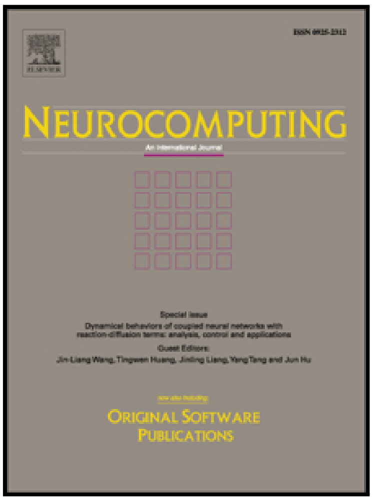


A Survey on Laplacian Eigenmaps Based Manifold Learning Methods

Bo Li , Yan-Rui Li , Xiao-Long Zhang

PII: S0925-2312(18)31264-5
DOI: https://doi.org/10.1016/j.neucom.2018.06.077
Reference: NEUCOM 20088

To appear in: *Neurocomputing*

Received date: 13 March 2018
Revised date: 28 May 2018
Accepted date: 6 June 2018

A Survey on Laplacian Eigenmaps Based Manifold Learning Methods

Bo Li [1,2] , Yan-Rui Li [1,2] , and Xiao-Long Zhang [1,2]

[1]School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, Hubei, 430065, China

[2] Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan, Hubei, 430065, China

Abstract: As a well known nonlinear dimensionality reduction method, Laplacian Eigenmaps (LE) aims to find low dimensional representations of the original high dimensional data by preserving the local geometry between them. LE has attracted great attentions because of its capability of offering useful results on a broader range of manifolds. However, when applying it to some real-world data, several limitations have been exposed such as uneven data sampling, out-of-sample problem, small sample size, discriminant feature extraction and selection, etc. In order to overcome these problems, a large number of extensions to LE have been made. So in this paper, we make a systematical survey on these extended versions of LE. Firstly, we divide these LE based dimensionality reduction approaches into several subtypes according to different motivations to address the issues existed in the original LE. Then we successively discuss them from strategies, advantages or disadvantages to performance evaluations. At last, the future works are also suggested after some conclusions are drawn.

Keywords: Dimensionality reduction, Laplacian eigenmaps, Manifold learning

1. Introduction

In the fields of machine learning, pattern recognition and data mining, the original data to be handled are often characterized by high dimensionality, which easily results in the curse of dimensionality problem. Thus many kinds of methods have to be presented to reduce the dimensions of the original data. Dimensionality reduction aims to find low dimensional compact projections of the original data, which can make the following contributions. One is to reduce the computational expense by using the low dimensional representations of the original data to model different statistical leaning approaches. Another is that dimensionality reduction can preserve the geometry structure of the original data as much as possible in the low dimensional subspace, which makes it more suitable for data visualization. In addition, dimensionality reduction also benefits to explore more discriminant features especially in pattern recognition and data mining, where it acts as feature extraction or feature selection for data classification. So far, more and more dimensionality reduction approaches have been put forward, which are widely applied in a large number of fields as computer vision [1, 212], biometric recognition [3-4], multi-media retrieval [5-6], natural language processing [7-8] and bioinformatics [9-10].

Table 1 Partition of Dimensionality Reduction Methods

Partition		Classical methods
The linear	Linear transformation	PCA, LDA, ICA
The nonlinear	With explicit transformation	Kernel PCA, Kernel FDA,
	Without explicit transformation	BP Neural networks ISOMAP, LLE, LE

Dimensionality reduction methods can be categorized into the linear and the nonlinear learning models. Represented by the classical principal component analysis (PCA) [11], linear discriminant analysis (LDA) [12] and independent component analysis (ICA) [208], the most notable property of linear dimensionality reduction methods is that there is a linear transformation between the original data and their low dimensional embeddings. At the same time, for nonlinear dimensionality reduction algorithms, two subtypes can also be partitioned based on whether or not an explicit nonlinear projection function exists between the original data and their representations, which are listed in Table 1. Kernel methods [14-16] are regarded to the first kind because there are different kernel functions between them such as inner product kernel, Gaussian kernel, polynomial kernel, exponential Kernel, Laplacian Kernel, etc [182]. But for manifold learning, they can be treated as another kind of nonlinear dimensionality reduction approaches, where no explicit projection function can be found between the original data and their embeddings, thus out-of-sample problem naturally appears to them. As to neural networks [160, 171], they are also introduced to reduce the dimensionality of the original data. It must be noted that the weights between nodes of different lays will be optimized and fixed by training neural networks and it does not occur to neural networks out-of-sample problem. However, similar to manifold learning methods, there is still no explicit projection function between the original data and low dimensional embeddings when using neural networks. In this sense, neural networks can also be classified to the kind of nonlinear dimensionality reduction without explicit transformation. In neural networks, some learning strategies as back-propagation (BP) [161-162] and radial basis probabilistic [20-21, 205-206] have been presented with extensive applications, moreover, they are also newly adopted to find roots of polynomial successfully [17-19]. Recently, deep learning neural networks including auto-encoder [22], convolution neural networks (CNN) [23], canonical correlation analysis networks (CCANets) [24] as well as generative adversarial networks (GAN) [25] have around great interests of many researchers due to their high performances on data classification, which makes their applications possible on biometric identification [26-27], action recognition [28-29] and object detection [30, 163], etc.

During last decade, manifold learning methods have grown explosively, which can be divided into two kinds. One kind aims to find a subspace by preserving the local geometry structure such as manifold charting (MC) [31], locally linear embedding (LLE) [32], Laplacian eigenmaps (LE) [33], local tangent space alignment (LTSA) [34], Hessian locally linear embedding (HLLE) [35], Riemannian manifold learning (RML) [36], conformal eigenmaps [165], landmark multi-dimensional scaling (LMDS) [166], locally linear coordination (LLC) [167], Gaussian mixed model(GMM) [168] and stochastic neighbor embedding (SNE) [169-170]. Dissimilar to the former, another kind tries to preserve the global characteristics of the original data in a low dimensional subspace, for example isometric mapping (ISOMAP) [37], diffusion map (DM) [38], maximum variance unfolding (MVU) [39] and Logmap [164]. Among all these manifold learning methods, LE is favored by so many researchers because it is computationally simpler and capable of offering useful results on a broader range of manifolds [172-173].

However, when applying the original LE to real-world data, some limitations such as uneven data embeddings, out-of-sample problem and small sample size problem are easily exposed. Therefore, the original LE has to be extended to avoid these problems. In addition, LE functions well for data dimensionality reduction, which can also be exploited to extract or select discriminant features for data classification and label prediction. Thus classification-oriented LE based methods have also been developed, where data similarity metric learning, discriminant learning as well as multi-manifold learning including the supervised and the unsupervised versions have been paid attentions to. Moreover, in order to extract or select features from high dimensional data more discriminatively, except orthogonality, some other constraints as uncorrelation and sparseness are also imposed on LE based methods. They are all contained in the LE based manifold learning methods. In the following, the above mentioned issues, to which the original LE has to be modified to address, will be described one by one.

1) Uneven data

Based on the assumption that the original data are densely and evenly distributed, with the optimal parameters, the low dimensional representations of the original data can be well found using LE. In real-world data, the assumption is not always the case. For example, in some situations, some data are missing, which looks like some holes in the original data. Under such circumstance, the traditional LE shows inability to well preserve the local geometry structure in the low dimensional subspace. In other words, the local relations between nodes in the original data will be greatly destroyed in the low dimensional embeddings. Thus how to handle uneven data including those with holes still needs further studying.

2) Out-of-sample problem

Between the original data and their low dimensional representations, there is no an explicit projection function in the original LE, which often results in out-of-sample problem. For any new observation, its projection in the low dimensional space cannot be easily obtained with those known embedding results because no projection function exists. In order to find its projection, the traditional LE should be carried out to all data as well as the new coming point again, thus much more computational cost will be paid especially when applying it to large scale data pattern recognition. So it demands some extensions to LE to solve out-of-sample problem.

3) Small sample size problem

In fact, the embeddings of the original data are the corresponding eigenvectors associated to a generalized eigenvalue decomposition in LE, which easily incurs irreversible ill-conditioned eigenproblems especially for small sample size problem. On this occasion, some matrix in the generalized eigenvalue decomposition will be irreversible. Eigen-decomposition cannot be used to achieve the low dimensional projections. As a substitution, it can transform the matrix to be positive-definite to avoid small sample size problem. Moreover, it is also a expectancy that some other techniques are presented to solve the problem.

4) Metric learning

In the original LE, the nearest neighbor graph is firstly constructed, where Euclidean distance between point to pint is involved in determining k nearest neighbors for any point. However, when LE is applied to data classification, the metric is not enough to mine more discriminant or non-redundant information. Thus it is looking forward to some other distance metrics to improve the performance of data classification.

5) Discriminant learning

LE is an unsupervised dimensionality reduction method and no class label information is contained. However, it has been validated that class label information makes contributions to learn some discriminant information hidden in the original data. Therefore, some supervised strategies have been introduced for supervised learning, which results in many kinds of discriminant feature extraction methods based on LE.

6) Multi-manifold learning

Provided that all the data reside on a single continuous manifold, the original LE can be adopted to explore the low dimensional manifold structure hidden in the high dimensional data. However, for data classification, it is not the case. It is often assumed that the data with the same class locate on one manifold and those with different labels will be sampled from the corresponding manifolds, which leads to an issue of multi-manifold learning. Using the original LE, the locality of all manifolds can be well quantified and preserved, which cannot make it sure to separate data with varied class labels as much as possible. Hence, it deserves to make further research on multi-manifold learning models instead of one single manifold.

7) Constraints

From the original LE, it can be found that low dimensional projections of the original data will be obtained by solving an objective function, where just a simply constraint is appended to it and no other constraints are considered. However, it is generally believed that constraints on the low dimensional embeddings are helpful for feature extraction, which will improve the performance of final data classification. Thus in order to extract more discriminant features from the high dimensional data, it is encouraged to constrain the objective function of the original LE from the viewpoint of statistics learning.

The rest of the paper is organized as follows: Section 2 simply reviews the original LE and its relations to LLE. A systematical survey on LE based methods is made according to the issues mentioned above, where their strategies, advantages or disadvantages and some comparison results are also offered. In Section 4, the research directions in future are also suggested accompanied with some conclusions.

2 Reviews on LE

LE desires to find low dimensional embeddings of the original data by preserving the similarity relations between local points which can be determined using k nearest neighbors criterion or super-ball criterion [40]. For any point, its neighborhood points will be firstly selected with Euclidean distance or Cosine distance as measurements. Then repeat the process to all the points and a nearest neighbor graph can be constructed, which is just the first step of the original LE.

In step2, the weights between any two nodes in the nearest neighbor graph will be assigned to either simply value or value of a heat kernel function. For the first mode, they are directly set to a simple value 1 or 0 as follows:

$$S_{ij} = \begin{cases} 1 & X_i \in N(X_j)\ or\ X_j \in N(X_i) \\ 0 & otherwise \end{cases} \tag{1}$$

where $N(X_j)$ and $N(X_i)$ denote the neighborhood of point X_j and the neighborhood of point X_i, respectively.

In most cases, the weights between nodes are also set to the value of a heat kernel function, which are stated below.

$$S_{ij} = \begin{cases} e^{\frac{-\|X_i - X_j\|^2}{\beta}} & X_i \in N(X_j)\ or\ X_j \in N(X_i) \\ 0 & otherwise \end{cases} \tag{2}$$

where β is a constant.

In step 3, LE aims to explore a low dimensional subspace where the local relations including the neighborhood and the weights are all well preserved. It can be approached by modeling the following objective function.

$$\begin{aligned} &\min \sum_{ij} \|Y_i - Y_j\|^2 S_{ij} \\ &s.t.\ YDY^T = nI \end{aligned} \tag{3}$$

where $D_{ii} = \sum_j S_{ij}$, Y_i and Y_j are low dimensional representations of the original points X_i and X_j, respectively.

The above objective function can be deduced to

$$\begin{aligned} &\min YLY^T \\ &s.t.\ YDY^T = nI \end{aligned} \tag{4}$$

where L is graph Laplacian and the corresponding derivation is stated below.

$$\begin{aligned} &\sum_{ij} \|Y_i - Y_j\|^2 S_{ij} = \sum_{ij} (Y_i Y_i^T - Y_i Y_j^T - Y_j Y_i^T + Y_j Y_j^T) S_{ij} \\ &= 2\sum_{ij} (Y_i Y_i^T - Y_i Y_j^T) S_{ij} = 2Y(D-S)Y^T = 2YLY^T \end{aligned} \tag{5}$$

Thus it can be easily found that the low dimensional representations, i.e. Y, are eigenvectors associated to the bottom d eighenvalues of the following generalized decomposition problem.

$$LY_i = \lambda_i DY_i \tag{6}$$

where λ_i means the corresponding generalized eigenvalue.

Compared to LLE, if the weights between nodes in the nearest graph are set to the value of the following function, LLE is just LE.

$$S_{ij} = \begin{cases} W_{ij} + W_{ij}^T - W_{ij}W_{ij}^T & X_i \in N(X_j) \ or \ X_j \in N(X_i) \\ 0 & otherwise \end{cases} \tag{7}$$

where W_{ij} expresses the least locally linear reconstruction weights obtained from the following objective function.

$$\varepsilon\left(W_{ij}\right) = \min \left\| X_i - \sum_{j=1}^{k} W_{ij} X_j \right\|^2 \tag{8}$$

Due to the fact that W_{ij} is sum-to-one, thus we can change Eqn.(8) to

$$\begin{aligned} \varepsilon\left(W_{ij}\right) &= \min \left\| \sum_{j=1}^{k} W_{ij}(X_i - X_j) \right\|^2 = \min \left\{ \sum_{j=1}^{k} W_{ij}(X_i - X_j) \bullet \sum_{t=1}^{k} W_{it}(X_i - X_t) \right\} \\ &= \min \sum_{j=1,t=1}^{k} W_{ij} W_{it} G_{jt} \end{aligned} \tag{9}$$

where G_{jt} denotes a local gram matrix and is defined as follows:

$$G_{jt} = (X_i - X_j) \bullet (X_i - X_t) \tag{10}$$

Using Lagrange multiplier, the solution of the above objective function is listed below.

$$W_{ij} = \frac{\sum_{t=1}^{k} G_{jt}^{-1}}{\sum_{m=1}^{t} \sum_{l=1}^{k} G_{lm}^{-1}} \tag{11}$$

3. Survey of LE based methods

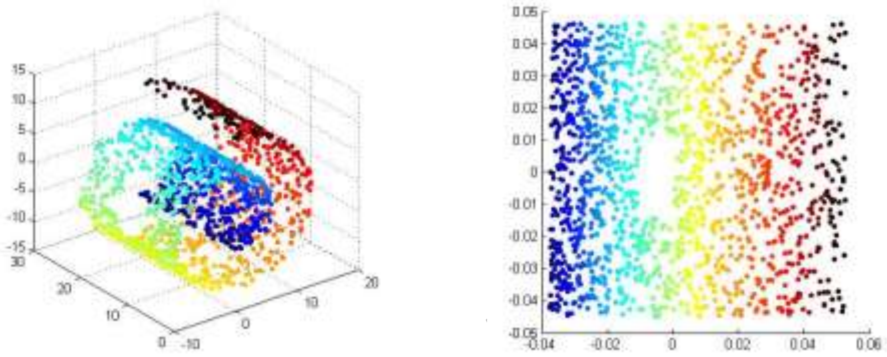
As mentioned above, several limitations existed in the original LE, which hinder its applications on data visualization, data mining and pattern classification efficiently. Thus all kinds of extensions have been made to overcome these problems including the nearest neighbors selection, unevenly data sampling, out-of-sample problem, small sample size problem, metric learning, discriminant learning, multi-manifold identification and statistics constraints. It must be pointed out that when constructing the neighborhood graph, both LE and LLE have shared the same strategies such as smoothing [174-175] and parameters setting [176-178], which were reviewed in Ref.[179-180]. So in this paper, we will not concentrate on the issue of the nearest neighbors selection. In the following, LE based dimensionality reduction methods will be systematical surveyed according to the rest issues, where they are also categorized into some more subtypes. Table.2 concludes theses issues with the further subdivisions. Moreover, the corresponding classical method is listed for each subdivision, which can also be found in Table.2.

Table 2 Categorization of extensions to LE

Categorization		Classical methods
Uneven data		HLE
Out-of-sample problem	Linear approximation	LPP
	Kernel transformation	Kernel LPP
	Tensor representation	TSA
	Incremental leaning	Ref.[49]
	Neural networks	RNN
	Extreme learning machine	Ref.[59]
Small sample size	Preprocessing	LLP
	Perturbation	Ref.[204]
	Margin	DMML
Metric learning	P2S distance	NFSE
	S2S distance	FSDML
	Manifold to manifold distance	MMD
	Manifold margin	DMML
Discriminant learning	Discirminant weights learning	LDP
	Discriminant graph learning	MFA
	Label based similarity measure	CMVM
	Mixed learning models	LPP/MMC
Multi-manifold learning	Unsupervised	UDP
	Supervised	MFA
Constraints	Orthogonal	OLPP
	Global uncorrelation	LUDP
	Local uncorrelation	CDNE
	Sparse	SPP

2.1 Extensions to LE concerning Uneven data

When using k nearest neighbors to approach the local patches in manifold, it is assumed that data are densely distributed. However, it is not always the case. Unevenly data especially with holes may be sampled. Thus for these data, Hessian LE (HLE) method was put forward [34]. HLE is a variant of LE that can deal with unevenly sampled data. In the proposed Hessian LE, the ‘curviness’ of the high-dimensional manifold, characterized by matrix H_e, will be minimized in the expected low dimensional subspace with a locally isometric constraint. The curviness of the manifold is quantified by means of the local Hessian at every data point, which is modeled in its local tangent space and invariant to differences in the positions of data points. By making eigen-decomposition to H_e, HLE will naturally explore a low dimensional subspace where the local geometry structure in the original data can be well preserved. Fig.1 shows the original data with a hole and their embeddings using HLE, by which the performance of HLE to deal with uneven data can be validated.



(a)

(b)

Fig.1: (a) the Swiss-roll data with hole; (b) 2D embeddings of the original data using HLE

2.2 Extensions to LE concerning out-of-sample problem

Out-of-sample problem often occurs to LE when nonlinear dimensionality reduction methods are employed to real-world data [41-43,50-54]. For LE, there is also no a projection function between the original data and their low dimensional embeddings, thus for any new coming point, how to find its embedding is still an open problem because the embeddings of the training data cannot be well taken full advantage of. Fortunately, some methods using different tricks have been presented to conquer out-of-sample problem.

2.2.1 linear approximation to LE

Linear approximation to LE is widely used to overcome out-of-sample problem. Although the original data are nonlinearly projected into a low dimensional subspace, it is suggested that a linear approximation can be adopted to replace the nonlinear projection, i.e. $Y_t = A^T X_t$, where A denotes the linear transformation matrix. Then a new objective function will be deduced as follows:

$$\begin{aligned} &\min A^T XLX^T A \\ &s.t.\ A^T XDX^T A = nI \end{aligned} \tag{12}$$

Compared to the original LE, the solutions can be transformed to find an optimal linear projection A ,by which the low dimensional embeddings of the existed points and the new coming one can be intuitively obtained with the linear transformation $Y = A^T X$.

Locality preserving projection (LPP) [44-45] is a classical linear version of LE, where a linear transformation is recommended between the original data and their projections. On one hand, the problem of out-of-sample is naturally avoided. On the other hand, the efficiency using LPP for data classification is also improved. Unfortunately, due to introducing the linearity compulsively, the global nonlinear geometry in data may be destroyed, which makes it failure to detect the nonlinear geometry structure and cannot well carry out geometry-aware learning.

2.3.2 Kernel extensions to LE

Due to the fact that kernel functions are introduced to make nonlinear projections to the original data, it can easily find the embedding of any new coming data using an explicit kernel function, thus out-of-sample problem will be naturally avoided. The objective function with kernel extension to the original LE is stated below:

$$\begin{aligned} &\min \alpha^T KLK^T \alpha \\ &s.t.\ \alpha^T KDK^T \alpha = nI \end{aligned} \tag{13}$$

where $K = \phi(X_i) \bullet \phi(X_j)$.

α are solutions of the objective function listed above, which can be introduced to calculate the mappings as:

$$Y = \sum_i \alpha_i \phi(X_i) \tag{14}$$

Kernel LPP (KLPP) [46-47] is a kernel extension to LPP by nonlinearly mapping the original data into a reducing kernel Hilbert space (RKHS) with some kernel functions [181]. Thus similar objective function with kernel forms will be modeled to find the corresponding optimal subspace. KLPP can improve the performance for features learning with more computational burden. Moreover, it also brings some problems such as kernel function selection and parameter setting.

2.3.3 Tensor representation to LE

As mentioned above, both the linearization and kernel versions to LE consider to represent the original data point with vectors. However, some other forms such as tensor can also be used to rewrite the original data, i.e. $X \in \Re^{1\times 2\times \ldots \times n}$. Similar to LPP, another linearization with tensor representation, i.e. $Y_i = X_i \times_1 A^1 \times_2 A^2 \times \ldots \times_n A^n$, is also introduced, thus the original objective function of LE should be transformed to the following formulation:

$$\begin{aligned} &\min \left\| X_i \times_1 A^1 \times_2 A^2 \times \ldots \times_n A^n - X_j \times_1 A^1 \times_2 A^2 \times \ldots \times_n A^n \right\|^2 W_{ij} \\ &s.t.\ f\left(A^1, A^2, \ldots, A^n\right) = d \end{aligned} \tag{15}$$

From this objective function, we can obtain the optimal tensor representation, with which any point can be easily projected.

Tensor subspace analysis (TSA) [48] aims to explore the intrinsic local geometrical structure embedded in a tensor space, where the low dimensional tensors can be obtained by making projections. Thus TSA is viewed to a tensor learning method, where any image can be treated as a second order tensor. Then LE will be applied to these data with tensor forms. TSA is much computational simple and robust to noises and outliers.

2.3.4 Incremental learning to LE

Some researchers take incremental learning to solve out-of-sample problem existed in the original LE. When carrying out incremental learning to LE, firstly, the similarity matrix for weighting adjacent relations will be updated. In other words, with a new observation x_{n+1},the similarity matrix will be changed into the scale of $(n+1)\times(n+1)$.Then the low dimensional embeddings of the original data will be changed accordingly.

$$Y_{n+1} = \left. \begin{bmatrix} & W_{(n+1)1} & \\ (Y_1, \ldots, Y_n)(& \ldots &) \\ & W_{(n+1)n} & \end{bmatrix} \middle/ \sum_{i=1}^{n} W_{(n+1)i} \right. \tag{16}$$

The above incremental learning is a differential based model [49]. There is another mode for incremental learning, i.e. sub-manifold analysis [49]. In this method, LE is firstly used to obtain a low dimensional subspace with the eigenvectors associated to some bottom eigenvalues, then map the new coming point into the subspace with the least local reconstruction trick and obtain the corresponding linear reconstruction weights, at last, the projection of the new coming point can be computed to linear combination of the linear reconstruction weights and the embeddings of its k nearest neighbors, which is stated below:

$$Y_{n+1} = \sum_{i=1}^{k} c_i Y_{N(X_i)} \tag{17}$$

where $Y_{N(X_i)}$ are the known embeddings of k nearest neighbors for new coming point X_{n+1} , c_i is the corresponding linear reconstruction weight obtained from the following objective function.

$$\begin{aligned} &\min \left\| X_{n+1} - \sum_{i=1}^{k} c_i V_i \right\|^2 \\ &s.t. \sum_{i=1}^{k} c_i = 1 \end{aligned} \tag{18}$$

where $V = [V_1, V_2, ..., V_d]$ are some eigenvectors spanned the low dimensional subspace by using LE to the original data.

There are also some other incremental learning methods [55-58], which aim to obtain the projection of any new coming point with the existed embeddings of training data. They have show their performances for out-of-sample problem. However, more computational cost must be paid when applying them.

2.3.4 Using Neural networks

As described above, neural networks can also be introduced to solve out-of-sample problem existed in LE. Some neural network based methods are more computationally efficient, for example, quickNet [61-62] , random vector feature link (RVFL) [63-65] and random neural networks (RNN) [66], all of which are within the family of randomness-based learning networks. All these methods act differently on the way to optimize parameters. Recently, deep neural networks are also employed to overcome out-of-sample problem [67].

2.3.5 Using extreme learning machine

Extreme learning machine (ELM) has the property that can randomly choose hidden nodes and analytically determine the output weights. Moreover, ELM tends to learn good generalization to new observations with extremely fast speed [183].

ELM is also introduced to obtain out-of-sample embeddings [183, 59-60]. The output for an input can be represented to

$$f_L(X) = \sum_{i=1}^{L} \beta_i h_i(X) = h(X)\beta \tag{19}$$

where β is a output weight vector, each element in $h = [h_1, ..., h_L]$ is the value of a sigmoid function.

$$h_i = \frac{1}{1 + \exp(-a_i X + b_i)} \tag{20}$$

let H be matrix

$$H = \begin{bmatrix} h_1(X_1) & \dots & h_L(X_1) \\ \dots & \dots & \dots \\ h_1(X_L) & \dots & h_L(X_L) \end{bmatrix} \tag{21}$$

Thus β can be computed as follows:

$$\beta = (\frac{I}{c} + H^T H)^{-1} H^T Y \tag{22}$$

where c is a constant.

For any new coming point X_{n+1}, its projection is computed to:

$$Y_{n+1} = \frac{\beta}{1 + \exp(-X_{n+1}A - 1^T B)} \tag{23}$$

As a conclusion, the types for solving out-of-sample problem accompanied with their representatives are listed in Table 3. In addition, both advantages and disadvantages are also summarized and displayed in Table 3.

Table 3 Categorization of out-of-sample extensions of LE

Categorization	Classical methods	Advantages	Disadvantages
Linear approximation	LPP, OLPP, UDP	Robust, high efficiency	Nonlinearity destroyed
Kernel transformation	KLPP	Robust, nonlinear	Kernel function selection, parameter setting
Tensor representation	TSA	high efficiency	
Incremental learning	Ref.[49]	High performance	Computational expensive
Neural networks	QuickNet, RVFL, RNN	Robust , nonlinear computationally efficient	Uneasy for training
Extreme learning machine	Ref.[59]	Robust, nonlinear computationally efficient	Uneasy for training

3.3 Extensions to LE concerning small sample size problem

Similar to the ways to solve small sample size problem in traditional LDA, LE based methods have made extensions from the following tricks.

3.3.1 Preprocessing

Preprocessing is often adopted by LE based method to handle small sample size problem. Due to the fact that some matrixes are non positive-definite, which will result in ill eigen-decomposition problem when modeling them to a Fisher form. Thus a preprocessing is introduced in advance by carrying out PCA to the original data to reduce their dimensions, from which these matrixes will be made to be positive-definite. In LPP and UDP, PCA is firstly applied to the original data to obtain some features with lower dimensions, on which LPP or UDP can run to mine the expected features.

3.3.2 Perturbation

Another technique to make the matrixes mentioned above positive-definite is to add a perturbation to them [204], thus LE based manifold learning methods can be exploited by constructing different Fisher models on the matrixes after perturbation. However, it also brings another problem about how to determine the parameter optimally.

3.3.3 Margin

In order to avoid small sample size problem in LE based methods, another widely used strategy is to transform the form of Fisher to difference in the corresponding objective functions, where the non positive-definite matrixes will not show much impacts on eigen-decomposition. Benefited from this way, discrimiant multi-manifold leaning (DMML) successfully avoids small sample size problem and makes contributions on multiple manifolds identification where just a single image is taken as training sample [122].

3.4 Extensions to LE concerning discriminant learning

The original LE is an unsupervised manifold learning method, where no class label information is taken into account. However, supervised information shows heavy impacts on discriminant feature extraction. Thus when constructing the nearest neighbor graph, class information can be borrowed for discriminant analysis. It is well known that a graph consists of nodes and some weighted edges between them. On one hand, class labels are recommended to weight the edges in k nearest neighbor graph, by which the discrimination will be improved. On the other hand, labels are also considered to select k nearest neighborhood points, where nodes with the same class are chosen. Under such circumstance, LE will minimize the similarity between within-class data, which helps to cluster those data. In addition, the class label information can also be exploited to measure the label similarity and dissimilarity between neighborhood points or other points, by which points will be distinguished either differently labeled or identically labeled. Meanwhile, although the original LE is an unsupervised dimensionality reduction method, some supervised tricks are also carried out by combining LE to other supervised methods, i.e. LDA, thus the disciminant features will be extracted under supervision from the original high dimensional data.

Based on the original LE, some supervised extensions to it have been made. In the following , we will successively survey LE based discriminant learning dimensionality reduction methods according to the rank as weights based supervised learning, supervised neighbors selection, label based similarity measure and mixed learning models.

3.4.1 weights based supervised learning

In the original LE, the similarity between any two neighborhood points is set to either simply value or a heat kernel value. Thus in order to improve the discrimination of these weights, some supervised LE based methods are presented to define new weights between neighborhood points using both data label relation and local information, where nearest neighborhood points for any point can be classified into the inter-class nearest neighbors and the intra-class nearest ones. The weights between them are set differently. Local discriminant projection (LDP) sets the weights to [68]:

$$W_{ij} = \begin{cases} \exp(-\frac{d^2(X_i, X_j)}{\beta})\left(1+\exp(-\frac{d^2(X_i, X_j)}{\beta})\right) & \text{If both } X_i \text{ and } X_j \text{ are } k \text{ nearest neighbors each other and have the same label;} \\ \exp(-\frac{d^2(X_i, X_j)}{\beta})\left(1-\exp(-\frac{d^2(X_i, X_j)}{\beta})\right) & \text{If both } X_i \text{ and } X_j \text{ are } k \text{ nearest neighbors each other and have different labels;} \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

In orthogonal discriminant projection(ODP), another weight function is offered [69].

$$W_{ij} = \begin{cases} \exp(-\frac{d^2(X_i, X_j)}{\beta}) & \text{If both } X_i \text{ and } X_j \text{ are } k \text{ nearest neighbors each other and have the same label;} \\ \exp(-\frac{d^2(X_i, X_j)}{\beta})\left(1-\exp(-\frac{d^2(X_i, X_j)}{\beta})\right) & \text{If both } X_i \text{ and } X_j \text{ are } k \text{ nearest neighbors each other and have different labels;} \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

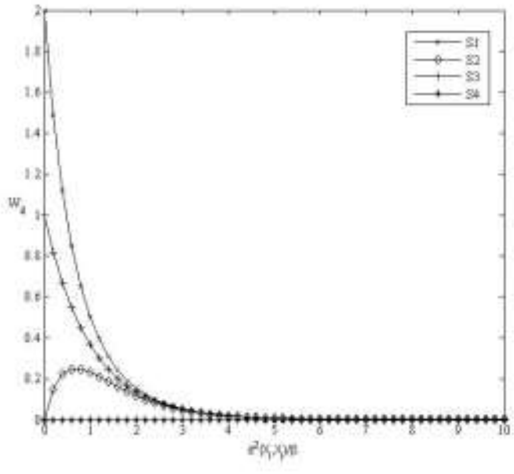


Fig.2 Typical plot of W_{ij} as a function of $d^2(X_i, X_j)/\beta$ in ODP(S2, S3, S4) and LDP(S1, S2, S4)

The above figure shows the curves of weight functions regard to Euclidean distance between point X_i and X_j ,where some discriminant information may be implied.

In Table 4, we report the performance comparison on AR[185] and Yale[184] face data sets, where LPP, LDP and ODP are involved. For LPP, it is a unsupervised method for weight setting. To the contrary, both LDP and ODP assign the weights between nodes by taking class information and local relations into account. From this viewpoint, the performances of LDP and ODP on two face data sets are superior to those of LPP, which can also be found from Table.4.

Table 4 Performance comparison on AR and Yale face data with LPP, LDP and ODP

Data sets	LPP	LDP	ODP
AR	85.33%	89.42%	91.33%
Yale	84%	87.33%	86%

In another distance-weighted LPP method [70], the weights between any two nearest neighbors are also adjusted by considering the relations of the original facial expression data, where a metric P represented the distance between expression kinds is constructed to modify the original Euclidean distance. It is stated below:

$$D_{ij}^{*} = D_{ij} + \frac{P\left(e_i, e_j\right)}{\max\left(P\left(e_i, e_j\right)\right)} \max\left(D_{ij}\right) \tag{26}$$

where D_{ij} denotes Euclidean distance between points X_i and X_j .

With the above defined weights, spectrum mapping can be employed to find low dimensional embeddings of the original data.

3.4.2 neighborhood based supervised learning

Recently, more and more supervised tricks are used for supervised construction of neighborhood. Although combined to LLE, these tricks are also meaningful to LE. Thus they will be described in the following. In the original LLE, the manifold local geometry is usually explored using k nearest neighbors graph, where Euclidean distance is employed. In most cases, some points with different labels may also have shorter Euclidean distances than those with the same class when constructing the nearest neighbors graph, which leads to wrong neighborhoods for classification that contain neighbors with different classes. To address the problem, a method to adjust neighborhood weights is advanced by taking the class information into account, where the distance between any two points belonging to different classes is defined to be relatively larger than their Euclidean distance while those distances between two points with the same label are preserved. The work was first presented by de Ridder et al. [71], where only the Euclidean distances between points belonging to different classes are simply enlarged by adding a constant. Instead of enlarging the between-class distances, Wen et al. utilized a nonlinear function to shorten the within-class distances [72], which shows similar impacts on recognition performance. These methods just either enlarge between-class distances or shrink within-class distances. Thus Zhang brought forward an enhanced supervised LLE model by reducing within-class distance and expanding between-class distance simultaneously [73]. Combined to class information, these methods endeavor to increase the accuracy of LLE by adjusting the distances between neighborhood points rather than by selecting the neighborhoods points. So Hui et al. [74] and Zhao et al. [75] imposed a strict constraint on construction of k nearest neighbors graph that only points with the same class can be considered to be neighbors. Nevertheless, when points are not sampled densely, the neighborhoods points determined by the method mentioned above will be not enough to explore the manifold geometry structure. Thus, Han et al. proposed a method to make a supplement [76]. According to the ascending Euclidean distances, the same class samples are firstly predefined as neighborhood points, and then the remaining neighbors are searched from points with different classes. Later, Zhang and Zhao introduced the probability-based distance that can enlarge the Euclidean distance for labeled and unlabelled points [77-78]. However, this modified version of LLE just takes advantage of class information to adjust the distances between points or to select the neighborhood points in k nearest neighbor graph, where more parameters are introduced with the augment of the application difficulty.

In most cases, an intra-class graph and an inter-class graph will be constructed by using both class information and local information [79-84]. In the intra-class graph, for any point, its neighborhood points must be sampled from those points with the same class, where the neighbors are points with the sorted bottom Euclidean distances to it. In other words, the intra-class neighborhood consists of points labeled identically. Similarly, for any point in the inter-class graph, its neighborhood points also have the ranked bottom Euclidean distances to it, however, the biggest difference is that the class labels of those inter-class neighbors must be various to it. Specially, those inter-class neighbors should have the same class label.

3.4.3 Label based similarity measure

In the original LE, the similarity between any two nodes in the k nearest graph can be set to some values. As described in subsection 2.4.1, class information is also taken to set the weights discriminatively. The weights setting as these two modes cannot make a significant distinction between the intra-class data and the inter-class data. Thus a similarity matrix and a dissimilarity matrix using the label information are respectively defined as follows:

$$S_{ij}=\begin{cases}1 & X_i \text{ and } X_j \text{ are with the same label}\\ 0 & \text{otherwise}\end{cases} \tag{27}$$

$$nonS_{ij}=\begin{cases}1 & \mathrm{X_i} \text{ and } \mathrm{X_j} \text{are varied labelled}\\ 0 & \text{otherwise}\end{cases} \tag{28}$$

Using these label based similarity and dissimilarity, some extensions to LE are also made for discriminant learning. Both maximum variance mapping (MVP) [85] and constrained maximum variance mapping (CMVM) [86] introduce the dissimilarity matrix to learn the variance between the inter-class data, then a low dimensional subspace will be explored by maximizing the variance and by preserving the locality. In feature space distance metric learning (FSDML) [87], the similarity and the dissimilarity are all absorbed to model an intra-class graph and an inter-class graph for data classification, respectively.

3.4.4 Mixed learning models

Generalized Fisher framework (GFF) is a hybrid supervised model where scatters both in LDA and in UDP are all considered [88]. In LDA, two scatters, i.e. the between-class scatter and the within-class scatter, are formulated based on the data class information to represent the between-class data apartness and the within-class data compactness.

$$S_B=\sum_{i=1}^{c} n_i(m_i-m)(m_i-m)^T \tag{29}$$

$$S_W=\sum_{i=1}^{c}(\sum_{j=1}^{n_i}(X_i^j-m_i)(X_i^j-m_i)^T) \tag{30}$$

Instead of data labels, local information and non-local information are made full advantage of to model a local scatter and a non-local scatter, by which the data clustering and separability can be approached individually.

$$S_L=\frac{1}{2}\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}s_{ij}\left(X_i-X_j\right)\left(X_i-X_j\right)^T \tag{31}$$

$$S_N=\frac{1}{2}\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(1-s_{ij})\left(X_i-X_j\right)\left(X_i-X_j\right)^T \tag{32}$$

It is the fact that both scatters S_W and S_L are often taken as measurements for data similarity, thus in GFF, a similarity scatter S_s is defined to linear combination of S_W and S_L :

$$S_S = \beta S_W + (1-\beta) S_L \tag{33}$$

With the same manner to the similarity scatter , a dissimilarity scatter can also be formulated to:

$$S_D = \alpha S_B + (1-\alpha) S_N \tag{34}$$

GFF aims to find a subspace where the similarity scatter will be minimized and the dissimilarity scatter will be maximized, simultaneously. From above scatters, it can be found that the original LDA and UDP are mixed for subspace location. Moreover, GFF can also be LDA or UDP by setting the parameters accordingly.

Locality preserving projection based on maximum marginal criterion (LPP/MMC) [89] is another LE based method with LPP and LDA. LPP/MMC seeks to maximize the difference, rather than the ratio, between the locality preserving between-class scatter and locality preserving within-class scatter. DLPP/MMC is theoretically elegant and can derive its discriminant vectors from both the range of the locality preserving between-class scatter and the range space of locality preserving within-class scatter.

Mixing the local version of MMC and LPP, local maximal margin discriminant embedding (LMMDE) [90] constructs an objective function which combines the local scatter and local margin weighted with some coefficients, by which a subspace will be explored with the maximum local margin and the minimum locality, simultaneously.

This technique is also adopted by some other classical linear methods including LDA, where different local versions of the between-class scatter and the within-class scatter are reasoned to form the corresponding Fisher criterions. Using traditional LDA to all the local patches on manifold data, local Fisher discriminant analysis (LFDA) [79], non-parametric discriminant analysis (NDA) [80] and reconstructive discriminant analysis (RDA) [81-82] maximize the trace ratio of the local inter-class graph scatter to the local intra-class graph scatter to find an optimal subspace. However, in some cases, small sample size problem occurs to them because their objective functions are of trace ratio formula. In order to overcome the problem, different local margin criterions are established. Locality sensitive discriminant analysis (LSDA) [83] deduces a local margin to difference between the local inter-class graph scatter to the local intra-class graph scatter. Another manifold learning based method called local discriminant embedding (LDE) [84] constructs two graphs using data neighborhood and class relations, where two graph Laplacians are taken to model a local manifold margin to find the low-dimensional embeddings.

3.5 Extensions to LE concerning metric learning

In the original manifold learning method, Euclidean distance between point to point is always used, based on which some promising results on synthetic and real-world data have been possessed. However, in some methods, some other distance metrics are also recommended to quantify the distance between point to point, point to subspace, subspace to subspace or manifold to manifold. So in the following, we categorize these distance metrics into point to space distance (P2S), space to space (S2S) distance and manifold to manifold (M2M) distance, which will be discussed step by step.

3.5.1 Point to feature space distance

When computing P2P distance metric, only information of two single points is involved in. However, in traditional manifold learning method as LLE, the reconstruction error can be expressed to Euclidean distance between any point to its weighted mean of k nearest neighbors, where local relations including neighborhood points and the least reconstruction weights are all contained. That is to say, a synthetic point represented by the weighted mean of k nearest will take full advantage of local geometry in the original neighborhood, where the information of more points than just one point is exploited. Motivated by this idea, other distance metrics that measure Euclidean distance between a point to a set composed of some points are proposed, which is named point to space (P2S) distance. If the number of points consisting the set is 1, it will be P2P distance. If the number of points consisting the set is 2, it will be point to line (P2L) distance. Similarly, set the number to 3, a point to plane (P2PL) distance metric can be obtained. Then extend these special cases into a general form, a P2S distance will be derived [91]. In the following, Table.5 shows different P2S distance metrics with varied number of point set.

Table.5 Different Distance metrics with varied k

k	P2S
1	Point to point distance metric
2	Point to line distance metric
3	Point to plane distance metric
other	Point to space distance metric

It has been validated that any point can be mapped into a space composed of some data points and the distance between P2S will be Euclidean distance between the point to its projection in the space, which can be expressed to

$$D_{P2S} = \left\| X_i - f^{(p)}(X_i) \right\|_2 \tag{35}$$

where $f^{(p)}(X_i)$ denotes the projection of point X_i to space consisting of points $X_j\,(j = 1, 2, ..., k)$.

The projection can also be computed as

$$f^{(p)}(X_i) = \sum_{j=1}^{k} l_j X_j \tag{36}$$

$$\sum_{j=1}^{k} l_j = 1 \tag{37}$$

Moreover, $l_j\,(j = 1, 2, ..., k)$ is the corresponding mapping coefficient [91].

Recently, the nearest linear combination (NLC) approaches [92-94] including the nearest feature line (NFL) distance metric and the nearest feature plane (NFP) distance metric have been validated to explore information contained in more than one feature point within the same class. In some methods, the space is composed of some points with the same label which can be used to characterize the within-class data. Thus the space is also named feature space. In k nearest neighbor classifier, P2L or P2S distance metric show their superiority to P2P distance metric for recognition accuracy [95-96]. Moreover, either P2L or P2S distance metric has also been introduced to extract features from high dimensional data. Uncorrelated disciminant nearest feature line analysis (UDNFLA) [97] uses P2L distance metric to seek a low dimensional feature subspace under a uncorrelated constraint such that all the distances based on the intra-class feature line will be minimized and the total distances based on the inter-class feature line will be maximized, simultaneously. Nearest feature space embedding (NFSE) [91] defines both a local between-class scatter and a local within-class scatter using P2S distance metric, by which an objective function with ratio form can be constructed to find a low dimensional subspace.

3.5.2 S2S distance metric

As mentioned in the above subsection, more information will be contained in a space consisting of more points than just one point. Thus in order to fully utilize the geometry of the original data, P2S distance is extended to S2S distance, which is defined to:

$$D_{S2S} = \left\| f^{(p)}(X_i) - f^{(p)}(X_j) \right\| \tag{38}$$

FSDML is a supervised method where S2S distance and label based similarity are all absorbed to construct an intra-class graph and an inert-class graph, from which two graph Laplacians can be achieved for feature extraction. Moreover, FSDML is also robust to noises and outliers. But its computational burden in FSDML should be high, most of which is paid for S2S distance metric.

In FSDML, feature space is composed of some points and no more constraints are imposed on them. However, if these points are orthogonal each other, they will span an orthogonal subspace. There are some metrics which can scale the distance between two orthogonal subspace. One is principal angle, which is defined as follows:

$$\cos(\theta) = \max_{X_i \in S_1} \max_{X_j \in S_2} X_i^T X_j \tag{39}$$

where X_i and X_j are the corresponding points in orthogonal subspace S_1 and S_2, respectively.

Principal angle can also be viewed as canonical correlation kernel [98-100] and is deduced to:

$$\cos(\theta_k) = \lambda_k \tag{40}$$

where λ_k is the singular value for $P_1^T P_2$, it is also treated as canonical correlation, moreover, P_1 and P_2 are orthogonal bases composed of S_1 and S_2, respectively.

Due to the fact that any point on a Grassmannian manifold will be an orthogonal subspace [101-103]. Thus the distance between two orthogonal subspace will be transformed to geodesic distance on a Grassmannian manifold.

Geodesic flow kernel (GFK) [186] is a new kernel based method, which makes full use of the Grassmannian manifold geometry structure information by integrating an infinite number of subspaces that characterize changes in geometric and statistical properties on Grassmannian manifold. GFK shows performance on computationally advantageous, automatically inferring important algorithmic parameters without requiring extensive cross validation.

3.5.3 manifold to manifold(M2M) distance

For point set on a manifold, their distances will be geodesic distances along the manifold. If any two sets are respective on different manifolds, M2M distance will be recommended as measurement. There are also some M2M distance metrics used in LE based method.

In non-parameter discriminant multi-manifold learning (NDML) [104-105], M2M distance metric is defined to the sum of distances between any point to mean of its k inter-class nearest neighbors.

$$D = \left\| X_i - \bar{X}_i \right\| \tag{41}$$

where D denotes manifold distance and $\bar{X}_i$ is the weighted mean of k inter-class nearest neighbors for any point x_i .

In fact, the M2M distance metric defined in NDML is similar to P2S distance metric. The difference between them is that the weighted mean is considered to a point. Meanwhile, in P2S distance, the weights are not equal, however, the weights are identical each other in NDML.

M2M distance is also newly defined in manifold to manifold distance (MMD) method [106-107]. Based on canonical correlation, firstly a variation distance is obtained:

$$D_v = r \Big/ \sum_{k=1}^{r} \lambda_k \tag{42}$$

And then, another example distance is also formulated

$$D_E = \frac{\|X_i\| \|X_j\|}{X_i^T X_j} \tag{43}$$

On the basis of these two distances, M2M distance metric can be obtained.

$$D = \alpha D_v + (1-\alpha) D_E \tag{44}$$

For some special cases, M2M distance can be changed into geodesic distance on a Riemannian manifold where data are symmetric positive definite. Thus logarithm Euclidean distance (LED) [108-111], affine invariance Riemann metric (AIRM) [112-114] and projection metric (PM) [115-116] can be used as measurements, which help to explore the original data geometry.

Geometry-aware methods [110-111] attempt to make discriminant classification while preserving as much as possible the geometry of the original data points, which can be measured by geodesic distance on Riemann manifold. In Ref. [111], a closed form solution for geometry-aware is approached for substantial dimensionality reduction without affecting the classification accuracy.

3.5.4 manifold margin

LDA is a traditional linear dimensionality reduction method, which has been widely used for high dimensional data feature extraction. However, there are also some shortcomings including small sample size problem in LDA. The problem occurs to LDA when data number is smaller than data dimensions, which results in the irreversible within-class scatter. In order to overcome the problem, some techniques have been proposed [117-121]. Among them, a straight-forward method is to change the ration form of the original LDA into a difference form, which is defined to margin between various classes. Thus small sample size problem can be naturally avoided. Unfortunately, the similar dilemma also emerges in LE based methods. Motivated by the idea in LDA to solve small sample size problem with margin, some manifold margin criterions are also presented. In some cases, the manifold margins are globally formulated. Mostly, the proposed manifold margins in LE based method are reasoned locally. Beginning with a globally defined manifold margin, then the local version to it will also be surveyed.

Discriminant multiple manifold learning (DMML) [122] is a LE based dimensionality reduction method by maximizing a global manifold margin. Just like the margin in maximum margin criterion (MMC)[123], which is defined to difference of the between-class scatter to the within-class scatter, DMML also deduces a manifold margin directly. In UDP, the local scatter and non-local scatter are constructed to measure high order statistics of data within one manifold and data from different manifold, respectively. Thus manifold margin in DMML, is directly set to difference between non-local scatter to local scatter as follows:

$$S_M = S_N - S_L \qquad (45)$$

However, in some other LE based manifold learning methods, manifold margin, by which the apartness between manifolds can be quantified, is not directly formulated to difference of two global scatters of the total points. To the contrary, just some marginal points on manifolds are considered to construct the manifold margin locally. In marginal Fisher analysis(MFA) [124], the manifold margin is locally expressed by some marginal points distributed on varied labeled manifolds. Because just part of marginal points are used for modeling the manifold margin, the computational cost will be paid less than that of globally defined ones. But it is also lead to another problem that how to find the marginal points. The strategy in support vector machine (SVM) [125-127] can be introduced to search manifold marginal points. It is often implemented by using k nearest neighbor criterion to search them from different labeling manifold data. Thus the sum of all the minimum distances between any two marginal points can be taken as the manifold margin.

In Table.6, some performances by using NFSE, FSDML, MMD and DMML on benchmark data sets such as FERET[187], Yale and CMU PIE[188] are reported. Among these comparison methods, P2S distance metric is used in NFSE and S2S distance metric is introduced in FSDML, respectively. As to MMD and DMML, M2M distance metric and manifold margin are modeled, respectively. Thus the results in Table 6 can also offer some comparisons for these four distance metrics. From Table.6, it can find that M2M distance shows their outperformance compared to P2S, S2S and manifold margin on these three face data sets.

Table 6 Performance comparison with NFSE, FSDML, MMD and DMML

Data sets	NFSE	FSDML	MMD	DMML
FFERET	79.88%	82.52%	84.32%	80.23%
Yale	90.4%	94.53%	96.47%	90.67%
CMU PIE	94.19%	94.99%	97.12%	91.94%

3.6 Extensions to LE concerning multi-manifold learning

It is assumed that data located on a single manifold are labeled with the same class and data belonging to different manifolds will be accordingly varied labeled. For examples, one person's face images are considered on one manifold and other person face images will reside on the corresponding manifolds. Therefore, face recognition problem will be intuitively transferred to different manifolds identification. In order to achieve optimal results for classification, the explored embeddings related to the corresponding manifolds should be located as far as possible in a low dimensional subspace, which results in an issue of "classification-oriented multi-manifold learning" [146]. The problem cannot be solved by some current manifold learning based algorithms and their simply supervised extensions because they all just pay attentions to "locality" without considering "non-locality" or variances among manifolds. Consequently, multi-manifold discriminant learning methods with diverse multiple manifold metrics have to be brought forward [147-154], which can be partitioned into unsupervised models and supervised ones.

Table 7 Categorization of extensions to LE based multi-manifold learning

Categorization		Classical method
Unsupervised multi-manifold learning		UDP
Supervised multi-manifold learning	Dicriminant graph based	MFA
	Manifold metric based	MMD

3.6.1 Unsupervised multi-manifold learning

When dealing with multiply manifold learning, in some methods, class information is not made use of. The manifold similarity is just measured according to some information unrelated to class labels between the original data. On behalf of unsupervised multi-manifold learning method, UDP use the non-local relations between any two points to measure the dissimilarity of different manifolds. UDP can also be treated as a simply versions of LPP under the condition that the local density is uniform [155]. In addition, some unsupervised versions to UDP have also been made. By projecting the original data into a kernel space, Kernel UDP constructs a model to find a subspace with the maximum kernel non-local scatter and the minimum kernel local scatter[156]. As the transformation from LPP to OLPP, an orthogonal constraint is also imposed to UDP, which is named orthogonal UDP(OUDP) and is presented to explore an orthogonal subspace[157]. In order to fast its speed when applying UDP, 2DUDP is also proposed [158-159].

3.6.2 Supervised multi-manifold learning

Due to introducing the class information, k nearest neighbor graph, which is necessarily modeled to detect the local relations of the original data in all the manifold leaning methods, can be constructed under the supervision of data labels. Thus both an inter-class graph and an intra-class graph will be cut from the original one, which characterize the data sampled from varied labeled manifolds and data located on one manifold, respectively. In such situations, multi-manifold learning will be viewed as discriminant graph learning. Namely, by using spectrum mapping to both graphs, multi-manifold learning will be implemented, which poses an issue as discriminant graph based multi-manifold learning. In addition, based on the assumption of data multi-manifold distribution, the similarity metric of them becomes the key for multi-manifold identification, hence another kind of multi-manifold learning can be called metric learning based one, which desires to maximum the distance between manifolds and to minimize the locality of them.

1) Discriminant graph based methods

There are many discriminant graph based multi-manifold learning methods [190-197]. As a representative, MFA shows its capability of discirminant learning. However, in MFA, not only local geometry information but also data labels are all considered to construct both an intrinsic graph and a penalty graph, which is composed of the local intra-class data and the local marginal inter-class data, respectively. In the intrinsic graph, its scatter can be quantified using graph Laplacian just like the local scatter defined in UDP. Conversely, the penalty graph consists of marginal points sampled from those with k nearest distances. Thus in the penalty graph, the weight matrix and the corresponding scatter can also be formed. By applying spectrum embeddings to both the intrinsic graph scatter and the penalty graph scatter, a linear subspace will be explored for data classification by MFA.

All the discriminant graph based multi-manifold learning methods have shared a lots in common. Firstly, local relations and label information are all considered when constructing both graphs. Secondly, two scatters related to different graphs can be used to characterize the compactness and apartness of manifolds either by spectrum mapping or by other tricks. Thirdly, a Fisher framework can be founded to search a low dimensional subspace for discrimnant feature extraction and final classification. This kind multi-manifold learning methods have been validated efficient for discriminant learning such as MFA and its extensions. However, due to its Fisher form in the corresponding objective function, small sample size problem often appears to them, which hampers their applications on some real-world data sets.

2) Multi-manifold metric learning

Metric learning based multi-manifold learning pays more attentions to manifold similarity, by which the apartness between manifolds can be measured and quantified. As mentioned above, P2P, P2S, S2S, M2M, manifold margin and geodesic distances can be used to measure dissimilarity existing in different manifolds. Moreover, some other similarity metrics are also modeled, based on which multi-manifold learning dimensionality reduction or feature extraction methods will be presented [198-203]. From all these metric learning based approaches, some characteristics can be concluded. The similarity metrics between manifolds play critical roles in them because they can be introduced to measure the variance of differently distributed manifold data. Thus manifolds can be separated as far as possible by maximizing the similarity based matrixes. In addition, manifold locality also shows some contributions to multi-manifold discriminant learning, which has been validated by the unsupervised manifold leaning method, i.e. LPP. Thus in metric learning based multi-manifold methods, not only the similarity matrixes but also manifold locality are all involved in modeling different objective functions for discirminant subspace location.

3.7 Extensions to LE concerning constraints

By using LE to the original data, a subspace will be explored which is spanned by the eigenvectors related to Laplacian matrix. Whether or not the desired subspace is orthogonal is not concerned. However, some constraints are attached to LE or LPP to extract features with more performance. For example, an orthogonal constraint is appended to LPP to obtain discirminant features with orthogality, uncorrelation is also introduced to reduce redundancy in the extracted features. Moreover, taken as a constraint, L1-norm is also used in LE based methods to explore the sparse features from the original data. Purposed by these ideas, the original LE has been extended with various constraint versions. In the following, they will be surveyed from the orthogonal, the globally uncorrelated, the locally uncorrelated to the sparse.

3.7.1 orthogonal constraint

The representative of LE based methods with orthogonal constraint is orthogonal LPP [128], which was proposed by Cai et al. in 2006. OLPP is an extension to LPP, where some constraints between projection vectors are modeled to the objective function of LPP, i.e.

$$A_d^T A_1^T = A_d^T A_2^T = ... = A_d^T A_{d-1}^T \tag{46}$$

Then the objective function of OLPP can be formulated as:

$$\begin{aligned} &\min \frac{A^T S_L A}{A^T XDX^T A} \\ &s.t.\ A_d^T A_1^T = A_d^T A_2^T = ... = A_d^T A_{d-1}^T \end{aligned} \tag{47}$$

OLPP is superior to LPP because the features extracted by OLPP have additional orthogonal property compared to LPP. However, it is the orthogonal constraint that OLPP is also of more complexity than the original LPP.

3.7.2 globally uncorrelated constraint

Except the orthogonal constraint on the subspace, other constraint such as uncorrelation is also put forward. Recently, the statistical uncorrelated characteristic between vectors is also concentrated by some researchers [129-134], the reason lies in that the uncorrelated features extracted from the original data contain the minimum redundancy compared to those orthogonal ones from the viewpoint of statistics [133]. The vectors uncorrelation has the following definition:

$$A_i^T \Gamma A_j^T = 0 (i \neq j) \tag{48}$$

where Γ is a matrix.

Uncorrelation constraints can be divided into the globally uncorrelated and the locally uncorrelated, which just depends on matrix Γ. If Γ is firstly globally modeled, the proposed constraint on it will be global uncorrelation. On the contrary, If Γ is defined by locality learning, no matter what final matrix can be deviated from, i.e. the global or the local, it will be a locally uncorrelated constraint.

In some modifications to LE, both constraints as orthogonal and global uncorrelated are all involved in low dimensional subspace exploring. Uncorrelated discrimiannt locality preserving projection

(UDLPP) designs a globally uncorrelated constraint on the total scatter, which is just the covariance of all the original data [135]. Moreover, it is also extended to a reproducing kernel Hilbert space, thus the kernel version of the global constraint is formulated.

Uncorrelated discriminant nearest feature line analysis (UDNFLA) is also an example for globally uncorrelated uncorrelation [136]. Both UDLPP and UDNFLA are uncorrelated about a matrix with the same formulation, i.e. the total scatter. However, in UDNFLA, the total scatter is defined based on P2L distance metric rather than P2P distance metric.

3.7.3 locally uncorrelated constraint

Since data geometry can be explored by locality learning, especially in manifold learning based methods, it will be vital to extract the locally statistical uncorrelated discriminant information for further classification. Thus locally statistical uncorrelated criterions are worth further demonstration. The locally uncorrelation has the following expression.

$$\int a_i^T (X_i - Local(\bar{X}_i))(X_i - Local(\bar{X}_i))^T a_j dX_i = 0 \tag{49}$$

where $Local(\bar{X}_i)$ is defined to local mean of k nearest neighbors of point X_i .

Based on the above definition, local uncorrelated discriminant transformation(LUDT)[137] uses mean of the weighted nearest neighbors to computed $Local(\bar{X}_i)$, thus an uncorrelation will be locally constructed. At last, it can be turned into vector uncorrelation about a globally defined matrix. However, the constraint in LUDT is also viewed to be a local one.

Note that when computing any point’s projection in its feature space, the projection can be expressed to the weighted mean of points composed of its nearest feature space. Moreover, the sum of these weights are one. Namely, the projection can also be taken as the local mean. Thus another locally uncorrealtion will be formulated, where P2S distance metric is contained[189].

In LUDT, the local mean can be easily determined. But in some cases, it is a challenge to model a local mean using sparse samples with high dimensions. So in local uncorrelated discriminant projection (LUDP) [138], the locally uncorrelation is not defined using local mean but using the local relation based similarity to weigh a scatter.

Table 8 shows performance comparisons using methods with orthogonal, globally and locally uncorrelated constraints. Among all these comparison methods, on one hand, LDUP and Local uncorrelated subspace learning (LUSL) [189] are local uncorrelation constrained. Differing to them, UDLPP is just constrained on a globally uncorrelation. NFSE is with orthogonal constraint. On the other hand, either UDLPP or LDUP uses P2P distance as measurement. However, P2S distance metric is adopted in both NFSE and LUSL.

Table 8 Mean performances(%) with standard deviations using UDLPP, LUDP, NFSE, CDNE and LUSL on FERET face data

Methods	3Trains	4Trains	5Trains
UDLPP	78.64 ± 1.52	80.76 ± 0.95	82.37 ± 1.21
LDUP	79.21 ± 1.83	81.37 ± 1.94	83.28 ± 0.96
NFSE	75.67 ± 1.83	79.88 ± 2.12	82.87 ± 1.32
LUSL	81.56 ± 1.13	83.75 ± 0.78	86.12 ± 0.56

From the comparison results, it can be concluded that method with local uncorrelation is superior to those with global and orthogonal uncorrelation. Moreover, it can also be found that the performances of methods with P2S distance metric are better than those with P2P distance metric.

3.7.4 Sparse constraint

In the past few years, sparse representations of signals have received a great deal of attentions[139-142], which can be solved by sparse representation for searching the most compact representation in terms of sparse linear combination of an over-complete dictionary. Compared to methods based on orthogonal constraint, sparse representation usually offers better performance with its capacity for efficient signal reconstruction. In order to obtain the sparse representation to a signal, the following formulation will be constructed.

$$\begin{aligned}&\min \|S_i\|_1\\ &s.t.\ X_i = XS_i\end{aligned} \tag{50}$$

where S_i denotes the sparse representation coefficients.

From the above function, a sparse matrix S can be obtained, which is also used to measure the similarity between any two points. Thus based on spare matrix S , a sparse based local scatter will be reasoned. With a linear combination of the proposed sparse based local scatter and a modified maximum margin, an objective function can be formed in sparse locality preserving discriminative projections (SLPDP) [143], by which an optimal subspace is found. From SLPDP, it can also conclude that the similarity between any two points is sparsely measured. Moreover, the sparse representation relations are well preserved in methods such as sparsity preserving projection (SPP)[144] and discriminant sparse neighborhood preserving embedding (DSNPE) [145].

4 Conclusions

It is well known that Laplacian eigenmaps is a nonlinear manifold learning method, which can be used for dimensionality reduction or feature extraction. Due to its efficiency on mining the local geometry from the original data, LE has been widely applied for high dimensional data analysis including face recognition[3, 213], palmprint recognition[13, 207], image processing[2, 211] and tumor gene expressive data prediction[209-210]. However, LE also exists some limitations such as uneven data projection, out-of-sample problem and small sample size problem, which hinder its applications on real-world data. When the original LE is exploited for data classification, some attempts are also made to improve its capability of discriminant learning including metric learning, supervised learning, multi-manifold learning and constraint models. Thus in this paper, we have surveyed the research progress on LE based methods both from the issues on overcoming the problems in LE and issues on discriminant learning. Moreover, some comparisons are also made to evaluate the performances of methods motivated by different purposes. On the basis of these works, some future directions on LE based methods can be concluded as follows:

1) LE based similarity metric learning in deep domain adaptation. Recently, deep domain adaptation becomes a hot topic in machine learning especially in deep learning, where the data similarity between source and target must be measured as a loss function to train deep neural networks. Thus the metric learning proposed in LE based methods can also be introduced to weigh domain shifts between source data and target data in deep domain adaptation.

2) LE based graph theory for information entropy. In real-world data, the probability distributions of them are always unknown, however, they are also very important for further learning. Thus all kind of statistical methods are recommended to solve the problem, among which graph model is one widely used trick. For example, by random walking graph, von Neumann entropy can be approximated by normalized graph Laplacian spectrum. Thus LE based graph theory will be expected to function more in information theory.

3)Geometry-aware multi-manifold learning. Recently, geometry-ware dimensionality reduction methods have been paid more attentions. In these geometry-ware methods, the geometry existing in the original data is further explored and well preserved in the low dimensional subspace. Moreover, the traditional multi-manifold similarity can be transferred to geodesic distance on a Riemann manifold, by which multi-manifold can be learning discriminatively. Thus it will be desired to develop LE based geometry-ware dimensionality reduction methods.

Acknowledgments

This work was supported in part by the grants of Natural Science Foundation of China (61572381, 61273303 and 61602349) and China Post-doctoral Science Foundation (2016M601646)

References

1. E. Roy Davies. Machine Vision: Theory, Algorithms, Practicalities. Morgan Kaufmann, 2005
2. Z.L.Sun, D.S.Huang, Y.M. Cheung, J. Liu, and G.B. Huang, “Using FCMC, FVS and PCA techniques for feature extraction of multispectral images,” IEEE Geoscience and Remote Sensing Letters, vol.2, no.2, pp.108-112, 2005.
3. J.X. Mi, D.S.Huang, B. Wang, and X. Zhu, “The nearest-farthest subspace classification for face recognition,” Neurocomputing, vol.113, pp.241-250, 2013
4. B. Li, C.H. Zheng, and D.S.Huang, “Locally linear discriminant embedding: An efficient method for face recognition,” Pattern Recognition, vol.41, no.12, pp. 3813-3821, 2008.
5. H. Zhang, X. Gao, P. Wu, X. Xin, A cross-media distance metric learning framework based on multi-view correlation mining and matching, World Wide Web, 19(2)(2016)181-197.
6. H. Zhang, P. Wu, A. Beck, Z.J. Zhang, X.Y. Gao, Adaptive incremental learning of image semantics with application to social robot, Neurocomputing, 173(2016) 93-101.
7. M. Liu, L. Zhang , H. Hu , L. Nie , J. Dai, A classification model for semantic entailment recognition with feature combination, Neurocomputing, 208(2016) 127-135
8. M. Liu, L. Jiang, and H. Hu. Automatic extraction and visualization of semantic relations between medical entities from medicine instructions. Multimedia Tools and Application. 76(8) (2017) 10555-10573.
9. D.S.Huang, C.H. Zheng, Independent component analysis based penalized discriminant method for tumor classification using gene expression data, Bioinformatics, 22(15)(2006) 1855-1862
10. B. Li, B.B.Tian, X.L.Zhang, X.P. Zhang, Locally Linear Representation Fisher Criterion Based Tumor Gene Expressive Data Classification, Computers in Biology and Medicine,53(10)(2014) 48-54
11. I.T. Jolliffe. Principal component analysis, second ed., Springer-Verlag, Berlin, 2002.
12. A.M. Martinez, A.C. Kak, PCA versus LDA, IEEE Trans. Pattern Anal. Mach. Intell., 23(2)(2001) 228-233.
13. L. Shang, D.S.Huang, J. X. Du, and C. H. Zheng, " Palmprint recognition using FastICA algorithm and radial basis probabilistic neural network," Neurocomputing, vol.69, nos.13-15, pp. 1782-1786, 2006
14. B. Scholkopf, A. Smola, K.R. Muller, Nonlinear Component Analysis as a Kernel Eigenvalue Problem, Neural Computation, 10(5)(1998)1299-1319.
15. Y. Wen, L. He, P. Shi, Face recognition using difference vector plus KPCA, Digital Signal Process, 22(2012)140-146.
16. J. Yang, Z. Jin, J. Yang, D. Zhang, A. F. Frangi, Essence of Kernel Fisher Discriminant: KPCA plus LDA, Pattern Recognition, 37(10)(2004) 2097-2100.

17. D.S. Huang, Horace H.S.Ip, Z.R. Chi, A neural root finder of polynomials based on root moments, Neural Comp., 16(8)(2004)1721-1762.
18. D.S. Huang, A constructive approach for finding arbitrary roots of polynomials by neural networks, IEEE Trans. on Neural Netw., 15(2)(2004)477-491.
19. D.S. Huang, Z.R. Chi, W.C. Siu, A case study for constrained learning neural root finders, Applied Mathematics and Computation, 165(3)(2005)699-718.
20. D.S. Huang, Radial basis probabilistic neural networks: Model and application, Int. Journal of Pattern Recognit., and Artificial Intell., 13(7)(1999)1083-1101.
21. D.S. Huang, J.X. Du, A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks, IEEE Trans. Neural Networks, 19(12)(2008)2099-2115.
22. G. E. Hinton, R. R. Salakhutdinov, Reducing the Dimensionality of Data with Neural Networks , Science, 313(5786)(2006)504:507
23. S. Ji, W. Xu, M. Yang, K. Yu, 3D Convolutional Neural Networks for Human Action Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35 (1)(2013) 221–231
24. X. Yang , W. Liu , D. Tao, and J. Cheng, Canonical correlation analysis networks for two-view image recognition, Information Science, 385(2017) 338-352
25. I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al. Generative Adversarial Networks, In :Proceedings of Advances in Neural Information Processing Systems, pp.2672-2680, 2014
26. Y. Sun, X. Wang, X. Tang, Hybrid Deep Learning for Face Verification, In :Proceedings of 2013 IEEE International Conference on Computer Vision, pp.1489-1496, 2013.
27. Y. Sun, X. Wang, X. Tang, Deep Learning Face Representation by Joint Identification-Verification, In: Proceedings of 2014 IEEE International Conference on Computer Vision and Pattern Recognition, pp.1-9,2014.
28. K. Charalampous, A. Gasteratos, On-line deep learning method for action recognition, Pattern Analysis & Applications, 19(2)(2016)337-354.
29. K. Simonyan and A. Zisserman. Two-Stream Convolutional Networks for Action Recognition in Videos. In: Proceedings of Advances in Neural Information Processing Systems (NIPS 2014), PP.1-11, 2014.
30. T. Wilkinson, A. Brun, A Novel Word Segmentation Method Based on Object Detection and Deep Learning, In:Proceedings of International Symposium on Visual Computing, pp 231-240,2015
31. M. Brand, Charting a manifold, In proceedings of Neural Information Processing Systems, 2002.
32. S.T. Roweis, L.K. Saul, Nonlinear Dimensionality Reduction By Locally Linear Embedding, Science, 290(2000) 2323-2326.
33. M. Belkin, P. Niyogi, Laplacian Eigenmaps for Dimensionality Reduction and Data Representation, Neural Computation, 15(6)(2003)1373-1396.
34. D. Donoho, C. Grams, Hessian Eigenmaps: Locally linear embedding techniques for high dimensional data, Proceeding s of National Academy of Science, 100(2003) 5591-5595.
35. Z. Zhang, H. Zha, Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment, SIAM J. Scientific Computing, 26(1)(2005)313-338.
36. T. Lin, H.B. Zha, Riemannian manifold learning, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(5)(2008)796-809.
37. J.B. Tenenbaum, V. de Silva, J.C. Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction, Science, 290(2000) 2319-2323.

38. R.R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, S. W. Zucker, Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps, PNAS, 102(2005)7426–7431
39. K.Q. Weinberger, L.K. Saul, An introduction to nonlinear dimensionality reduction by maximum variance unfolding," In: American Association for Artificial Intelligence, 2006.
40. P. Hall, B. U. Park, R.J. Samworth. Choice of neighbor order in nearest-neighbor classification. Annals of Statistics. 36 (5)(2008): 2135–2152.
41. W.F. Schmidt , M. Kraaijveld , R.P. Duin , et al. , Feedforward neural networks with random weights, in: Proceedings of International Conference on Pattern Recognition, pp. 1-4 , 1992
42. H. Strange, R. Zwiggelaar, A generalised solution to the out-of-sample extension problem in manifold learning, in: Proceedings of AAAI Conference on Artificial Intelligence, 2011
43. .H. Strange , R. Zwiggelaar , Open Problems in Spectral Dimensionality Reduction, Springer Briefs in Computer Science, 2014
44. X. He, S. Yang, Y. Hu, P. Niyogi,, and H. J. Zhang, Face Recognition Using Laplacianfaces, IEEE Trans. Pattern Analysis and Machine Intelligence, 27(3)(2005)328-340.
45. X. He and P. Niyogi, Locality preserving projections. Advances in Neural Information Processing Systems, pp. 153–160, 2003
46. J.B. Li, J. S. Pan, and S. C. Chu .Kernel class-wise locality preserving projection. Information Science, 178(7)(2008):1825–1835
47. G. Feng, D. Hu, D. Zhang, and Z. Zhou. An alternative formulation of kernel LPP with application to image recognition.Neurocomputing, 69(13-15)(2006): 1733-1738
48. X. He, D. Cai, and P. Niyogi, Tensor Subspace Analysis. Advances in Neural Information Processing Systems, pp.499-506, 2005
49. P. Jia, J. Yin, X. Huang, and D. Hu. Out-of-sample algorithm of Laplacian eigenmaps applied to dimensionality reduction, http://www.paper.edu.cn,2008.
50. S. Mousazadeh and I. Cohen. Out-of-sample extension of band-limited functions on homogeneous manifolds using diffusion maps. Elsevier North-Holland, 2015.
51. M. Tang, F. Nie, and R. Jain. A graph regularized dimension reduction method for out-of-sample data. Neurocomputing, 225(2017):58-63
52. Y. Han, Z. Xu, Z. Ma Z, and Z. Huang. Image classification with manifold learning for out-of-sample data. Signal Processing, 93(8)(2013):2169-2177.
53. K. Yu , T. Zhang , Y. Gong, Nonlinear learning using local coordinate coding, in: Proceedings of Advances in Neural Information Processing Systems, 2009, pp. 223–231 .
54. K. Zhang , J. Kwok , Density-weighted Nyström method for computing large kernel eigen systems, Neural Comput. 21 (2009) 121–146 .
55. H. Li, H. Jiang, R. Barrio, and L. Chen. Incremental manifold learning by spectral embedding methods. Pattern Recognition Letters, 32(10)(2011):1447-1455.
56. P. Jia, J. Yin J, X. Huang, and D. Hu. Incremental Laplacian eigenmaps by preserving adjacent information between data points. Pattern Recognition Letters, 30(16)(2009):1457-1463.
57. Q. Xuand J. Hu. Fault feature extraction method for compressor based on improved incremental Laplacian eigenmap algorithm. Chinese Journal of Scientific Instrument, 34(4)(2013):791-796.
58. H.M.Yang and H.E. Pi-Lian. Image Recognition Based on Incremental Laplacian Eigenmap and SVM. Computer Simulation, 24(11)(2007):221-223.

59. A. M. Quispe, C. Petitjean and L. Heutte. Extreme learning machine for out-of-sample extension in Laplacian eigenmaps. Pattern recognition letters, 74(2016)68-73

60. L. Yang, S. Yang, S, Li S, and J. Jiao. Incremental Laplacian Regularization Extreme Learning Machine for Online Learning. Applied Soft Computing, 59(2017):546-555

61. H. White. An additional hidden unit test for neglected nonlinearity in multi- layer feedforward networks, in: Proceedings of International Joint Conference on Neural Networks, pp. 451-455,1989

62. H. White. Approximate nonlinear forecasting methods, Handb. Econ. Forecast. 1 (2006):459-512 .

63. C.P. Chen.A rapid supervised learning neural network for function interpolation and approximation, IEEE Trans. Neural Netw. 7 (1996) :1220-1230 .

64. B. Igelnik and Y.H. Pao.Stochastic choice of basis functions in adaptive function approximation and the functional-link net, IEEE Trans. Neural Netw. 6 (1995) :1320-1329 .

65. Y.H. Pao , G.H. Park , and D.J. Sobajic , Learning and generalization characteristics of the random vector functional-link net, Neurocomputing 6 (1994):163-180.

66. W.F. Schmidt , M. Kraaijveld , R.P. Duin, Feedforward neural networks with random weights, in: Proceedings of International Conference on Pattern Recognition, pp. 1-4, 1992

67. A. Jansen, G. Sell, and V. Lyzinski V. Scalable out-of-sample extension of graph embeddings using deep neural networks[J]. Pattern Recognition Letters,94(2017):1-6

68. H. Zhao, S, Sun, Z. Jing, and J. Yang, Local Structure Based Supervised Feature Extraction, Pattern Recognition, 39(2006)1546-1550.

69. B. Li, C. Wang, and D.S.Huang, Supervised feature extraction based on orthogonal discriminant projection, Neurocomputing, 73(1-3)(2009): 191-196

70. D. Jin and B. Li. Distance-Weighted Manifold Learning In Facial Expression Recognition. In: Proceedings of International Conference on Industrial Electronics and Applications, pp.1776-1780, 2016

71. D. de Ridder, O. Kouropteva, O. Okun. “Supervised locally linear embedding,” In: Proceedings of ICANN, pp 333–341,2003

72. G. Wen, L. Jiang. “Clustering-based locally linear embedding,” In: Proceedings of 2006 IEEE international conference on systems, man and cybernetics, pp 4192–4196, 2006

73. S. Zhang. “Enhanced supervised locally linear embedding,” Pattern Recognition Letter, 30(13)(2009):1208-1218

74. K. Hui, C. Wang. “Clustering-based locally linear embedding,” In: Proceedings of ICPR, 2008

75. Q. Zhao, D. Zhang, H. Lu. “Supervised LLE in ICA space for facial expression recognition,” In: Proceedings of ICNNB’05, pp 1970–1975, 2005

76. P.Y. Han, A.T. J. Beng, W.E. Kiong. “Neighborhood discriminant locally linear embedding in face recognition,” In: Proceedings of CGIV2008, pp 223–228, 2008

77. Z. Zhang, L. Zhao. “Probability-based locally linear embedding for classification,” In: Proceedings of FSKD, pp 243–247, 2007

78. L. Zhao, Z. Zhang. “Supervised locally linear embedding with probability-based distance for classification,” Computation Mathematics Application, 57(6):919-926, 2009

79. M. Sugiyama, Local Fisher Discriminant Analysis for Supervised Dimensionality Reduction, In proceedings of International Conference on Machine Learning, 2006, pp.905-912

80. Z. Li, D.Lin, X. Tang, Nonparametric Discriminant Analysis for Face Recognition, IEEE Transactions on Pattern analysis and machine intelligence, 31(4)(2009)755-761.

81. J. Yang, Z. Lou, Z. Jin, J. Yang, Minimal Local Reconstruction Error Measure Based Discriminant Feature Extraction and Classification, In proceedings of Conference on Computer Vision and Pattern Recognition, pp. 1-6, 2008.

82. Y. Chen, Z. Jin. Reconstructive discriminant analysis: A feature extraction method induced from linear regression classification, Neurocomputing, 87(2012)41-50.

83. D. Cai, X. He, K. Zhou, J. Han, H. Bao, Locality Sensitive Discriminant Analysis, In proceedings of the 20th international joint conference on Artificial intelligence, pp.708-713, 2007

84. H.T. Chen, H.W. Chang, T.L. Liu, Local Discriminant Embedding and Its Variants, In Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition, pp.846-853, 2005

85. T. Zhang, J. Yang, H. Wang, C. Du, Maximum variance projection for face recognition, Opt. Eng. 46 (6) (2007) 1–8.

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

87. B. Li，Z.T. Fan， and X. L. Zhang. “Feature Space Distance Metric Learning for Discriminant Graph Embedding”. International Joint Conference on Neural Networks (IJCNN2016), pp.1992-1998, 2016

88. B. Li and D. S. Huang. Maximum Generalized Fisher Criterion, In: proceedings of IEEE International Conference on Computer Science and Information Technology, PP.349-352, 2010.

89. M. Wan, S. Gai, and J. Shao. Local graph embedding based on maximum margin criterion (LGE/MMC) for face recognition Informatica, 36(1)(2012):103-112

90. Huang P, Tang Z, Chen C, and Z. Yang. Local Maximal Margin Discriminant Embedding for Face Recognition[J]. Journal of Visual Communication & Image Representation, 25(2)(2014):296-305.

91. Y. N. Chen, C. C. Han, C. T. Wang, and K. C. Fan, Face recognition using nearest feature space embedding, ”IEEE Transactions on Transactions Pattern Analysis and Machine Intelligence, 33(6)(2011):1073-1086

92. S.Z. Li, Face Recognition Based on Nearest Linear Combinations, Proc. Computer Vision and Pattern Recognition, 1998.

93. S.Z. Li and J. Lu, Face Recognition Using the Nearest Feature Line Method,” IEEE Transactions on Neural Networks, 10(2)(1999): 439-433

94. S.Z. Li, K.L. Chan, and C.L. Wang, “Performance Evaluation of the Nearest Feature Line Method in Image Classification and Retrieval,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11)(2000): 1335-1339

95. J.T. Chien, C.C. Wu, Discriminant Waveletfaces and Nearest Feature Classifiers for Face Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(12)(2002) 1644-1649.

96. Y. Pang, Y. Yuan, and X. Li. Iterative Subspace Analysis Based on Feature Line Distance, IEEE Transactions on Image Processing, 18(4)(2009) 903-907.

97. J. Lu, Y. P. Tan, Uncorrelated Discriminant Nearest Feature Line Analysis for Face Recognition, IEEE signal processing letter, 17(2)(2010)185-188

98. P.L. Lai and C. Fyfe. Kernel and Nonlinear Canonical Correlation Analysis[J]. International Journal of Neural Systems, 10(5)(2000):614-619

99. S.Y. Huang, M.H. Lee, and C.K. Hsiao. Nonlinear measures of association with kernel canonical correlation analysis and applications. Journal of Statistical Planning & Inference, 139(7)(2009):2162-2174.

100. B. Fortuna.Kernel canonical correlation analysis with applications. In proceedings of Sikdd, pp.12-15, 2004

101. M. Brück, X. Du, j. Park, and C.L. Terng. The submanifold geometries associated to Grassmannian systems. American Mathematical Society, 2002.

102. R. Fioresi. Quantum Deformation of the Grassmannian Manifold, Journal of Algebra, 214(2)(1999):418-447.

103. Q. Wang, J. Gao, and H. L. Grassmannian Manifold Optimization Assisted Sparse Spectral Clustering, In:Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.3145-3153, 2017

104. B. Li, J. Li, and X.P. Zhang, Nonparametric Discriminant Multi-manifold Learning for Dimensionality Reduction, Neurocomputing, 152(3)(2015): 121-126

105. B. Li, J. Li, and J. Liu. "Nonparametric Discriminant Multi-manifold Learning," In: Proceedings of International Conference on Intelligent Computing(ICIC 2014), pp. 113-119,2014

106.R. Wang, S. Shan, X. Chen, and W. Gao. Manifold-manifold distance and its application to face recognition with image sets. IEEE Transactions on Image Processing, 21(10)(2012):4466-4479.

107.R. Wang, S. Shan, X. Chen, and W. Gao.. Manifold-Manifold Distance with application to face recognition based on image set. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.1-8, 2008

108. V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric Means In A Novel Vector Space Structure On Symmetric Positive-definite Matrices. Siam Journal on Matrix Analysis & Applications, 29(1)(2007):328-347.

109. F. Yger and M. Sugiyama. Supervised LogEuclidean Metric Learning for Symmetric Positive Definite Matrices. Computer Science, 2015.

110. M. Harandi, M. Salzmann, and R. Hartley R. Dimensionality Reduction on SPD Manifolds: The Emergence of Geometry-Aware Methods. IEEE Transactions on Pattern Analysis & Machine Intelligence, 40(1)(2018):48-62.

111.M. Harandi, M. Salzmann, and R. HartleyR. From Manifold to Manifold: Geometry-Aware Dimensionality Reduction for SPD Matrices, In proceedings of ECCV 2014, pp.17-32, 2014

112. M. Hagedoorn and R.C. Veltkamp. Reliable and Efficient Pattern Matching Using an Affine Invariant Metric. International Journal of Computer Vision, 31(2-3)(1999):203-225.

113. M. Werman and D. Weinshall. Similarity and affine invariant distances between 2D point sets. IEEE Transactions on Pattern Analysis & Machine Intelligence, 17(8) (1995):810-814.

114. V. V. Makeev. On one affine-invariant metric on the class of convex plane compacts. Journal of Mathematical Sciences, 110(1999):194–197.

115.Z. Huang, R. Wang, S. Shan, and X. Chen. Projection Metric Learning on Grassmann Manifold with Application to Video based Face Recognition. In: Proceedings of Computer Vision and Pattern Recognition, pp.140-149, 2015

116. M. Yukawa, K. Slavakis, and I. Yamada I. Adaptive Parallel Quadratic-Metric Projection Algorithms. IEEE Trans Audio Speech & Language Processing, 15(5)(2007):1665-1680.

117. P. Howlanda, J. Wangb, H. Parkc, Solving the small sample size problem in face recognition using generalized discriminant analysis, Pattern Recognition,39 (2006) 277-287.

118. W. Zheng, L. Zhao, C. Zou, An efficient algorithm to solve the small sample size problem for LDA, Pattern Recognition 37 (2004) 1077-1079.

119. J. Ye, R. Janardan, C.H. Park, H. Park, An optimization criterion for generalized discriminant analysis on undersampled problems, IEEE Trans. Pattern Anal. Mach. Intell. 26 (8) (2004) 982-994.

120. J. Ye, Q. Li, A two-stage linear discriminant analysis via QR-decomposition, IEEE Trans. Pattern Anal. Mach. Intell. 27 (6) (2005) 929-941.

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

122. J. Lu , Y. P. Tan, G. Wang, Discriminative multi-manifold analysis for face recognition from a single training sample per person, IEEE Transactions on Pattern analysis and machine intelligence, 35(1)(2013) 39-51.

123. H. Li, T. Jiang, K. Zhang, Efficient and robust feature extraction by maximum margin criterion, IEEE Trans. Neural Networks, 17 (1) (2006) 157–165

124. S. Yan, D. Xu, B. Zhang, and H.J. Zhang, “Graph Embedding: A General Framework for Dimensionality Reduction,” IEEE Trans. Pattern Anal. Mach. Intell., 29(1)(2007):40-51

125. M.M.Adankon and M. Cheriet M. Support Vector Machine. Computer Science, 1(4)(2002):1-28

126. G. Cauwenberghs and T. Poggio. Incremental and decremental support vector machine learning. In:ProceedingsofInternational Conference on Neural Information Processing Systems, pp.388-394, 2000

127. S. Amari and S. Wu. Improving support vector machine classifiers by modifying kernel functions. Neural Networks, 1999, 12(6)(1999):783-789.

128. D. Cai, X. He, J. Han, and H. Zhang, “Orthogonal Laplacianfaces for Face Recognition,” IEEE Trans. Image Process., 15(11)(2006):3609-3614

129. Z. Jin, J. Yang, Z. Hu and Z. Lou. Face Recognition Based on the Uncorrelated Discrimination Transformation, Pattern Recognition, 34(7)(2001): 1405-1416

130. X. Y. Jing, D. Zhang and Y. Y. Tang. An Improved LDA Approach, IEEE Trans. on Systems, Man, and Cybernetics-part B: Cybernetics, 34(5)(2004):1942-195

131. J. Ye, T. Li, T. Xiong and R. Janardan. Using Uncorrelated Discriminant Analysis for Tissue Classification with Gene Expression Data, IEEE/ACM Trans. on Computational Biology and Bioinformatics, 1(4)(2004):181-190

132. J. Yang, J. Yang and D. Zhang. What’s wrong with Fisher Criterion? Pattern Recognition, 35(11)(2002):2665-2668

133. Z. Jin, J. Y. Yang, Z. M. Tang, Z. S. Hu. A theorem on the uncorrelated optimal discriminant vectors, Pattern Recognition, 34(10)(2001):2041–2047

134. J. Ye, R. Janardan, Q. Li. Feature reduction via generialized uncorrelated linear discriminant analysis, IEEE Trans. On Knowledge and Data Engineering, 18(10)(2006):1312-1322

135. X. Yu and X.Wang, Uncorrelated discriminant locality preserving projections, IEEE Signal Process.Lett.15(2008)361-364

136.J. Lu and Y.P. Tan. Uncorrelated Discriminant Nearest Feature Line Analysis for Face Recognition. IEEE Signal Processing Letters, 17(2)(2009):185-188.

137. X. Jing, S. Li, D. Zhang, and J. Yang. Face recognition based on local uncorrelated and weighted global uncorrelated discriminant transforms. In: Proceedings of IEEE International Conference on Image Processing, pp.3049-3052, 2011

138. Y. Chen, W. S. Zheng, X. H. Xu, J. H. Lai. Discriminant subspace learning constrained by locally

statistical uncorrelation for face recognition, Neural Networks, 42(1)(2013):28-43

139. J. Wright, A. Yang, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210-227.
140. Robert Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, 58(1):267–288, 1996.
141. I. Drori and D. Donoho, “Solution of L1 minimization problems by LARS/Homotopy methods,” in ICASSP, 3(2006) 636–639.
142. L. Zhang, M.Yang, Z. Feng and D. Zhang. On the Dimensionality Reduction for Sparse Representation based Face Recognition. In: Proceedings of International Conference on Pattern Recognition , pp.1237- 1240, 2010
143. J. Zhang, J. Wang, and X. Cai. Sparse locality preserving discriminative projections for face recognition Neurocomputing, 260(2017):321-330.
144. L. Qiao, S. Chen, and X. Tan, “Sparsity preserving projections with applications to face recognition,” Pattern Recognition, 43 (1)(2010) 331-341.
145. J. Gui, Z. Sun, W. Jia, and Y. Lei. Discriminant sparse neighborhood preserving embedding for face recognition. Pattern Recognition, 2012, 45(8)(2012):2884-2893.
146. J. Yang, D. Zhang, J. Y. Yang, and B. Niu, “Globally Maximizing, Locally Minimizing: Unsupervised Discriminant Projection with Application to Face and Palm Biometrics,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 4, pp. 650-664, Apr. 2007.
147. N. Zheng, X. Guo, L. Qi, and L. Guan. Two-dimensional discriminant multi-manifolds locality preserving projection for facial expression recognition. In: Proceedings of Conference of the European Chapter of the Association for Computational Linguistics, pp.398-405, 2009
148. X. Guo, Y. Tie, L. Qi and L. Guan, A Novel Semi-Supervised Dimensionality Reduction Framework for Multi-Manifold Learning, IEEE MultiMedia, 99(2009): 1-1
149. W. Yang, C. Sun and L. Zhang, Face Recognition Using a Multi-manifold Discriminant Analysis Method, In: Proceedings of 2010 20th International Conference on Pattern Recognition, pp. 527-530, 2010.
150. J. Jiang, R. Hu, Z. Han, K. Huang and T. Lu, Graph discriminant analysis on multi-manifold (GDAMM): A novel super-resolution method for face recognition, In: Proceedings of 2012 19th IEEE International Conference on Image Processing, pp. 1465-1468, 2012
151. Y. Wang, Y. Jiang, Y. Wu and Z. H. Zhou, Spectral Clustering on Multiple Manifolds, IEEE Transactions on Neural Networks, 22(7)(2011): 1149-1161
152. H. Hu, Sparse Discriminative Multimanifold Grassmannian Analysis for Face Recognition With Image Sets, IEEE Transactions on Circuits and Systems for Video Technology, 25(10)(2015): 1599-1611
153. L. Huang, J. Lu, and Y. P. Tan. Multi-manifold metric learning for face recognition based on image sets. Journal of Visual Communication & Image Representation, 25(7)(2014)1774-1783.
154. J. Li, Y. Wu, J. Zhao, and K. Lu. Multi-manifold Sparse Graph Embedding for Multi-modal Image Classification[J]. Neurocomputing, 173(P3)(2016):501-510.
155. W. Deng, J. Hu, J. Guo, H. Zhang, and C. Zhang, Comments on 'Globally Maximizing, Locally Minimizing: Unsupervised Discriminant Projection with Application to Face and Palm Biometrics, IEEE Trans. Pattern Anal. Mach. Intell., 29(8)(2007):1503-1504
156. W. Yang, C. Sun, J. Yang, and K. Ricanek. Face Recognition Using Kernel UDP. Neural Processing Letters, 34(2)(2010):177-192.

157. Q. Wang, R. Zhang, H. Pan, and S. Lou. Face recognition based on Log-Gabor and orthogonal UDP. Computer Science, 21(2011): 714-719
158. Y. Li, G. He, and J. Yang. 2DUDP: Novel method of feature extraction based on image matrix. In: Proceedings of Control Conference, pp.490-494,2008
159. J. Jiang, Z. Feng, H. Xiong, and S. Tian. The Method of Feature Extraction Based on Modular Two Dimension Unsupervised Discriminant Projection, In: Proceedings of International Conference on Electric Information & Control Engineering, pp.691-695,2012
160. S. Haykin. Neural Networks: A Comprehensive Foundation (3rd Edition). Macmillan, 1998.
161. J. S. Bing and J. Choi. Back-Propagation Neural Networks. Springer US, 1995.
162. H. C. Hsin, C. C. Li, M. Sun, and R. J. Sclabassi. An adaptive training algorithm for back-propagation neural networks. IEEE Transactions on Systems Man & Cybernetics, 25(3)(1992):512-514.
163. Z. Cai, Q. Fan, R.S. Feris, and N, Vasconcelos. A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection. In: Proceedings of European Conference on Computer Vision. Springer, pp.354-370, 2016
164. A. Brun, C. F. Westin, M. Herberthson, and H. Knutsson, Fast manifold learning based on Riemannian normal coordinates, In: Proceedings of 14th Scandinavian Confonference on Image Analysis, pp. 920-929, 2005.
165. F. Sha and L. K. Saul, Analysis and extension of spectral methods for nonlinear dimensionality reduction, In: Proceedings of 22nd International conference on Machine Learning, pp. 785-792, 2005
166. L. Yang, Locally multidimensional scaling for nonlinear dimensionality reduction, In: Proceedings of 18th International conference on Pattern Recognition, pp. 202-205, 2006.
167. Y. W. Teh and S. T. Roweis. Automatic alignment of hidden representations, In: Proceedings of Advances in Neural Information Processing Systems, pp. 841-848, 2002.
168. Rasmussen C E. The infinite Gaussian mixture model. In: Proceedings of International Conference on Neural Information Processing Systems. MIT Press, 1999:554-560.
169. G. E. Hinton and S. T. Roweis, Stochastic neighbor embedding, In: Proceedings of Advances in Neural Information Processing Systems, pp. 833-840, 2002
170. L. van der Maaten and G. E. Hinton, Visualizing data using t-SNE, Journal of Machine Learning Research, 9(2008):2579-2605
171. D.S.Huang, Systematic Theory of Neural Networks for Pattern Recognition, House of Electronic Industry of China, 1996.
172. C. Chen, L. Zhang, J. Bu, C. Wang， and W Chen. Constrained Laplacian Eigenmap for dimensionality reduction. Neurocomputing, 73(4–6)(2010):951-958.
173. S. T. Tu, J. Y. Chen, W. Yang, and S. Hu. Laplacian Eigenmaps-Based Polarimetric Dimensionality Reduction for SAR Image Classification. IEEE Transactions on Geoscience & Remote Sensing, 50(1)(2011):170-179.
174. J. Park, Z. Zhang, H. Zha and R. Kasturi, Local smoothing for manifold learning, In : Proceedings of International Conference on Computer Vision and Pattern Recognition, pp.52-59, 2004
175. P. Qiu, The local piecewisely linear kernel smoothing procedure fitting jump regression surfaces, Technometrics, 46(1) (2004):8-98.
176. P. W. Holland and R. E. Welsch, Robust Regression using Iteratively Reweighted Least-Squares, Communications in statistics, 6(9)(1977):813-827
177. C. Hou, J. Wang, Y. Wu, and D. Yi, Local linear transformation embedding, Neurocomputing,

72 (2009): 2368-2378.

178. C. Hou, C. Zhang, Y. Wu, and Y. Jiao, Stable local dimensionality reduction approaches, Pattern Recognition, 42 (2009): 2054-2066.
179. J. CHEN and Z. MA. Locally linear embedding: A review. International Journal of Pattern Recognition & Artificial Intelligence, 25(07)(2011):985-1008.
180. J. Chen and Y. Liu. Locally linear embedding: a survey. Kluwer Academic Publishers, 2011.
181. J. W., Xu, A. R.. C. Paiva, I. Park I, and J. C. Principe. A Reproducing Kernel Hilbert Space Framework for Information-Theoretic Learning. IEEE Transactions on Signal Processing, 56(12)(2008):5891-5902.
182. B. Yekkehkhany, A. Safari, S. Homayoun, and M. Hasanlou. a Comparison Study of Different Kernel Functions for Svm-Based Classification of Multi-Temporal Polarimetry SAR Data. Information Sciences, 3(2)(2014):281-285.
183. G. B. Huang, Q. Y. Zhu, and C. K. Siew. Extreme learning machine: Theory and applications. Neurocomputing, 70(1)(2006):489-501.
184. Yale University Face Database, _http://cvc.yale.edu/projects/yalefaces/yalefaces.html, 2002.
185. http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html.
186. B. Gong, Y. Shi, F. Sha, and K. Grauman. Geodesic flow kernel for unsupervised domain adaptation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.2066-2073, 2012
187. P.J. Phillips, The Facial Recognition Technology (FERET) Database, http://www.itl.nist.gov/iad/humanid/feret/feret_ master.html, 2006
188. T. Sim, S. Baker, and M. Bsat The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces tech. report CMU-RI-TR-01-02, Robotics Institute, Carnegie Mellon University, Jan. 2001.
189. B. Li, Y. Peng, and X. Wang, Local uncorrelated subspace learning with Point to Feature Space Distance Metric. Submitted, 2017.
190. M. D. Yuan, D. Z. Feng, Y. Shi, and C. Xiao. Adaptive graph orthogonal discriminant embedding: an improved graph embedding method. Neural Comput & Applic, https://doi.org/10.1007/s00521-018-3374-8, 2018
191. X. Liu, H. Lu, and H. Luo. Smooth Multi-Manifold Embedding for Robust Identity-Independent Head Pose Estimation. In: Proceedings of International Conference on Computer Analysis of Images and Patterns, pp.66-73, 2009
192. G. Feng, D. Zhang, J. Yang, and D. Hu. A theoretical framework for matrix-based feature extraction algorithm with its application to image recognition. International Journal of Image & Graphics, 2008, 08(01)(2008):0800294-.
193. W. Yang, C. Sun C, and Z. Lei. A multi-manifold discriminant analysis method for image feature extraction. Pattern Recognition, 44(8)(2011):1649-1657.
194. J. Jiang, R. Hu, Z. Han, K. Huang, and K. Lu. Graph discriminant analysis on multi-manifold (GDAMM): A novel super-resolution method for face recognition. In: Proceedings of IEEE International Conference on Image Processing. pp.1465-1468,2013
195. J. See and M. F. A. Fauzi. Learning Neighborhood Discriminative Manifolds for Video-Based Face Recognition. In: Proceedings of International Conference on Image Analysis and Processing, pp.247-256, 2011
196. Y. L. Yu and L. M. Zhang. Orthogonal MFA and uncorrelated MFA. Pattern Recognition &

Artificial Intelligence, 21(2008): 603-608

197. C. Zhao, Z. Lai,, D. Miao, Z . Wei , and C. Liu. Graph embedding discriminant analysis for face recognition. Neural Computing & Applications, 24(7-8)(2014):1697-1706.
198. R. Wang and X. Chen. Manifold Discriminant Analysis. In: Proceedings of International Conference on Computer Vision and Pattern Recognition, pp.429-436, 2009
199. M. Wa. Maximum inter-class and marginal discriminant embedding (MIMDE) for feature extraction and classification. Neural Computing & Applications, 21(7)(2012):1737-1743.
200. J. He, D. Wu, N. Xiong, and C. Wu.. Orthogonal margin discriminant projection for dimensionality reduction. Journal of Supercomputing, 72(6)(2016):2095-2110.
201. B. Raducanu and F. Dornaika F. A supervised non-linear dimensionality reduction approach for manifold learning[J]. Pattern Recognition, 45(6)(2012):2432-2444.
202. A.W. Fitzgibbon and A. Zisserman. Joint manifold distance: a new approach to appearance based clustering. In: Proceedings of Computer Vision and Pattern Recognition, pp.1-8, 2003.
203. E. Kokiopoulou and P. Frossard. Minimum distance between pattern transformation manifolds: algorithm and applications. IEEE Transactions on Pattern Analysis & Machine Intelligence, 31(7)(2009):1225-1238.
204. J. M. Guo. The Laplacian spectral radius of a graph under perturbation. Computers & Mathematics with Applications, 54(5)(2007):709-720.
205. D.S.Huang and W.B.Zhao. Determining the centers of radial basis probabilistic neural networks by recursive orthogonal least square algorithms, Applied Mathematics and Computation, 162(1)(2005): 461-473
206. W.B.Zhao, D.S.Huang, J.Y. Du, and L.M. Wang, Genetic optimization of radial basis probabilistic neural networks, International Journal of Pattern Recognition and Artificial Intelligence, 18(8)(2004): 1473-1500
207. Z.Q. Zhao and D.S.Huang, Palmprint recognition with 2DPCA+PCA based on modular neural networks, Neurocomputing, 71(1-3)(2007): 448-454
208. D.S.Huang and J.X. Mi, A new constrained independent component analysis method, IEEE Trans. On Neural Networks, 18(5)(2007): 1532-1535
209. K. H. Liu, and D.S.Huang, Cancer classification using rotation forest, Computers in Biology and Medicine, 38(5)(2008):601-610
210. C. H. Zheng, D.S.Huang, and L. Shang, Feature selection in independent component subspace for microarray data classification, Neurocomputing, 69(16-18)(2006): 2407-2410
211. Z. L. Sun, D.S.Huang, and Y. M. Cheung, Extracting nonlinear features for multispectral images by FCMC and KPCA, Digital Signal Processing, 15(4)(2005): 331-346
212. J. X. Du, D.S.Huang, X. F. Wang, and X. Gu, Computer-aided plant species identification (CAPSI) based on leaf shape matching technique, Transactions of the Institute of Measurement and Control, 28(3)(2006): 275-284
213. C. Y. Lu and D.S.Huang, Optimized projections for sparse representation based classification, Neurocomputing, 113(2013): 213-219

Bo Li received his M.Sc. and PhD degree in Mechanical and Electronic Engineering from Wuhan University of Technology in 2003, Pattern Recognition and Intelligent System from University of Science and Technology of China in 2008, respectively. Now, he is an associated professor at School of Computer Science and Technology, Wuhan University of Science and Technology. He is also a research associated in Ryerson University. His research interests include machine learning, pattern recognition, image processing and bioinformatics.

Yan-Rui Li received his B.S. in Computer Science and technology from Wuhan University of science and Technology in 2013. Now, he is a master candidate in School of Computer science and technology, Wuhan University of science and Technology. Her research interests include machine learning, image processing.

Xiao-Long Zhang received his B.S. and M.S. in Department of Computer Science, Northeastern University, China in 1985 and 1988, respectively, and his Ph.D. in Department of Computer Science from Tokyo Institute of Technology in 1998. Now, he is a professor at School of Computer Science and Technology, Wuhan University of Science and Technology. He is also a senior member of Chinese Association for Artificial Intelligence. His research interests include machine learning, knowledge discovery with big data, data mining and bioinformatics.