Laplacian MinMax Discriminant Projection and its Applications

Zhonglong Zheng and Xueping Chang

Department of Computer Science, Zhejiang Normal University, Jinhua, China
Email: zhonglong@sjtu.org

Jie Yang

Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China

A new algorithm, Laplacian MinMax Discriminant Projection (LMMDP), is proposed in this paper for supervised dimensionality reduction. LMMDP aims at learning a discriminant linear transformation. Specifically, we define the within-class scatter and the between-class scatter using similarities which are based on pairwise distances in sample space. After the transformation, the considered pairwise samples within the same class are as close as possible, while those between classes are as far as possible. The structural information of classes is contained in the within-class and the between-class Laplacian matrices. Therefore, the discriminant projection subspace can be derived by controlling the structural evolution of Laplacian matrices. The performance on several data sets demonstrates the competence of the proposed algorithm.

ACM Classification: I.5

Keywords: Manifold Learning; Dimensionality Reduction; Supervised Learning; Discriminant Analysis

1. INTRODUCTION

Dimensionality reduction has attracted tremendous attention in the pattern recognition community over the past few decades and many new algorithms have been developed. Among these algorithms, linear dimensionality reduction is widely spread for its simplicity and effectiveness. Principal component analysis (PCA), as a classic linear method for unsupervised dimensionality reduction, aims at learning a kind of subspaces where the maximum covariance of all training samples are preserved (Turk,1991). Locality Preserving Projections, as another typical approach for unsupervised dimensionality reduction, seeks projections to preserve the local structure of the sample space (He, 2005). However, unsupervised learning algorithms cannot properly model the underlying structures and characteristics of different classes (Zhao, 2007). Discriminant features are often obtained by supervised dimensionality reduction. Linear discriminant analysis (LDA) is one of the most popular supervised techniques for classification (Fukunaga, 1990; Belhumeur, 1997). LDA aims at learning discriminant subspace where the within-class scatter is minimized and the between-class scatter of samples is maximized at the same time. Many improved LDAs up to date have demonstrated competitive performance in object classification (Howland, 2004; Liu, 2007; Martinez, 2006; Wang and Tang, 2004a; Yang, 2005).

Manuscript received: 15 April 2008
Communicating Editor: Tele Tan

As stated in Zhao (2007), the primary formulation of LDA does not always hold in non-Euclidean Spaces. In this paper, we propose a new supervised dimensionality reduction algorithm, Laplacian MinMax Discriminant Projection (LMMDP), for discriminant feature extraction. The spirit of LMMDP is largely motivated by the traditional LDA, Laplacian Eigenmaps (LE) and the nearest neighbourhood selection strategy (Weinberger, 2006; Nie, 2007). In our algorithm, we only focus on the farthest neighbourhood within the same class and nearest neighbourhood between different classes at the same time. We formulate the within-class scatter and the between-class scatter by means of similarity criterions which were commonly used in LE and LPP (Locality Preserving Projection). The extended within-class scatter and the between-class scatter are governed by the within-class Laplacian matrix and the between class Laplacian matrix. Generally, LDA can be regarded as a special case of LMMDP. Therefore, LMMDP not only conquers the non-Euclidean space problem but also provides an alternative way to find potential better discriminant subspaces.

The paper is organized as follows. In Section 2, we provide a brief introduction of the related work. In Section 3, the proposed Laplacian MinMax Discriminant Projection is described in detail. The experimental results and the performance comparisons are presented in Section 4. Section 5 covers some conclusions.

2. OVERVIEW OF LINEAR DISCRIMINANT ANALYSIS

Let $X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{D\times n}$ denote a data set matrix which consists of n samples $\{x_i\}_{i=1}^n \in \mathbb{R}^D$. Linear dimensionality reduction algorithms focus on constructing a small number, d, of features by applying a linear transformation $W \in \mathbb{R}^{D\times d}$ that maps each sample data x_i of X to the corresponding vector $\{y_i\}_{i=1}^n \in \mathbb{R}^d$ in d-dimensional space as follows:

$$W:\ x_i \in \mathbb{R}^D \rightarrow y_i = W^T x_i \in \mathbb{R}^d. \tag{1}$$

Assume that the matrix X contains c classes, and is ordered such that samples appear by class

$$X = [x_1^1, \ldots, x_{c_1}^1, \ldots, x_1^c, \ldots, x_{c_c}^c]. \tag{2}$$

In traditional LDA, two scatter matrices, i.e., within-class matrix and between-class matrix are defined as follows (Fukunaga, 1990):

$$S_w = \frac{1}{n}\sum_{i=1}^{c}\sum_{x\in X_i}(x-\bar{m}_i)(x-\bar{m}_i)^T \tag{3}$$

$$S_b = \frac{1}{n}\sum_{i=1}^{c} n_i(\bar{m}_i-\bar{m})(\bar{m}_i-\bar{m})^T \tag{4}$$

where n_i is the number of samples in the *i–th* class X_i, $\bar{m}_i$ is the mean vector of the *i–th* class, and $\bar{m}$ is the mean vector of total samples. It follows from the definition that $trace(S_w)$ measures the within-class compactness, and $trace(S_b)$ measures the between-class separation.

The optimal transformation matrix W obtained by traditional LDA is computed as follows (Fukunaga, 1990):

$$W_{opt} = \arg\max_w \frac{tr(W^T S_b W)}{tr(W^T S_w W)}. \tag{5}$$

To solve the above optimization problem, the traditional LDA computes the following generalized eigenvalue equation

$$S_b w_i = \lambda S_w w_i, \tag{6}$$

and takes the d eigenvectors that are associated with the d largest eigenvalues $\lambda_i, i = 1, \ldots, d$.

3. LAPLACIAN MINMAX DISCRIMINANT PROJECTIONS

Let x_i^s denote the *i–th* sample in *s–th* class. We formulate Equation 1 as follows:

$$y_i^s = W^T x_i^s \qquad s.t. \qquad W^T W = I \tag{7}$$

Assume that each x_i^s have two kinds of neighbourhoods: within-class neighbourhood $\mathfrak{n}_{fw}$ and between-class neighbourhood $\mathfrak{n}_{nb}$. $\mathfrak{n}_{fw}$ implies the set of k_{fw} farthest sample points of (x_i^s) in the same class. $\mathfrak{n}_{nb}$ implies the set of k_{nb} nearest sample points of (x_i^s) in different classes. Then we obtain $1 \le k_{fw} \le n_i - 1$, and $1 \le k_{nb} \le n - n_i$, where n_i is the number of samples in class i.

After the transformation derived from LMMDP, we hope that the considered sample points will be closer within the same class, and those between different classes will be farther as shown in Figure 1.

3.1 Discriminant Within-Class Scatter

We define the within-class scatter of class s as $\mathfrak{W}_s$ as follows:

$$\mathfrak{W}_s = \sum_{i=1}^{c_s} \alpha_i^s \sum_{j=1}^{k_{fw}} \| y_i^s - y_j^s \|^2 \tag{8}$$

where y_j^s implies its corresponding original sample x_j^s belonging to $\mathfrak{n}_{fw}$ neighbourhood of (y_i^s)'s corresponding original sample x_i^s. α_i^s is the weight, defined as

$$\alpha_i^s = exp\left(-\frac{\| x_i^s - \overline{x}_{\mathfrak{n}_{fw}(x_i^s)} \|^2}{t}\right), \quad i = 1, \ldots, c_s \tag{9}$$

where t is the time variable as that of in LE and LPP algorithms (Belkin, 2003; He, 2005). It suffices to note that we adopt the distance measure $\mathbb{R}^D$ (in Equation 9) by Euclidean norm for the simplification reason. In fact, the distance measure can be other norms depending on the metric of the original sample space which may be Euclidean or non-Euclidean.

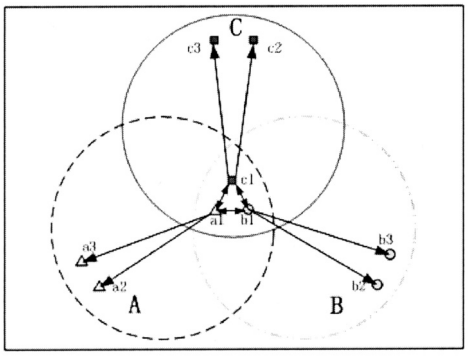


Figure 1: There are three classes A, B and C illustrated above. In class A, a2 and a3 are a1's within-class $\mathfrak{n}_{fw}$ neighbourhood. b1 and b2 are a1's between-class $\mathfrak{n}_{nb}$ neighbourhood. After the transformation, we try to make a1 close to a2 and a3, while far from b1 and c1. For class B and C, the analysis is similar to that of class A.

The above Eq.(8) can be re-formulated as follows to obtain its compact form:

$$\begin{aligned}\mathfrak{W}_s &= \sum_{i=1}^{c_s} \alpha_i^s tr\{\sum_{j=1}^{k_{fw}} (y_i^s - y_j^s)(y_i^s - y_j^s)^T\} \\ &= \sum_{i=1}^{c_s} \alpha_i^s tr\{\sum_{j=1}^{k_{fw}} (y_i^s (y_i^s)^T - 2y_i^s (y_j^s)^T + y_j^s (y_j^s)^T\} \\ &= \sum_{i=1}^{c_s} \alpha_i^s \{k_{fw} tr(y_i^s (y_i^s)^T) - 2tr(y_i^s \sum_{j=1}^{k_{fw}} (y_j^s)^T) + tr\sum_{j=1}^{k_{fw}} (y_j^s (y_j^s)^T)\} \\ &= \sum_{i=1}^{c_s} \alpha_i^s \{k_{fw} tr(y_i^s (y_i^s)^T) - 2tr(y_i^s e_{k_{fw}}^T Y_{\mathfrak{fw}(y_i^s)}^T) + tr(Y_{\mathfrak{fw}(y_i^s)} Y_{\mathfrak{fw}(y_i^s)}^T)\}\end{aligned} \tag{10}$$

where $Y_{\mathfrak{fw}(y_i^s)}$ consists of $\mathfrak{n}_{fw}$ neighbourhood of y_i^s, and $e_{k_{fw}}$ is all one column vector of length k_{fw}. Let, $Y_s = (y_1^s, y_2^s, \ldots, y_{c_s}^s)$, $\alpha_s = diag(\alpha_1^s, \alpha_2^s, \ldots, \alpha_{c_s}^s)$, then Equation 10 can be re-written as follows:

$$\begin{aligned}\mathfrak{W}_s &= \sum_{i=1}^{c_s} \{k_{fw} \alpha_i^s tr(y_i^s (y_i^s)^T) - 2\alpha_i^s tr(y_i^s e_{k_{fw}}^T Y_{\mathfrak{fw}(y_i^s)}^T) + \alpha_i^s tr(Y_{\mathfrak{fw}(y_i^s)} Y_{\mathfrak{fw}(y_i^s)}^T)\} \\ &= k_{fw} tr(Y_s \alpha_s Y_s^T) - 2tr(Y_s \alpha_s E_{k_{fw}} Y_{\mathfrak{fw}(y_i^s)}^T)) + tr(Y_{\mathfrak{fw}(y_i^s)} A_s Y_{\mathfrak{fw}(y_i^s)}^T)\end{aligned} \tag{11}$$

where $A_s = diag(\underbrace{\alpha_s, \alpha_s, \ldots, \alpha_s}_{k_{fw}})$. There exists a 0–1 indicator matrix $S_{\mathfrak{fw}_{(y_i^s)}}$ satisfying

$$Y_{\mathfrak{fw}(y_i^s)} = Y_s S_{\mathfrak{fw}_{(y_i^s)}}. \tag{12}$$

Substituting Equation 12 to Equation 11 gives

$$\begin{aligned}\mathfrak{W}_s &= k_{fw} tr(Y_s \alpha_s Y_s^T) - 2tr(Y_s \alpha_s E_{k_{fw}} S_{\mathfrak{fw}_{(y_i^s)}}^T Y_s^T) + tr(Y_s S_{\mathfrak{fw}_{(y_i^s)}} A_s S_{\mathfrak{fw}_{(y_i^s)}}^T Y_s^T) \\ &= tr(Y_s L_s Y_s^T),\end{aligned} \tag{13}$$

where

$$L_s = k_{fw} \alpha_s - 2\alpha_s E_{k_{fw}} S_{\mathfrak{fw}_{(y_i^s)}}^T + S_{\mathfrak{fw}_{(y_i^s)}} A_s S_{\mathfrak{fw}_{(y_i^s)}}^T. \tag{14}$$

Let $\mathfrak{W}_s$ denote the total within-class scatter of all samples, then we obtain

$$\mathfrak{W} = \sum_{s=1}^{c} \mathfrak{W}_s = \sum_{s=1}^{c} tr(Y_s L_s Y_s^T) \tag{15}$$

There exists a 0–1 indicator matrix P_s satisfying $Y_s = YP_s$. For supervised learning, the known class information is recorded in column vector of P_s. Substituting Equation 15 to Equation 14 gives

$$\mathfrak{W} = \sum_{s=1}^{c} YP_s L_s P_s^T Y^T = tr(Y\mathfrak{L}_{\mathfrak{W}} Y^T), \tag{16}$$

where

$$\mathfrak{L}_{\mathfrak{W}} = \sum_{s=1}^{c} P_s L_s P_s^T \tag{17}$$

can be viewed as the *within-class Laplacian matrix*.

Taking $Y = W^T X$ into account, we re-write Equation 16 as follows

$$\mathfrak{W} = tr(W^T \mathfrak{D}_{\mathfrak{W}} W). \tag{18}$$

where $\mathfrak{D}_{\mathfrak{W}} = X \mathfrak{L}_{\mathfrak{W}} X^T$ is the total *within-class scatter matrix*.

3.2 Discriminant Between-Class Scatter

Let B_s denote the between-class scatter of class s. B_s is defined as

$$B_s = \sum_{i=1}^{c_s} \alpha_i^s \sum_{j=1}^{k_{nb}} \| y_i^s - y_j \|^2, \tag{19}$$

where y_j implies its corresponding original sample x_j^s belonging to $\mathfrak{n}_{nb}$ neighbourhood of (y_i^s)'s corresponding original sample x_i^s. α_i^s is the weight, defined as

$$\alpha_i^s = exp(-\frac{\| x_i^s - \overline{x}_{\mathfrak{n}_{nb}(x_i^s)} \|^2}{t}), \qquad i = 1, \ldots, c_s. \tag{20}$$

By the similar deduction as that in Equation 10 and 11, B_s can be formulated as follows

$$B_s = k_{nb} tr(Y_s \alpha_s Y_s^T) - 2tr(Y_s \alpha_s E_{k_{nb}} Y_{\mathfrak{nb}(y^s)}^T)) + tr(Y_{\mathfrak{nb}(y^s)} A_s Y_{\mathfrak{nb}(y^s)}^T), \tag{21}$$

where $Y_{\mathrm{nb}(y^s)}$ consists of $\mathfrak{n}_{nb}$ neighbourhood of each data in y^s.

There exists a 0–1 indicator matrix $S_{\mathrm{nb}_{(y^s)}}$ satisfying $Y_{\mathrm{nb}(y^s)}) = YS_{\mathrm{nb}_{(y^s)}}$. Equation 21 can be re-written as

$$\begin{aligned} B_s &= k_{nb} tr(YP_s \alpha_s P_s^T Y^T) - 2tr(YP_s \alpha_s E_{k_{nb}} S_{\mathfrak{nb}_{(y^s)}}^T Y^T) + tr(YS_{\mathfrak{nb}_{(y^s)}} A_s S_{\mathfrak{nb}_{(y^s)}}^T Y^T) \\ &= tr(YL_{bs} Y^T), \end{aligned} \tag{22}$$

where $L_{bs} = k_{nb} P_s \alpha_s P_s^T - 2P_s \alpha_s E_{k_{nb}} S_{\mathfrak{nb}_{(y^s)}}^T + S_{\mathfrak{nb}_{(y^s)}} A_s S_{\mathfrak{nb}_{(y^s)}}^T$.

Let $\mathfrak{B}$ denote the total between-class scatter of all classes, then we obtain

$$\begin{aligned} \mathfrak{B} &= \sum_{s=1}^{c} B_s = \sum_{s=1}^{c} tr(YL_{bs} Y^T) \\ &= tr(Y \mathfrak{L}_{\mathfrak{B}} Y^T), \end{aligned} \tag{23}$$

where

$$\mathfrak{L}_{\mathfrak{B}} = \sum_{s=1}^{c} L_{bs} \tag{24}$$

can be viewed as the *between-class Laplacian matrix*.

Taking $Y = W^T X$ into account, we re-write Equation 23 as follows

$$\mathfrak{B} = tr(W^T \mathfrak{D}_{\mathfrak{B}} W). \tag{25}$$

where $\mathfrak{D}_{\mathfrak{B}} = X \mathfrak{L}_{\mathfrak{B}} X^T$ is the total *between-class scatter matrix*.

3.3 Discriminant Projection

We construct the following Fisher criterion

$$f(W) = \arg\max_w \frac{\mathfrak{B}}{\mathfrak{W}} = \frac{tr(W^T \mathfrak{D}_{\mathfrak{B}} W)}{tr(W^T \mathfrak{D}_{\mathfrak{W}} W)}. \tag{26}$$

To solve the above optimization problem, we take the similar approach used in the traditional LDA. We take the d eigenvectors derived from the following generalized eigenvalue analysis

$$\mathfrak{D}_{\mathfrak{B}} w_i = \lambda_i \mathfrak{D}_{\mathfrak{w}} w_i \tag{27}$$

that are associated with the d largest eigenvalues $\lambda_i, i = 1,\ldots,d$.

The idea of Laplacian MinMax (minimize within-class scatter and maximize between-class scatter) discriminant projection is to make the samples within the same class cluster as compact as possible and samples between classed separate as far as possible in Laplacian aspect. Different from (Nie, 2007; Yan, 2005), LMMDP mainly focuses on samples far apart within the same class and samples nearby between different classes. It is easy to understand that the overlap of different classes is mainly due to the outlier samples in classes. Therefore, the outliers usually lead to misclassifications. The minimization of Equation 18 means pulling the outliers in the same class near the class centre, to some extent. When the minimization of Equation 18 is combined with the maximization of Equation 23, the projections obtained by the proposed LMMDP, i.e., Equation 26, demonstrate discriminant power. In addition, it should be noted that the distance between samples in the original sample space are measured Euclidean norm for simplification reasons in this paper. In fact, the distance measure can be other norms depending on the metric of the original sample space which may be Euclidean or non-Euclidean.

4. COMPARATIVE EXPERIMENTAL RESULTS

In this section, we investigate the use of LMMDP on several data sets including UCI (Available at http://www.ics.uci.edu/ mlearn/MLRepository.html), USPS (Available at http://www.kernel-machines.org/data) and PIE-CMU face data set (Sim, 2002). The data sets used in the paper belong to different fields in order to test the performance of LMMDP algorithm. We compare our proposed algorithm with PCA (Turk, 1991), LDA (Belhumeur, 1997), LPP (He, 2005) and Marginal Fisher Analysis (MFA) (Yan, 2005).

4.1. On UCI Sata Set

In this experiment, we perform on iris data taken from the UCI Machine Learning Repository. There are 150 samples of 3 classes (50 samples per class) in iris data set. We randomly select 20 samples per class for training and the remaining samples for testing. The average results are obtained over 50 random splits. All algorithms reduce the original samples to 2-dimensional space. The classification is based on k-nearest neighbour classifier. The experimental results are shown in Table 1. In terms of LMMDP algorithm, there are several parameters which should be set before the experiments. In the experiment, $k_{fw} = 15$, $k_{nb} = 20$, and the time variable $t = 10$.

Algorithm	PCA	LDA	LPP	MFA	LMMDP
Accuracy	95.112	95.391	95.391	95.383	95.891

Table 1: Recognition accuracy of different algorithms

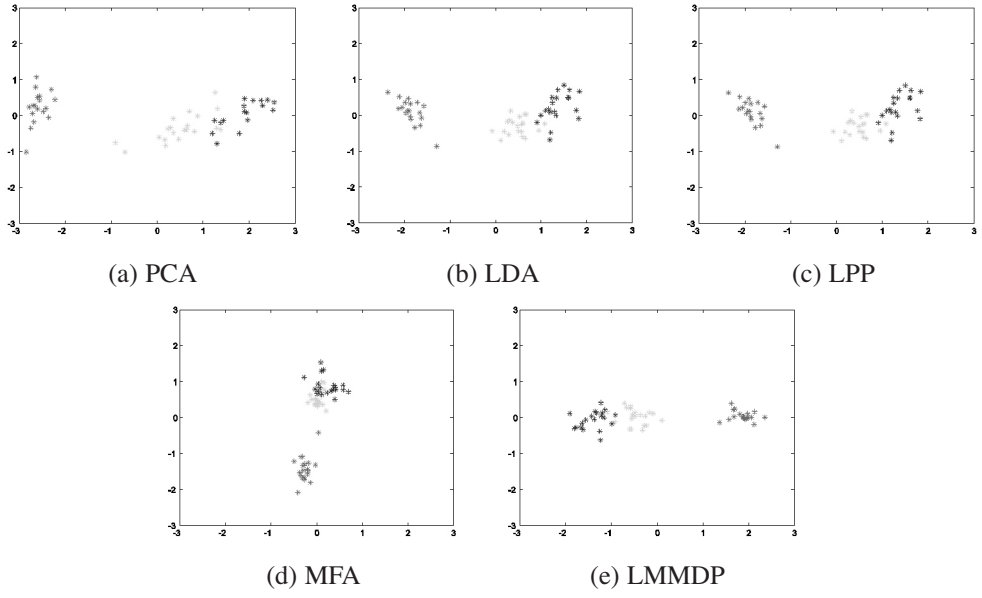
To demonstrate the performance of LMMDP algorithm, we randomly select one split from the 50 splits. The embedding results of LMMDP in 2D space, together with the other four algorithms, are shown in Figure 2.

As illustrated in Table 1, LMMDP algorithm outperforms other methods with a recognition rate of 93.891%. We can find that the within-class embedding result of LMMDP is more compact than those of the other four methods, as illustrated in Figure 2.

4.2. On USPS Data Set

In this experiment, we focus on the digit recognition task using the USPS data set. It contains normalized grey scale images size 16×16, divided into a training set of 7,291 images and a testing set of 2,007 images. For convenience, we randomly select 500 samples per class in the original training set as our training set. That is to say, there are 5,000 training samples in our training set. Our testing set is the same as the original testing set. The average results with corresponding reduced dimensions are obtained over 50 random splits. The classification is based on k-nearest neighbour classifier. The experimental results are shown in Table 2. In the experiment, the parameters of LMMDP algorithm are set as $k_{fw} = 105$, $k_{nb} = 215$, and the time variable $t = 10$.

On this data set as described in Table 2, LMMDP algorithm also makes a little more improvement than the others. It suffices to note that LDA fails because its available embedding number may be insufficient under this case.



(a) PCA (b) LDA (c) LPP

(d) MFA (e) LMMDP

Figure 2: Embedding results in 2-D space of PCA, LDA, LPP, MFA and LMMDP

Algorithm	PCA	LDA	LPP	MFA	LMMDP
Accuracy	85.071	79.521	86.336	87.368	87.821
Dims	51	9	39	31	53

Table 2: Best recognition accuracy of different algorithms

4.3. On PIE-CMU Face Data Set

The PIE-CMU face data set consists of 68 subjects with 41,368 face images (Sim, 2002). In this experiment, we select 40 subjects with 120 face images for each from the CMU data set, 60 images for training, and the other 60 images for testing. Before the experiment, faces in the images are detected by the face detection system described in Zheng (2006). The detected faces are converted to grey scale images and resized to 32×32. Some samples are shown in Figure 3. Totally, there are 2,400 images in the training set and the testing set, respectively.

It should be mentioned that we take PCA as a preprocessing step for LMMDP. The number of principal components is a free parameter to choose. As pointed out in Wang and Tang (2004b) and Wang and Tang (2006), the dimension of principal subspaces significantly affects the performance of recognition tasks. Besides, they confirmed that the optimal number lies in the interval [50,200]. Based on their work, we find the best dimension of PCA is 182. Therefore, we take 182 as the number of principal components in the following experiments.

For the sake of visualization, we illustrate algorithmic-faces derived from different algorithms, such as Eigenfaces from PCA, Fisherfaces from LDA and Laplacianfaces from LPP, in Figure 4. The special face-like images derived from MFA and LMMDP can be called MFAfaces and LMMDPfaces, respectively.

Figure 3: Some samples of CMU-PIE face data set

Figure 4: From the top row to the bottom row, the face-like images are Eigenfaces, Fisherfaces, Laplacianfaces, MFAfaces and LMMDPfaces, respectively.

Algorithm	PCA	LDA	LPP	MFA	LMMDP
Accuracy	69.786	79.510	79.533	83.328	85.480
Dims	180	39	91	58	104

Table 3: Best recognition accuracy of different algorithms

The average results with corresponding reduced dimensions are obtained over 50 random splits. The classification is also based on k-nearest neighbour classifier. The experimental results are shown in Table 3. In the experiment, the parameters of LMMDP algorithm are set as $k_{fw} = 23$, $k_{fw} = 36$, and the time variable $t = 10$.

4.4 On Local Features

Still, we investigate the performance of the proposed LMMDP algorithm on local features. Here, we exploit Gabor wavelets to extract local features. The 2D Gabor functions proposed by Daugman are local spatial band-pass filters that achieve the theoretical limit for conjoint resolution of information in the 2D spatial and 2D Fourier domains, that is, Gabor wavelets exhibit desirable characteristics of spatial locally and orientation selectivity (Daugman, 1980). Donato *et al* (1999) had shown through experiments that the Gabor wavelet representation gives better performance than other techniques for classifying facial actions.

The Gabor wavelets (kernels, filters) can be defined as:

$$\Psi_{\alpha,\beta}(z) = \frac{\| k_{\alpha,\beta}\|^2}{\sigma^2} e^{(-\| k_{\alpha,\beta}\|^2 \| z\|^2 / 2\sigma^2)} [e^{ik_{\alpha,\beta} z} - e^{-\sigma^2/2}] \tag{28}$$

where α and β define the orientation and scale of the Gabor kernels, $\|\cdot\|$ denotes the norm operator, $z = (x, y)$, and the wave vector $k_{\alpha,\beta}$ is defined as:

$$k_{\alpha,\beta} = k_\beta e^{i\phi_\alpha} \tag{29}$$

where $k_\beta = k_{max} / f^\beta$ and $\phi_\alpha = \pi\alpha / 8$. k_{max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain.

By scaling and rotation of $k_{\alpha,\beta}$, all self-similar Gabor kernels in Equation 28 can be generated from one filter, the mother wavelet. Each kernel is a product of a Gaussian envelope and a complex plane wave. In the square brackets in Equation 28, the first term and the second term denote the oscillatory part and the DC part of the kernel respectively. If the parameter σ has sufficiently large value, the effect of the DC term becomes negligible. Here the parameters of α and β are set to eight and five respectively. In this paper, we also set α and β to be eight and five. Figure 5 shows the real part of the Gabor kernels at five scales and eight orientations and their magnitudes, with the following parameters: $\alpha = \{0,1,.,7\}$, $\beta = 0,1,.,4$, $k_{max} = \pi / 2, f = 2$, $\phi = 2\pi$. The Gabor kernels show desirable performance of orientation selectivity, frequency and spatial locality.

In order to encompass all frequency and locality information as much as possible, this paper, (same as Liu, 2001), concatenated all the Gabor representations at the five scales and eight orientations. Before the concatenation, $Y_{\alpha,\beta}(z)$ is down-sampled by a factor ρ to reduce the space dimension, and normalized to zero mean and unit variance. We then construct a vector out of the $Y_{\alpha,\beta}(z)$ by concatenating its rows (or columns). Now let $Y^\rho_{\alpha,\beta}(z)$ denote the normalized vector $Y_{\alpha,\beta}(z)$ constructed from, the augmented Gabor feature vector Y^ρ is defined as:

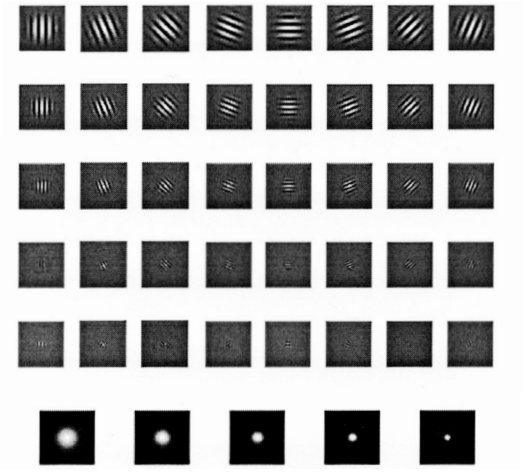


Figure 5: Gabor wavelets at 5 scales and 8 orientations

Algorithm	PCA	LDA	LPP	MFA	LMMDP
Accuracy	70.51	85.10	83.29	86.72	87.30
Dims	95	39	111	111	87

Table 4: Best recognition accuracy of different algorithms

$$Y^{\rho} = (Y^{\rho}_{\alpha,\beta} \mid \alpha = 0,.,7; \beta = 0,.,4) \tag{30}$$

Then Y^{ρ}, which is a row vector or column vector, serves as the original space performing different methods for recognition.

To test the performance of the LMMDP algorithm on local features, we apply LMMDP to a face recognition task based on PIE-CMU data set. In this experiment, the training and the testing set are the same as mentioned in subsection 4.3. The average results with corresponding reduced dimensions are obtained over 50 random splits. The classification is also based on k-nearest neighbour classifier. The experimental results are shown in Table 4.

5. CONCLUSION

In this paper, we have proposed a new method, Laplacian MinMax Discriminant Projection (LMMDP), based on Laplacian eigenmap and LDA algorithms for supervised dimensionality reduction. Using similarity weighted discriminant criterions, we define the within-class Laplacian matrix and between-class Laplacian matrix. LMMDP focuses on the farthest points (n_{fw}) within class and nearest points (n_{nb}) between class. n_{fw} is minimized, meanwhile n_{nb} is maximized. In comparison with the traditional LDA, LMMDP focuses more on the enhancement of the discriminability of local structures. Therefore, LMMDP has the flexibility of finding optimal discriminant subspaces.

Experiments are performed on several real data sets. Experimental results indicate that discriminant criterions formulated in LMMDP are more suitable for discriminant feature extraction, no

matter whether the sample space is Euclidean or not. The performance of LMMDP will be further enhanced by trying other improved LDA strategies. In addition, how to choose the best parameters of LMMDP will be an interesting direction for future study.

ACKNOWLEDGEMENT

The authors confirm that the research was supported by National Science Foundation of China (No.60805001).We thank reviewers for their insightful comments and constructive suggestions.

REFERENCES

BELHUMEUR, P.N., HESPANHA, J.P. and KRIEGMAN, D.J. (1997): Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7): 711–720.

BELKIN, M. and NIYOGI, P. (2003): Laplacian Eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15: 1373–1396.

DAUGMAN, J.G. (1980): Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20: 847–856.

DONATO, G., BARTLETT, M.S. and HAGER, J.C. (1999): Classifying facial actions. *IEEE Trans. Pattern Anal. Machine Intell.*, 21: 974–989.

FUKUNAGA, K. (1990): Introduction to statistical pattern recognition. Second Edition, Academic Press, Boston, MA.

HE, X., YAN, S., HU, P.N.Y.X. and ZHANG, H. (2005): Face recognition using Laplacianfaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(3): 328–340.

HOWLAND, P. and PARK, H. (2004): Generalizing discriminant analysis using the generalized singular value decomposition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(8): 995–1006.

LIU, C. (2007): Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(5): 725–737.

LIU, C. and WECHSLER, H. (2001): A Gabor feature classifier for face recognition. *Intl. Conf. Computer Vision*.

MARTINEZ, A. and ZHU, M. (2006): Where are linear feature extraction methods applicable. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(12): 1934–1944.

NIE, F., XIANG, S. and ZHANG, C. (2007): Neighbourhood MinMax Projections. IJCAI: 993–998.

SIM, T., BAKER, S. and BSAT, M. (2002): The CMU pose, illumination, and expression (PIE) database, in *Proceedings of the IEEE International Conference of Automatic Face and Gesture Recognition*.

TURK, M. and PENTLAND, A. (1991): Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1): 71–86.

WANG, X. and TANG, X. (2004a): Dual-space linear discriminant analysis for face recognition. *Proc. Intl. Conf. Computer Vision and Pattern Recognition*, 564–569.

WANG, X. and TANG, X. (2004b): A unified framework for subspace face recognition. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 26(9): 1222–1228.

WANG, X. and TANG, X. (2006): Random sampling for subspace face recognition. *International Journal of Computer Vision*, 70(1): 91–104.

WEINBERGER, K., BLITZER, J. and SAUL, L. (2006): Distance metric learning for large margin nearest neighbor classification. In *Advances in Neural Information Processing Systems,* 18: 475–1482.

YAN, S., XU, D., ZHANG, B. and ZHANG, H. (2005): Graph embedding: A general framework for dimensionality reduction. *Intl. Conf. Computer Vision and Pattern Recognition*.

YAN, S., XU, D., ZHANG, B., ZHANG, H., YANG, Q. and LIN, S. (2007): Graph embedding and extension: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1): 40–51.

YANG, J., FRANGI, A., YANG, J., ZHANG, D. and JIN, Z. (2005): KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(2): 230–244.

ZHAO, D., LIN, Z., XIAO, R. and TANG, X. (2007): Linear Laplacian discrimination for feature extraction. *Intl. Conf. Computer Vision and Pattern Recognition*, 1–7.

ZHENG, Z.L. and YANG, J. (2006): Supervised locality pursuit embedding for pattern classification, *Image and Vision Computing*, 24: 819–826.

BIOGRAPHICAL NOTES

Zhonglong Zheng received his PhD in Pattern Recognition and Artificial Intelligence from Shanghai JiaoTong University, China in 2005. Since 2006, he has been with the department of computer science of Zhejiang Normal University, China. His research interests include data understanding in machine learning, image analysis and pattern recognition. He is the author of more than 20 scientific publications. His current projects include the National Nature Science Foundation (No. 60805001).

Zhonglong Zheng

Xueping Chang received her BA degree in computer science from Xi'an JiaoTong University, China in 2005. She is now a graduate of the department of computer science, Zhejiang Normal University, China. Her research interests include dimensionality reduction and manifold learning.

Xueping Chang

Jie Yang received the PhD in computer science from the University of Hamburg, Germany in 1994. Dr Yang is now the professor and director of the Institute of Image Processing and Pattern Recognition of Shanghai JiaoTong University, China. His research interests include artificial intelligence, image processing, computer vision, data mining, and pattern recognition. He is the author of more than 80 scientific publications and books.

Jie Yang