recognition using FLDA with single training image per-person, Applied Mathematics and Computation 205 (12) (2008) 726-734.

[15] J.B. Tenenbaum, V.de. Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319-2323.

[16] S.T. Roweis and L.K. Saul. Nonlinear dimension reduction by locally linear embedding, Science, 290 (2000) 2323-2326.

[17] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation, Neural Computation 15 (6) (2003) 1373-1396.

[18] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using laplacianfaces, IEEE Transactions on Pattern Ana1ysis and Machine Intel1igence 27 (3) (2005) 328-340.

[19] J. Yang, D. Zhang, J. Yang, B. Niu, Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics, IEEE Transactions on Pattern Ana1ysis and Machine Intel1igence, 29 (4) (2007) 650-664.

[20] Y. Xu, A. Zhong, J. Yang, D. Zhang, LPP solution schemes for use with face recognition, Pattern

Recognition 43(12) (2010) 4165-4176.

[21] D. Zhao, Z. Lin, and X. Tang. Laplacian PCA and its applications, Proc. IEEE Conf. Computer Vision 2007, Rio de Janeiro, pp. 1-8, Oct.

[22] D. Zhao, Z. Liu, R. Xiao, X. Tang. Linear Laplacian Discrimination for Feature Extraction, Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2007.

[23] D. Hu, G. Feng, Z. Zhou, Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition, Pattern Recognition 40(1) (2007) 339-342.

[24] G. Feng, D. Zhang, J. Yang, D. Hu, A Theoretical framework for matrix-based feature extraction algorithms with its application to image recognition. International Journal of Image and Graphics, 8(1) (2008) 1-23.

[25] S. Yan, D. Xu, B. Zhang, and H. Zhang, Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Transaction on Pattern Ana1ysis and Machine Intelligence 29 (1) (2007) 40-51.

[26] H. Chen, H. Chang, and T. Liu, Local discriminant embedding and its variants, Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR2005). Vol.2, p.846-853.

[27] B. Li, D. S. Huang, C. Wang, K. H. Liu, Feature extraction using constrained maximum variance mapping, Pattern Recognition 41 (11) (2008) 3287-3294.

[28] H. Wang, S. Che, Z. Hu, W. Zheng, Locaility-preserved maximum information projection, IEEE Transactions on Neural Networks 19 (4) (2008) 571-585.

[29] M. Kirby and L. Sirovich, Application of the KL Procedure for the characterization of human faces, IEEE Transactions on Pattern Ana1ysis and Machine Intelligence 12 (1) (1990) 103-108.

[30] J.M. Lee, Riemannian manifolds: an introduction to curvature. Springer, 1997.

[31] M. S. Bartlett, J. R. Movellan and T.J.Sejnowski, Face recognition by independent component analysis, IEEE Transaction on Neural Network, 13 (6) (2002) 1450-1462.

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

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

[34] A.M. Martinez and R. Benavente, The AR Face Database <http://cobweb.ecn.purdue.edu/~aleix/ aleix_face_DB.html>

[35] A.M. Martinez and R. Benavente, The AR Face Database, CVC Technical Report #24, June 1998

[36] L. Zhang, L. Zhang, D. Zhang, and H. Zhu, Online Finger-Knuckle-Print Verification for Personal Authentication, Pattern Recognition, 43 (7) (2010) 2560-2571.

[37] L. Zhang, L. Zhang, D. Zhang, Finger-knuckle-print: a new biometric identifier, Proceedings of the IEEE International Conference on Image Processing, 2009.

[38] The Hong Kong PolyU FKP database: http://www4.comp.polyu.edu.hk/~biometrics/FKP.htm