Journal of Computers Vol. 30 No. 6, 2019, pp. 1-15
doi:10.3966/199115992019123006001

Face Recognition Based on Exponential Neighborhood Preserving Discriminant Embedding

Rui-Sheng Ran[1,2], Bin Fang[2*], Shou-Gui Zhang[3], Dong Qiu[4], Ji-Ye Qian[5]

[1] College of Computer and Information Science, Chongqing Normal University, Chongqing, 401331, China
rshran@cqnu.edu.cn

[2] College of Computer Science, Chongqing University, Chongqing, 400044, China
fb@cqu.edu.cn

[3] School of Mathematical Sciences, Chongqing Normal University, Chongqing, 401331, China
shgzhang@cqnu.edu.cn

[4] College of Mathematics and Physics, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China
dongqiumath@163.com

[5] State Grid Chongqing Electric Power Research Institute, Chongqing, 401123, China
qianjiye@cqu.edu.cn

Received 31 October 2017; Revised 14 June 2018; Accepted 1 August 2018

Abstract. As a manifold reduced dimensionality technique, neighborhood preserving discriminant embedding (NPDE) and its variant version have been proposed recently. But NPDE and its variant version have the so-called small-sample-size (SSS) problem. In this paper, the NPDE method is taken as the representative and an exponential neighborhood preserving discriminant embedding (ENPDE) is proposed to address the SSS problem. The main idea of ENPDE is that the matrix exponential is introduced to NPDE. ENPDE has two superiorities. First, ENPDE avoids the SSS problem. Second, ENPDE has a diffusion effect on the distance between samples belonging to different classes in the neighborhood, and then the discrimination property is emphasized. The experiments are conducted on three face databases: Yale, CMU-PIE and AR. The proposed ENPDE method is compared with the global method, including PCA, LDA, EDA, and the unsupervised and supervised neighborhood preserving embedding methods, including NPE, ENPE, NPDE, and the two-dimension NPDE methods, including 2DDNPE, B2DNPDE. The experiment results show that the performance of ENPDE are better than those of the above methods.

Keywords: face recognition, manifold learning, matrix exponential, neighborhood preserving discriminant embedding, the small-sample-size problem

1 Introduction

Face recognition is an active research area in pattern recognition. In face recognition, high-dimensional data usually includes redundant information and is computationally heavy. So, dimensionality reduction is an effective approach to deal with this problem. The most well-known dimensionality reduction methods are principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2]. However, PCA and LDA aim only to preserve the global structures of the image samples and cannot uncover the essential manifold structure of the image.

In the past decade, manifold learning is an active research area in pattern recognition field. A number

* Corresponding Author

of manifold learning algorithms have been developed, such as locally linear embedding (LLE) [3], Isomap [4], local tangent space alignment (LTSA) [5], Laplacian eigenmaps (LE) [6]. These manifold learning algorithms discover the intrinsic geometry structure of a data set and have been widely used in the past decade. Unfortunately, all of these algorithms suffer from the out-of-sample problem [7]. To address this problem, a linearization procedure is developed, which construct a linear map from the original data space to new low-dimensionality space. Representative ones are local preserving projection (LPP) [8] and neighborhood preserving embedding (NPE) [9]. LPP is a linearization version of LE and NPE is a linearization version of LLE. In some sense, NPE and LPP provide two different ways to linearly approximate the eigenfunctions of the Laplace Beltrami operator [9], which method of the two is better remains unsettled.

Due to its simplicity and effectiveness, NPE has become a popular method in pattern recognition field. But NPE is a type of unsupervised technique in which the class-specific information of data is not used. As a technique encoding the discriminant information on feature extraction, neighborhood preserving discriminant embedding (NPDE) [10-11] is proposed. Some of other discriminant neighborhood preserving embedding versions have also been investigated, such as [12-13]. NPDE and its variant versions are all based on the graph embedding and the Fisher's criterion, and then the class discriminant information is considered into NPE. The difference of all versions is that the objective function constructed is in different form. So NPDE and its variant versions are similar in essence. Compared with NPE, the class specific information during training phase for constructing projection directions is utilized, so NPDE has more discriminant power.

In the recent years, lots of improved versions of supervised NPE have been investigated. One way of NPDE extension is that the feature extraction is based on two-dimension image matrices. For example, the two-dimensional discriminant neighborhood preserving embedding method (2DDNPE) [14] and the bilateral two-dimension neighborhood preserving discriminant embedding method (B2DNPDE) [15]. Due to the two-dimension methods are based on the two-dimension image matrices, rather than one-dimension image vectors, these methods avoid the SSS problem. Based on this, the two methods are used to compare with the proposed ENPDE method in this paper. Another way of NPDE extension is to combining sparse representation with neighborhood preserving discriminant embedding, for example [16-17]. In addition, there are some research combining the tensor with neighborhood preserving embedding, for example, the tensor train neighborhood preserving embedding (TTNPE) [18], the discriminant analysis via jointly $L_{2,1}$-norm sparse tensor preserving embedding [19], etc.

In most cases, the dimension of the sample is much larger than the number of the samples, like LDA, the generalized eigenvalue problem of NPDE may be unsolvable. This is the so-called small-sample-size (SSS) problem. NPDE and its variant version have to suffer from the SSS problem. In general, PCA can be adopted to reduce the dimensionality of the original image, and then NPDE can be used to extract the image feature. However, the processing results in the null spaces of the neighborhood scatter matrices can be discarded. But the null spaces may contain some discriminant information. As a consequence, some of significant information, which is contained in the original data, may be lost in the low-dimensional embedded data.

As an effective method, exponential discriminant analysis (EDA) [20] is proposed to overcome the SSS problem of classical LDA. The main idea of EDA is that the matrix exponential is introduced to LDA. It replaces the between-class scatter matrix S_B with the corresponding matrix exponential $\exp(S_B)$, and replaces the within-class scatter matrix S_W with the corresponding matrix exponential $\exp(S_W)$, and so avoids the singularity of the matrix S_B and S_W.

Since EDA was proposed, it is widely applied to solve the SSS problem, especially in the manifold learning field. Many of manifold learning algorithms, such as LPP [8], NPE [9], DLPP [21], LDE [22], SDE [23], have the SSS problem. The local preserving projection (LPP) is a linear projective map that arise by solving a variational problem that optimally preserves the neighborhood structure of the data set [8]. LPP is also a type of unsupervised technique. And then, based on the analysis of LDA, the discriminant locality preserving projection (DLPP) is proposed to improve the performance of LPP [21]. The local discriminant embedding (LDE) dissociates the sub-manifold of each class from one another, and specifically derives the embedding for nearest neighbor classification in a low-dimensional Euclidean space [22]. The semi-supervised discriminant embedding (SDE) is the semi-supervised extension of LDE [23]. Generally, this type of methods have to deal with high dimensional data, so the

SSS problem very often occurs. Then, the EDA method is introduced to address this problem. The exponential NPE (ENPE) [24], the exponential LPP (ELPP) [25], the exponential DLPP (EDLPP) [26], the exponential LDE (ELDE) [27] and the exponential SDE (ESDE) [28] are proposed. These methods are the exponential versions of the corresponding methods. They avoid the SSS problem and show better performance in face recognition.

In this paper, for the SSS problem of NPDE and its variant versions, a general matrix exponential method is proposed to address the problem. Based on the similarity of NPDE and its variant version, the NPDE method is taken as the representative to discuss in this paper. For the other variant version, with the similar processing, the matrix exponential method expects to get the same benefits. For simplicity, we only discuss how to apply the matrix exponential method to NPDE.

The main idea of ENPDE is that the matrix exponential is introduced to NPDE. The advantages of the proposed ENPDE method are two aspects. Firstly, ENPDE avoids the SSS problem of NPDE. Secondly, with the matrix exponential transformation, the neighborhood scatter matrices of NPDE are transformed to a new space, and then the NPDE criterion is applied in such a space. This transformation has the effect of distance diffusion, i.e., the distance between samples belonging to different classes in the neighborhood is enlarged. This is what we want in pattern classification. So, we can believe that ENPDE will show advantageous performance over NPDE and some other methods. It is proved by the experiment results in section 5. To the best of our knowledge, this paper is the first work to discuss and address this problem.

The rest of this paper is organized as follows. In section 2, the neighborhood preserving discriminant embedding method is reviewed. In section 3, exponential neighborhood preserving discriminant embedding (ENPDE) is presented. Section 4 provides the theoretical analysis of the proposed ENPDE method. Experimental results are shown in section 5. Finally, section 6 concludes this study.

2 Review of NPDE

Let $\Phi^{w}=\sum_{i}\left|y_{i}-\sum_{j}w_{ij}^{w}y_{j}\right|^{2}=YM^{w}Y^{T}$ represents the input data as *D*-dimensional data point. Each data point belongs to one of *c* classes $\Phi^{w}=\sum_{i}\left|y_{i}-\sum_{j}w_{ij}^{w}y_{j}\right|^{2}=YM^{w}Y^{T}$. Let *r* (=*N*/*c*) denotes the number of training samples belonging to the *l*th class. The main idea of NPDE [10] is from Fisher's criterion and the class-specific information of data is used. NPDE seeks an optimal projection ***A*** to embed the *D*-dimensional data set ***X*** into a low *d*-dimensional data set $\Phi^{w}=\sum_{i}\left|y_{i}-\sum_{j}w_{ij}^{w}y_{j}\right|^{2}=YM^{w}Y^{T}$ $\Phi^{w}=\sum_{i}\left|y_{i}-\sum_{j}w_{ij}^{w}y_{j}\right|^{2}=YM^{w}Y^{T}$, namely, $\Phi^{w}=\sum_{i}\left|y_{i}-\sum_{j}w_{ij}^{w}y_{j}\right|^{2}=YM^{w}Y^{T}$ $\Phi^{w}=\sum_{i}\left|y_{i}-\sum_{j}w_{ij}^{w}y_{j}\right|^{2}=YM^{w}Y^{T}$. And the local neighborhood structure of the original data set ***X*** can be preserved.

The algorithm procedure of NPDE is stated below:

NPDE first construct an adjacency graph on the data set. There are two ways to construct the adjacency graph. One way is *k* nearest neighbors (KNN), and another way is $\Phi^{w}=\sum_{i}\left|y_{i}-\sum_{j}w_{ij}^{w}y_{j}\right|^{2}=YM^{w}Y^{T}$ -neighborhood. Usually, KNN method is adopted to construct an adjacency graph.

Formulation of within-neighborhood scatter.

Let $\Phi^{w}=\sum_{i}\left|y_{i}-\sum_{j}w_{ij}^{w}y_{j}\right|^{2}=YM^{w}Y^{T}$ be a within-class reconstruction weight that is used to reconstruct $\boldsymbol{x}_i$ from its neighbors $\Phi^{w}=\sum_{i}\left|y_{i}-\sum_{j}w_{ij}^{w}y_{j}\right|^{2}=YM^{w}Y^{T}$. $\Phi^{w}=\sum_{i}\left|y_{i}-\sum_{j}w_{ij}^{w}y_{j}\right|^{2}=YM^{w}Y^{T}$ if both the *j*th local neighbor and the *i*th sample is from the same class; and $\Phi^w = \sum_i \left| y_i - \sum_j w_{ij}^w y_j \right|^2 = YM^wY^T$ otherwise. And let the within-class reconstruct matrix $\Phi^w = \sum_i \left| y_i - \sum_j w_{ij}^w y_j \right|^2 = YM^wY^T$. The reconstruct matrix $\Phi^w = \sum_i \left| y_i - \sum_j w_{ij}^w y_j \right|^2 = YM^wY^T$ can be solved by minimizing the following objective function:

$$\Phi^w = \sum_i \left| y_i - \sum_j w_{ij}^w y_j \right|^2 = YM^wY^T \quad \Phi^w = \sum_i \left| y_i - \sum_j w_{ij}^w y_j \right|^2 = YM^wY^T, \tag{1}$$

with constraint $\Phi^w = \sum_i \left| y_i - \sum_j w_{ij}^w y_j \right|^2 = YM^wY^T$.

The cost function is defined as follows:

$$\Phi^w = \sum_i \left| y_i - \sum_j w_{ij}^w y_j \right|^2 = YM^wY^T = Y\left(I - W^w\right)^T \left(I - W^w\right)Y^T,$$

where $\Phi^b = A^T X\left(I - W^b\right)^T \left(I - W^b\right)X^T A$ with $\Phi^b = A^T X\left(I - W^b\right)^T \left(I - W^b\right)X^T A$ is an identity matrix. Since $Y = A^T X$, the above cost function may be reformulated as:

$$\Phi^w = A^T XM^w X^T A = A^T X\left(I - W^w\right)^T \left(I - W^w\right)X^T A, \tag{2}$$

Formulation of between-neighborhood scatter.

The between-neighborhood scatter is may be presented similar to within-neighborhood scatter. Let $\Phi^b = A^T X\left(I - W^b\right)^T \left(I - W^b\right)X^T A$ be a between-class reconstruction weight that is used to reconstruct $\boldsymbol{x}_i$ from its neighbors $\Phi^w = \sum_i \left| y_i - \sum_j w_{ij}^w y_j \right|^2 = YM^wY^T$ and the between-class reconstruct matrix $\Phi^b = A^T X\left(I - W^b\right)^T \left(I - W^b\right)X^T A$. The reconstruction errors are measured by minimizing the following objective function:

$$\Phi\left(W^b\right) = \sum_{i=1} \left| x_i - \sum_{j=1}^{k} w_{ij}^b x_j \right|^2 \tag{3}$$

with constraint $\Phi^b = A^T X\left(I - W^b\right)^T \left(I - W^b\right)X^T A$.

The cost function is defined as follows:

$$\Phi^b = \sum_i \left| y_i - \sum_j w_{ij}^b y_j \right|^2 = YM^bY^T = Y\left(I - W^b\right)^T \left(I - W^b\right)Y^T,$$

where $\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ with $\boldsymbol{I}$ an identity matrix. Since $Y = A^T X$, the cost function may be reformulated as:

$$\Phi^b = A^T XM^b X^T A = A^T X\left(I - W^b\right)^T \left(I - W^b\right)X^T A, \tag{4}$$

The optimal mapping transformation.

The objective function of NPDE is to minimize Φ^w and maximize Φ^b at the same time, i.e.,

$$A_{opt} = \arg\min_{A} \frac{|\Phi^{w}|}{|\Phi^{b}|} = \arg\min_{A} \frac{|A^{T}\Psi^{w}A|}{|A^{T}\Psi^{b}A|}, \quad \textbf{(5)}$$

where $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ and $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$.

As it is known, orthogonality is very important to discriminant analysis, because physically meaningless features through orthogonal transformation can be linked back to the same number of variables of the measurement space. The orthogonal NPDE has been investigated in [29-30]. Due to the importance of the orthogonality with respect to the discriminant analysis, some basic properties and crucial problems of the orthogonal discriminant analysis (ODA) methods have been explored and solved in [31]. Hence, the NPDE criterion may be defined by enforcing the projection matrix ***A*** in equation (5) to be orthogonal:

$$\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}. \quad \textbf{(6)}$$

According to the Rayleigh quotient and Lagrange multipliers method, the optimal mapping transformation vector may be obtained by solving the following generalized eigenvalue problem:

$$\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}. \quad \textbf{(7)}$$

Let the column vectors $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ be the solutions of equation (7), ordered according to their eigenvalues, $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$, and then the matrix $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ is the transformation matrix.

The detail about NPDE can be found in [10].

3 Exponential Neighborhood Preserving Discriminant Embedding (ENPDE)

3.1 Matrix Exponential

In this section, we firstly introduce the following definition and some properties of matrix exponential [32]. The matrix exponential is widely used in applications such as control theory, and Markov chain analysis. Given an $n\times n$ square matrix ***A***, its exponential is defined as:

$$\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}, \quad \textbf{(8)}$$

where ***I*** is the identity matrix. The properties of matrix exponential are listed as follows:

(1) $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ is a finite matrix.

(2) $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ is a full-rank matrix.

(3) If square matrix ***A*** commutes with the matrix ***B***, i.e., ***AB*** = ***BA***, then

$$\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}. \quad \textbf{(9)}$$

(4) For an arbitrary square matrix ***A***, there exists the inverse of $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$. This is given by

$$\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}. \quad \textbf{(10)}$$

(5) If ***T*** is a nonsingular matrix, then

$$\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}. \quad \textbf{(11)}$$

(6) If $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ are eigenvectors of ***A*** that correspond to the eigenvalues $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$, then $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ are
also eigenvectors of $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ that correspond to the eigenvalues $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$.

3.2 The SSS Problem of NPDE

About the ranks of the matrices $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ and $\Theta:\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ presented in section 2, the following conclusion holds:

Theorem 1. Let *N* be the number of samples, and *D* is dimension of the samples, if $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$, then the ranks of the matrices $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ and $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ are at most *N*–1, i.e. $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ and $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$.
Proof. According to the section 2, we have

$$\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}.$$

Note that is $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ an *N*-order matrix, and the row elements of the matrix $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ are with constraint $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$, so it easy to know that the rank of the matrix $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ is at most *N*–1. It is known that the maximum possible rank of the product of two matrices is smaller than or equal to the smaller of the ranks of the two matrices. And so the rank of the matrix $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ is at most *N*–1, and then the rank of $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ is at most *N*–1, i.e.,

$$\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}.$$

Similarly, it is easy to know that the rank of $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ is at most *N*–1, i.e.,

$$\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}.$$

Note that $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ and $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ are *D*-dimension matrices. According to Theorem 1, both the matrices $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ and $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ can be singular. This stems from the fact that, in most cases, the number of samples is much smaller than the dimension of the samples, i.e., $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$. This is known as the small-sample-size (SSS) problem and NPDE suffers from the difficulty.

3.3 ENPDE

According to the above Theorem 1, both the matrices $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ and $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ can be singular in most cases, it is from fact that there are some 0 eigenvalues in the matrices $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ and $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$. The exponential NPDE (ENPDE) is proposed to address the problem.

According to the section 2, the objective function of NPDE is equation (5):

$$A_{opt}=\arg\min_{A}\frac{\left|\Phi^{w}\right|}{\left|\Phi^{b}\right|}=\arg\min_{A}\frac{\left|A^{T}\Psi^{w}A\right|}{\left|A^{T}\Psi^{b}A\right|}.$$

In ENPDE, we replaces the matrices $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ and $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ with the matrix exponential $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ and $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$, i.e.,

$$\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}. \quad \textbf{(12)}$$

Furthermore, if the projection matrix ***A*** is imposed to be orthogonal, the equation (12) may be formulated as:

$$\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}. \quad \textbf{(13)}$$

According to the Rayleigh quotient and Lagrange multipliers method, the optimal mapping transformation vector may be obtained by solving the following generalized eigenvalue problem:

$$\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}. \quad \textbf{(14)}$$

And then, we select *d* generalized eigenvectors $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$ of the equation (14), associated with the largest *d* eigenvalues of the equation (14) and ordered according to eigenvalues, i.e., $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$. Let $\Theta:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^{n\times n}$, namely, the matrix ***A*** is composed of the *d* largest generalized eigenvectors, and then ***A*** is the optimal transformation matrix.

4 Theoretical Analysis of ENPDE

4.1 Solving the SSS Problem

From the section 3.2, NPDE method suffers from the SSS problem. To overcome the complication of singular matrices, PCA can be adopted to reduce the dimensionality of the original image to *r* dimension. So that the resulting matrix $\Psi_r^w = X_r M^w X_r^{\ T}$ and $\Psi_r^b = X_r M^b X_r^{\ T}$ are nonsingular, and then the optimal mapping transformation vector may be obtained by solving the following generalized eigenvalue problem:

$$\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}. \quad \textbf{(15)}$$

However, the null spaces of the matrices $\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ and $\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$, in which some of the eigenvalues of $\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ and $\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ are equal to 0, are often be discarded. The fact is from that the data $\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ instead of the original data $\boldsymbol{X}$ is used. But the null spaces may contain some discriminant information. Consequently, some of significant information, which is contained in the original data, may be lost in the low-dimensional mapped data $\boldsymbol{y}$, due to the processing PCA step.

Based on the section 3.3, ENPDE method sets the optical projection axes $\boldsymbol{a}_i$ to the eigenvectors of the following generalized eigenvalue problem, i.e., equation (14):

$$\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}.$$

Obviously, the matrices $\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ and $\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ are full-rank matrices. Furthermore, all the information that is contained in $\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ and $\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ can be extracted by the ENPDE method. Even when the SSS problem occurs, the information contained in the null-space of $\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ and $\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ will not be discarded. And so, we can believe that the ENPDE method has more discriminant power than NPDE.

4.2 Distance Diffusion Mapping

In fact, for ENPDE, there exists a non-linear mapping function $\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ such that the neighborhood preserving scatter matrices are mapped into a new space, i.e.,

$$\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}, \quad \textbf{(16)}$$

$$\boldsymbol{\Psi}^b \to \Theta\left(\boldsymbol{\Psi}^b\right) = \exp\left(\boldsymbol{\Psi}^b\right), \quad \textbf{(17)}$$

$$\boldsymbol{\Psi}^w \to \Theta\left(\boldsymbol{\Psi}^w\right) = \exp\left(\boldsymbol{\Psi}^w\right). \quad \textbf{(18)}$$

Recall that NPDE finds an optimal projection by simultaneously maximizing the between-class neighborhood cost function Φ^b and minimizing the within-class neighborhood cost function Φ^w in local neighborhoods. Note that Φ^b means that between-class scatter in the neighborhood, and Φ^w means that within-class scatter in the neighborhood. Using simple matrix algebra, the between-class neighborhood cost function Φ^b becomes:

$$\begin{aligned}\Phi^b &= \sum_i \left| \boldsymbol{y}_i - \sum_j w_{ij}^b \boldsymbol{y}_j \right|^2 = \sum_i \left(\left(\boldsymbol{A}^T \boldsymbol{x}_i\right) - \sum_j w_{ij}^b \left(\boldsymbol{A}^T \boldsymbol{x}_j\right) \right)^T \left(\left(\boldsymbol{A}^T \boldsymbol{x}_i\right) - \sum_j w_{ij}^b \left(\boldsymbol{A}^T \boldsymbol{x}_j\right) \right) \\ &= \sum_i \left(\boldsymbol{x}_i - \sum_j w_{ij}^b \boldsymbol{x}_j \right)^T \boldsymbol{A}\boldsymbol{A}^T \left(\boldsymbol{x}_i - \sum_j w_{ij}^b \boldsymbol{x}_j \right). \end{aligned} \quad \textbf{(19)}$$

From section 2, optimal transformation matrix $\boldsymbol{A}$ may be imposed the constraint condition that the $\boldsymbol{A}$ an orthogonal matrix, i.e., $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{I}$, then the equation (19) becomes:

$$\Phi^b = \sum_i \left(\boldsymbol{x}_i - \sum_j w_{ij}^b \boldsymbol{x}_j \right)^T \left(\boldsymbol{x}_i - \sum_j w_{ij}^b \boldsymbol{x}_j \right) = tr\left(\boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{W}^b\right)^T \left(\boldsymbol{I} - \boldsymbol{W}^b\right)\boldsymbol{X}^T \right) = tr\left(\boldsymbol{\Psi}^b\right), \quad \textbf{(20)}$$

where $tr(\cdot)$ is the trace of a matrix.

Similarly, the within-class neighborhood cost function Φ^w becomes:

$$\Phi^w = \sum_i \left| \boldsymbol{y}_i - \sum_j w_{ij}^w \boldsymbol{y}_j \right|^2 = tr\left(\boldsymbol{X}\left(\boldsymbol{I}-\boldsymbol{W}^w\right)^T \left(\boldsymbol{I}-\boldsymbol{W}^w\right)\boldsymbol{X}^T \right) = tr\left(\boldsymbol{\Psi}^w\right). \quad \textbf{(21)}$$

Since the neighborhood preserving scatter matrices $\boldsymbol{\Psi}^b$ and $\boldsymbol{\Psi}^w$ are symmetric and positive semi-definite, they can be written as $\boldsymbol{\Psi}^b = \boldsymbol{\varphi}_b^T \boldsymbol{\Lambda}_b \boldsymbol{\varphi}_b$ and $\boldsymbol{\Psi}^w = \boldsymbol{\varphi}_w^T \boldsymbol{\Lambda}_w \boldsymbol{\varphi}_w$. The matrices $\boldsymbol{\varphi}_b$ and $\boldsymbol{\varphi}_w$ are two orthogonal matrices, which contain the eigenvectors of the matrices $\boldsymbol{\Psi}^b$ and $\boldsymbol{\Psi}^w$ respectively. The matrices:

$$\begin{aligned} \boldsymbol{\Lambda}_b &= diag\left(\lambda_{b1}, \lambda_{b2}, \cdots, \lambda_{bD}\right), \\ \boldsymbol{\Lambda}_w &= diag\left(\lambda_{w1}, \lambda_{w2}, \cdots, \lambda_{wD}\right), \end{aligned} \quad \textbf{(22)}$$

are two diagonal matrices, where $\lambda_{bi}\left(i=1,2,\cdots,D\right)$ and $\lambda_{wi}\left(i=1,2,\cdots,D\right)$ are the eigenvalues of the matrices $\boldsymbol{\Psi}^b$ and $\boldsymbol{\Psi}^w$ respectively. And then, the between-class neighborhood cost function Φ^b and the within-class neighborhood cost function Φ^w become:

$$\Phi^b = \sum_i \left| \boldsymbol{y}_i - \sum_j w_{ij}^b \boldsymbol{y}_j \right|^2 = tr\left(\boldsymbol{\Psi}^b\right) = \lambda_{b1} + \lambda_{b2} + \cdots + \lambda_{bD}, \quad \textbf{(23)}$$

and

$$\Phi^w = \sum_i \left| \boldsymbol{y}_i - \sum_j w_{ij}^w \boldsymbol{y}_j \right|^2 = tr\left(\boldsymbol{\Psi}^w\right) = \lambda_{w1} + \lambda_{w2} + \cdots + \lambda_{wD}. \quad \textbf{(24)}$$

Note that $\boldsymbol{\Psi}^b$ and $\boldsymbol{\Psi}^w$ are semi-definite matrices, all of the eigenvalues are positive or equal to zero. By using the mappings, i.e., equation (17) and equation (18):

$$\boldsymbol{\Psi}^b \to \Theta\left(\boldsymbol{\Psi}^b\right) = \exp\left(\boldsymbol{\Psi}^b\right),$$

and

$$\boldsymbol{\Psi}^w \to \Theta\left(\boldsymbol{\Psi}^w\right) = \exp\left(\boldsymbol{\Psi}^w\right),$$

there will be an implicit mapping of samples, in the same input space, such that the cost function Φ^b and Φ^w are replaced by Φ_e^b and Φ_e^w,

$$\Phi_e^b = tr\left(\exp\left(\boldsymbol{\Psi}^b\right)\right) = e^{\lambda_{b1}} + e^{\lambda_{b2}} + \cdots + e^{\lambda_{bD}}, \quad \textbf{(25)}$$

$$\Phi_e^w = tr\left(\exp\left(\boldsymbol{\Psi}^w\right)\right) = e^{\lambda_{w1}} + e^{\lambda_{w2}} + \cdots + e^{\lambda_{wD}}. \quad \textbf{(26)}$$

In general, the distance between samples in different classes is bigger than the related distance between samples in the same class in the neighborhood, we have $\Phi^b > \Phi^w$. So, for most of the eigenvalues in equation (23) and equation (24), we can have the inequality $\lambda_{bi} > \lambda_{wi}$. And then we have $e^{\lambda_{bi}} > e^{\lambda_{wi}}$. So we have

$$\frac{e^{\lambda_{bi}}}{e^{\lambda_{wi}}} > \frac{\lambda_{bi}}{\lambda_{wi}}. \quad \textbf{(27)}$$

As shown, the non-linear mapping function Θ has the effect of distance diffusion. And the diffusion scale to the between-class distance is larger than that to the within-class distance in the neighborhood. Hence, the distances between different class samples in all the neighborhoods are enlarged. This is what we want for getting good discrimination.

4.3 Comparison with the Other Relevant Works

In this section, we compare the proposed ENPDE method with the other relevant works: (1) the classical and global dimensionality reduction methods, PCA [1], LDA [2] and EDA [20]. (2) The unsupervised and supervised neighborhood preserving embedding methods, including NPE [9], the exponential NPE (ENPE) [24] and NPDE [10]. (3) The recently proposed two-dimension NPDE methods, 2DDNPE [14] and B2DNPDE [15]. The feature extraction of the two methods are based on the two-dimension image matrices, not image vectors, so they avoided the SSS problem of NPDE effectively. The proposed ENPDE method is order to resolve the SSS problem of the NPDE method. And so, the two two-dimension methods are selected to compare with ENPDE.

PCA and LDA are global method and can only preserve the global structures of the image samples. As manifold reduced dimensionality technique, LPP, NPE, ENPE, NPDE and ENPDE can uncover the essential manifold structure of the image. These manifold algorithms have the better classification power, which have been proved in [8-10, 24].

The NPE and the recently proposed ENPE are unsupervised technique. NPDE and ENPDE are supervised technique, in which the class-specific information of data is used, so the classification performance of the proposed ENPDE is better than that of NPE and ENPE. From the sections 4.1 and 4.2, on the one hand, the proposed ENPDE method addressed the SSS problems of NPDE methods effectively, more information can be extracted by ENPDE in contrast to NPDE. On the other hand, importantly, with the matrix exponential transformation, ENPDE has the effect of distance diffusion, i.e., the distance between samples belonging to different classes in the neighborhood is enlarged. It is helpful for improving the performance of classification. And so, we can believe that ENPDE will show advantageous performance over NPDE.

The two-dimension NPDE methods, 2DDNPE and B2DNPDE, extract the image feature directly from image matrices and then avoid the SSS problem of NPDE. The ENPDE method not only avoids the SSS problem, but also has distance diffusion effect, and then the discrimination property is largely emphasized. And so, ENPDE will show better discrimination performance than 2DDNPE and B2DNPDE, which is proved in the experiment.

5 Experiment Results and Discussion

In this section, we evaluate the face recognition performance of the proposed ENPDE method. The experiments are made on the three face image databases: Yale [33], CMU-PIE [34] and AR [35].

5.1 Yale Face Database

The Yale face database was taken from the Yale Center for Computational Vision and Control. It contains 165 gray scale images of 15 individuals. The images demonstrate variation with the following expressions or configurations: (1) lighting (i.e., center light, left light, and right light); (2) with/without glasses; and (3) facial expressions (i.e., normal, happy, sad, sleepy, surprised, and winking). The original image size is 320×243 pixels. Some sample images of one subject from Yale face database are shown in the following Fig. 1. In this experiment, all images are aligned based on eye coordinates and are cropped and scaled to 24×24.

Fig. 1. Some sample images of one individual from Yale face database

5.2 CMU-PIE Face Database

The CMU-PIE face database contains 68 subjects, with a total of 41368 face images. The face images were captured across 13 different poses, under 43 different illumination conditions, and 4 different expressions. In our experiment, the subset C09 is chosen for testing, and each subject has 24 frontal face images that were taken under variations in pose, illumination, and expression. Some sample images of one subject from CMU–PIE face database are shown in the following Fig. 2. In our experiment, all face images are scaled to 28×28 pixels.

Fig. 2. Some sample images of one individual from PIE face database

5.3 AR Face Database

AR face database was created by Aleix Martinez and Robert Benavente in the Computer Vision Center (CVC) at the U.A.B. It contains over 4000 color images corresponding to 126 people's faces (70 men and 56 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions. The pictures were taken at the CVC under strictly controlled conditions. Each person participated in two sessions, separated by two weeks (14 days) time. The same pictures were taken in both sessions. As an example, some sample images of one individual from AR face database are shown in Fig. 3. In our experiment, each image is manually cropped and resized to 40×50 pixels.

Fig. 3. Some sample images of one individual from AR face database

From the analysis in section 4.3, in our experiments, the proposed ENPDE method is compared with the PCA, LDA, EDA, NPE, the exponential NPE (ENPE), NPDE, and the two-dimension NPDE methods, including 2DDNPE [14], B2DNPDE [15].

For the methods suffering from the SSS problem (LDA and NPDE), PCA technique is firstly used to reduce the dimension of the original image vector to avoid the singularity of the matrix. Where, we reserve 98% of the principal components in the PCA stage.

A random subset with p images for each individual is taken to form the training set, and the remaining images are used as the testing set.

For the PCA, LDA and EDA, when the training sample p and the subspace dimension are fixed, the average value of the 20 recognition accuracies, from the 20 times random splits, is regarded as the recognition ratio of the corresponding method in this case.

For the neighborhood preserving discriminant embedding methods, NPE, ENPE, NPDE and ENPDE, for a split, when the training sample p and the subspace dimension is fixed, the neighborhood parameter k is searched from $\{2,3,\cdots,N-1\}$ and with step size=10, where N is the training sample number. For convenience, denote $m=\left\lceil \frac{N-1}{10} \right\rceil+1$. There are m recognition accuracies corresponding to the m values of k for each split. For each random split, we report the top-1 recognition accuracy from the best parameter k configuration. And so, there are 20 maximal recognition accuracies, and then we get the average value of the 20 maximal recognition accuracies and regard it as the recognition rate of the corresponding method in this case.

Let the face image size is r×s. For the two-dimension methods 2DDNPE and B2DNPDE, the size of feature matrices reduced subspace are $r\times d$ and $d\times d$ respectively, where d is the number of selected eigenvectors. Similarly, when the training sample p and the subspace dimension d is fixed, the 20 times random train sample splits are made, and then the average value of the 20 recognition accuracies is regarded as the recognition rate of the corresponding method in this case.

In general, the recognition performance varies with the dimension of the face subspace. In the experiment, let the subspace dimension d is from a range of dimensions with the step size=10, i.e. from d=10, 20, $\cdots$, d_{last}. For every subspace dimension, the above process is repeated to calculate the recognition rate. Note that, for some databases, NPDE do not have a large number of eigenvectors. It is from the fact that they are suffered from the SSS problem. The maximum number of eigenvectors of these methods is depended on the retained fixed principal components of the PCA stage. This is related to the database and the training sample number. And LDA method may get the maximal C-1 subspace dimension, where C is the sample class numbers. But for the matrix exponential methods, including EDA and ENPDE, there is no intrinsic subspace dimensionality limitation.

The best average performance obtained by the above methods as well as the corresponding dimension is summarized in Table 1 to Table 3. The experiment results in Yale face database are illustrated in Table 1, where the training sample number is 5, 6, 7. The experiment results in CMU-PIE face database are illustrated in Table 2, where the training sample number is 8, 9, 10. The experiment results in AR face database are illustrated in Table 3, where the training sample number is 5, 6, 7. The number appearing in parenthesis is the optimal subspace dimension.

Table 1. Best average recognition accuracy on Yale database over 20 random splits. The number appearing in parenthesis corresponds to the optimal subspace dimensionality

Method	5 trains(%)	6 trains(%)	7 trains(%)
PCA	79.00(50)	80.67(90)	86.17(50)
LDA	80.56(14)	80.53(14)	86.40(14)
EDA	80.89(50)	84.40(90)	86.50(50)
NPE	82.59(30)	87.56(53)	85.00(40)
ENPE	83.67(30)	88.78(60)	87.00(50)
2DDNPE	84.00(24×5)	89.00(24×6)	90.25(24×8)
B2DNPDE	86.28(13×13)	88.73(11×11)	87.00(9×9)
NPDE	87.41(30)	88.89(50)	90.55(40)
ENPDE	88.52(30)	90.67(80)	92.44(40)

Table 2. Best average recognition accuracy on PIE face database over 20 random splits. The number appearing in parenthesis corresponds to the optimal subspace dimensionality

Method	8 trains(%)	9 trains(%)	10 trains(%)
PCA	79.18(80)	79.61(80)	84.43(90)
LDA	95.75(10)	94.96(10)	88.04(50)
EDA	82.37(80)	84.96(80)	89.39(50)
NPE	78.80(70)	74.12(50)	79.65(60)
ENPE	79.98(60)	76.32(70)	81.67(60)
2DDNPE	86.21(28×6)	87.33(28×9)	89.35(28×12)
B2DNPDE	88.23(12×12)	90.45(10×10)	91.37(11×11)
NPDE	91.95(72)	93.35(50)	93.55(74)
ENPDE	93.55(80)	96.27(70)	95.77(90)

Table 3. Best average recognition accuracy on AR face database over 20 random splits. The number appearing in parenthesis corresponds to the optimal subspace dimensionality

Method	5 trains(%)	6 trains(%)	7 trains(%)
PCA	88.08(100)	88.14(90)	90.50(100)
LDA	92.33(100)	94.70(70)	98.72(50)
EDA	93.39(100)	96.65(40)	98.07(50)
NPE	91.57(90)	92.08(70)	91.78(100)
ENPE	92.68(90)	94.01(80)	93.38(90)
2DDNPE	94.31(40×26)	95.53(40×29)	95.36(40×22)
B2DNPDE	94.23(50×50)	96.65(45×45)	96.37(47×47)
NPDE	97.78(90)	98.13(90)	98.07(100)
ENPDE	98.80(30)	99.17(90)	99.17(90)

The recognition rates (versus dimension) of the supervised methods, including LDA, EDA, NPDE and the proposed ENPDE are shown in Fig. 4 to Fig. 6. Where, 2DDNPE and B2DNPDE are also supervised methods, but the feature extraction of the two methods are based on the image matrices and then the subspaces are two-dimension, so the two methods are not compared with the other supervised methods in the form of curves. Fig. 4 is for Yale face database, where the training sample number is six. Fig. 5 is for

CMU-PIE face database, where the training sample number is ten. Fig. 6 is for AR face database, where the training sample number is five. These plots are the average over 20 random splits.

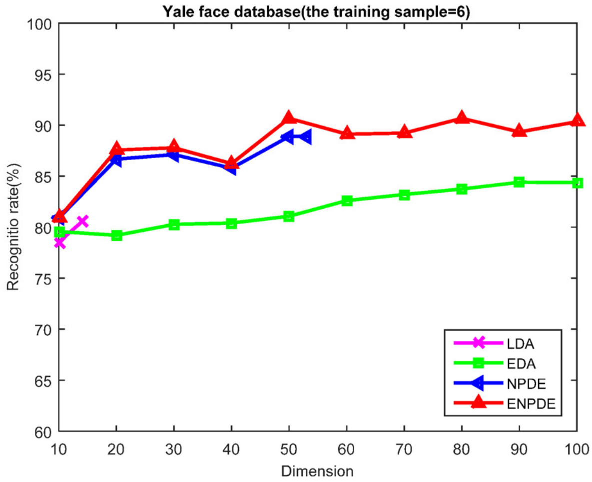

Fig. 4. Recognition accuracy (in percent) of the supervised methods versus the projected dimensions on the Yale face database (six training images)

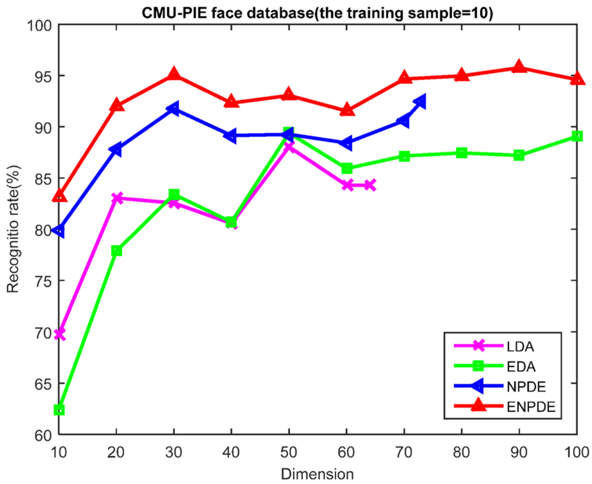

Fig. 5. Recognition accuracy (in percent) of the supervised methods versus the projected dimensions on the CMU-PIE face database (ten training images)

From the experiment results, we can observe that: (1) compared with NPDE and the other methods, ENPDE has better performance over all subspace dimensions in all the database; (2) in the different face database, ENPDE shows stable and robust in the all cases. This is what we want in face recognition; (3) the exponential methods, including the ENPDE, ENPE and EDA, have no dimensionality limit. However, LDA and NPDE have to suffer from the SSS problem, so PCA step has to be used to reduce the dimensionality of the original data and then the subspace dimension has to be limited, which are shown in Fig. 4 and Fig. 5. In AR face database, because the face image class is large, this problem is not shown in the range of 10-100 subspace dimensions.

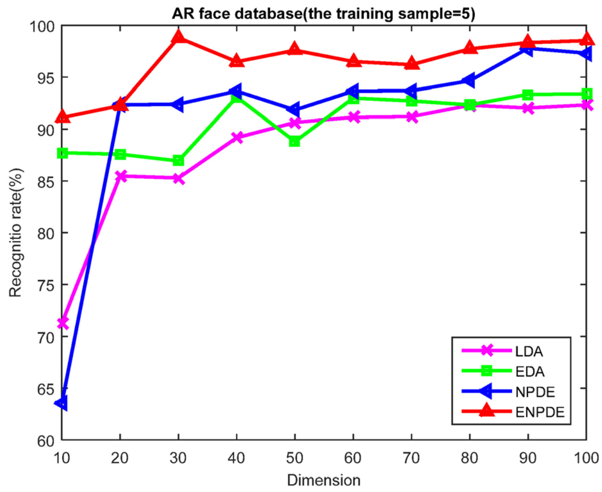

Fig. 6. Recognition accuracy (in percent) of the supervised methods versus the projected dimensions on the AR face database (five training images)

6 Conclusions

In this paper, the small-size-sample (SSS) problem of NPDE and its variant version has been investigated. A general exponential neighborhood preserving discriminant embedding (ENPDE) is proposed to improve NPDE method. The main idea of ENPDE is that the matrix exponential is introduced to NPDE. ENPDE has two superiorities: (1) unlike the NPDE method, the ENPDE method avoids the SSS problem, and it can extract more discriminant information. (2) ENPDE has the effect of distance diffusion mapping. With the help of distance diffusion mapping, the margin between samples belonging to different classes in the neighborhood is enlarged, which is helpful in improving the performance of classification. The experiments are conducted on three face databases: Yale, CMU-PIE and AR. In the experiments, ENPDE is compared with the global methods, including PCA, LDA, EDA, and the unsupervised and supervised neighborhood preserving embedding methods, including NPE, ENPE, NPDE, and the two-dimension NPDE methods, including 2DDNPE, B2DNPDE method. The experiment results have validated the effectiveness of the propose ENPDE method and proved that ENPDE has advantageous performance over the above methods in face recognition. The limitation of the ENPDE method is that the solution of the generalized eigenvalue problem of matrix exponential needs larger computation cost because of the characteristics of matrix exponential. So, it is expected that the better algorithm about the calculation of matrix exponential is proposed. The ENPDE method may be further researched. For example, a general exponential base $a(a>e)$, not the Euler number *e*, may be selected to build the matrix exponential. In fact, based on the property of the exponential function, the exponential base is larger, the distance diffusion mapping of the exponential NPDE method is stronger, and then the better classification performance may be gotten.

Acknowledgments

This research was supported by National Nature Science Foundation of China (61703062, 61703062), the Project of Natural Science Foundation Project of CQ CSTC of China (cstc2016jcyjA0419, cstc2017jcyjAX0316) and the School Fund Project of CQNU (16XLB006, 16XZH07).

References

[1] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3(1)(1991) 71-86.

[2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7)(1997) 711-720.

[3] S.T. Roweis, L.K. Saul, Nonlinear dimension reduction by locally linear embedding, Science 290(5500)(2000) 2323-2326.

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

[5] Z. Zhang, H. Zha, Principal manifolds and nonlinear dimension reduction via local tangent space alignment, SIAM Journal on Scientific Computing 26(1)(2005) 313-338.

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

[7] Y. Bengio, J. Paiement, P. Vincent, O. Delalleau, N. Roux, M. Ouimet, Out-of sample extensions for LLE, ISOMAP, MDS, Eigenmaps, and spectral clustering, in: Proc. Advances in Neural Information Processing Systems (NIPS), 2004.

[8] X.F. He, P. Niyogi, Locality preserving projections, in: Proc. Advances in Neural Information Processing Systems (NIPS), 2004.

[9] X.F. He, D. Cai, S. Yan, H. Zhang, Neighborhood preserving embedding, in: Proc. International Conference on Computer Vision (ICCV), 2005.

[10] P.Y. Han, A.T.B. Jin, F.S. Abas, Neighbourhood preserving discriminant embedding in face recognition, J. Vis. Commun. Image R 9(2009) 532-542.

[11] A.T.B. Jin, P.Y. Han., F.S. Abas, Face recognition based on neighborhood discriminant preserving embedding, in: Proc. 2008 10th Intl. Conf. on Control, Automation, Robotics and Vision (ICARCV), 2008.

[12] Y.M. Tian, Y.E. Yun, T.J. Ma, Discriminant neighborhood preserving embedding algorithm, Journal of Xidian University 38(3)(2011) 24-28.

[13] H. Du, X. Chai, F. Wang, F. Zhang, Face recognition method using neighborhood preserving discriminant embedding, Chinese Journal of Scientific Instrument 31(2010) 625-629.

[14] M. Pang, J.F. Jiang, C. Lin, B.H. Wang, Two dimensional discriminant neighborhood preserving embedding in face recognition, in: Proc. the 6th International Conference on Graphic and Image Processing(ICGIP), 2014.

[15] J.Z. Liang, C. Chen, Y.F. Yi, X.X. Xu, M. Ding, Bilateral two-dimensional neighborhood preserving discriminant embedding for face recognition, IEEE Access 5(2017) 17201-17212.

[16] Y.F. Yang, Y.P. Wang, X.S. Xue, Discriminant spare locality preserving projection for face recognition, Multimed Tool Appl. 76(2017) 2697-2712.

[17] B.W. Fang, Z.Q. Huang, Y. Li, Y. Wang, v-Support vector machine based on discriminant sparse neighborhood preserving embedding, Pattern Anal. Appl. 20(2017) 1077-1089.

[18] W.Q. Wang, V. Aggarwal, S. Aeron, Tensor train neighborhood preserving embedding, IEEE Trans. Signal Processing 66(10)(2018) 2724-2732.

[19] R.B. Huang, C. Liu, J.L. Zhou, Discriminant analysis via jointly L2,1-norm sparse tensor preserving embedding for image classification, J. Vis. Commun. Image R. 47(2017) 10-22.

[20] T.P. Zhang, B. Fang, Y. Tang, Z. Shang, B. Xu, Generalized discriminant analysis: a matrix exponential approach, IEEE Trans. System, Man, Cybern. B, Cybern. 40(1)(2010) 186-197.

[21] W.W. Yu, X.L. Teng, C.Q. Liu, Face recognition using discriminant locality preserving projections, Image Vis. Comput 24(2006) 239-248.

[22] H. Chen, H. Chang, T. Liu, Local discriminant embedding and its variants, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.(CVPR), 2005.

[23] H. Huang, J. Liu, Y. Pan, Semi-supervised marginal fisher analysis for hyperspectral image classification, ISPRS Ann. Photogrammetry-Remote Sens. Spatial Inf. Sci. 1(2012) 3.

[24] R.S. Ran, B. Fang, X.G. Wu, Exponential neighborhood preserving embedding for face recognition, IEICE Trans Info. Syst. E101-D (5)(2018) 1410-1420.

[25] S.Wang, H. Chen, X. Peng, C. Zhou, Exponential locality preserving projections for small sample size problem, Neurocomputing 74(17)(2011) 3654-3662.

[26] S.C. Huang, L. Zhuang, Exponential discriminant locality preserving projection for face recognition, Neurocomputing 208(2016) 373-377.

[27] F. Dornaika, A. Bosaghzadeh, Exponential local discriminant embedding and its application to face recognition, IEEE Trans on Cybernetics 43(3)(2013) 921-934.

[28] F. Dornaika, Y. El Traboulsi, Matrix exponential based semi-supervised discriminant embedding for image classification, Pattern Recognition 6(2017) 92-103.

[29] G.Q. Wang, N.F. Shi, X.J. Hou, Orthogonal neighborhood preserving discriminant projection for face recognition, in: the 2nd International Congress on Image and Signal Processing(CISP), 2009.

[30] S. Liu, Q.Q. Ruan, R.R. Ni, Orthogonal discriminant neighborhood preserving embedding for facial expression recognition, in: Proc. IEEE International Conference on Image Processing, 2010.

[31] Y. Wang, J.B. Xie, Y. Wu, Orthogonal discriminant analysis revisited, Pattern Recognition Letter 84(2016) 149-155.

[32] G.H. Golub, C.F.V. Loan, Matrix Computation, The Johns Hopkins Univ. Press, 1983.

[33] Database: Faces & Sketchs. <http://see.xidian.edu.cn/vipsl/database_Face.html>, 2006.

[34] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination, and expression database, IEEE Transactions on Pattern Analysis and Machine Intelligence 25(12)(2003) 1615-1618.

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