Matrix Exponential LPP for Face Recognition

Su-Jing Wang, Cheng-Cheng Jia, Hui-Ling Chen, Chun-Guang Zhou*
College of Computer Science and Technology, Jilin University, Changchun 130012, P. R. China

***Abstract*—Face recognition plays a important role in computer vision. Recent researches show that high dimensional face images lie on or close to a low dimensional manifold. LPP is a widely used manifold reduced dimensionality technique. But it suffers two problem: (1) Small Sample Size problem; (2)the performance is sensitive to the neighborhood size k. In order to address the problems, this paper proposed a Matrix Exponential LPP. To void the singular matrix, the proposed algorithm introduced the matrix exponential to obtain more valuable information for LPP. The experiments were conducted on two face database, Yale and Georgia Tech. And the results proved the performances of the proposed algorithm was better than that of LPP.**

I. Introduction

Automatic facial recognition has been a longstanding challenge in the field of computer vision and pattern recognition for several decades. A real face image usually has a high dimensional data. In order to deal with the high dimensional image data adequately, its dimensionality needs to be reduced. Dimensionality reduction is the transformation of high-dimensional data into a meaningful representation of reduced dimensionality. Principal Component Analysis (PCA)[1] and Linear Discriminant Analysis (LDA)[2] are two widely used techniques for reduced dimensionality. Recently, a number of research efforts have shown that the high dimensional image information in real world lies on or is close to a smooth nonlinear low dimensional manifold. However, both PCA and LDA fail to discover the underlying manifold structure, due to the fact that they aim only to preserve the global structures of the image samples. In order to uncover the essential manifold structure of the facial images, *laplacianfaces*[3] were obtained by using Locality Preserving Projections (LPP)[4] to preserve the local structure of image samples i.e.,the neighbor relationship between samples.

The neighbor relationship is measured by the artificial constructed adjacent graph. Usually, the most popular adjacent graph construction manner is based on the k nearest neighbor and $\epsilon-$lneighborhood criteria. Once a adjacent graph is constructed, the edge weights are assigned by various strategies such as 0-1 weights and heat kernel function. Unfortunately, such adjacent graph is artificially constructed in advance, and thus does not necessary uncover the intrinsic local geometric structure of the samples. To make things worse, the performance of LPP is seriously sensitive to the value of k. Even though the samples come from the same database and the numbers of each individual images are the same. To address the problem, some researches focus on how to construct the adjacent graph. Sample-dependent Graph [5] is constructed based on samples in question to determine neighbors of each sample and similarities between sample pairs, instead of predefining the same neighbor parameter k for all samples. Locally Discriminating Projection (LDP) [6] uses label information to construct the adjacent graph. Sparsity Preserving Projections (SPP) [7] aims to preserve the sparse reconstructive relationship of the samples, which is achieved by constructing the adjacent graph using a minimizing a L1 regularization-related objective function.

Another problem of LPP is the fact that it like LDA also suffers from the small sample size problem. This derives from that when the dimension of the sample is greater than the number of the samples, there is the singular of the matrices. To deal with the problem, *laplacianfaces*[3] uses PCA to reduce the dimension, and then applying the LPP. However, a potential problem is that the PCA criterion may not be compatible with the LPP criterion, thus the PCA step may discard the valuable information for LPP in the null space of $\mathbf{XLX}^T$. In order to address the issue, the Direct LPP [8] optimizes locality preserving criterion on high-dimensional images via simultaneously diagonalizing $\mathbf{XLX}^T$ and $\mathbf{XDX}^T$. Xu *et al.*[9] transforms $\mathbf{XLX}^T$ and $\mathbf{XDX}^T$ into the main space of $\mathbf{XDX}^T$, then find the optimal solution in the main space of $\mathbf{XDX}^T$. The above methods can extract at most $N-1$ dimensions feature vector.

To alleviate the above two problems of LPP: (1) the performance is sensitive to the value of k[10]; (2) the SSS problem, we propose a locality preserving projection based on matrix exponential. The rest of this paper is organized as follows: in Section II, we briefly review the LPP algorithm; in Section III, we give the background of matrix exponential, and introduce the Matrix Exponential LPP algorithm; in Section IV, the experiments are conducted on two public face databases: Yale and Georgia Tech face database, and the results are analyzed which show that the performance of Matrix Exponential LPP is better than that of LPP, especially for face database with different poses and cluttered background like the Georgia Tech face database; finally in Section V, conclusions are drawn.

II. Locality Preserving Projections

Given a set of N samples $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}, \mathbf{x}_i \in \mathbb{R}^D$, we attempt to find a transformation matrix $\mathbf{W}$ of size $D \times d$ to map: $\mathbf{y}_i = \mathbf{W}^T\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^d$, such that $\mathbf{y}_i$ easier to be distinguished in the projective subspace.

Locality Preserving Projections (LPP)[4] attempts to preserve the local structure of the samples in the low-dimensional projected subspace as much as possible. The local structure of the samples is measured by constructing the adjacency graph G. There are two ways to construct G: $\varepsilon-$ neighborhoods and

k nearest neighbors. The similarity matrix $\mathbf{S}$ is defined by the following two ways:

1) 0-1 ways

$$S_{ij} = \begin{cases} 1 & \text{nodes } i \text{ and } j \text{ are connected in } G \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

2) Heat kernel

$$S_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{2t^2}} & \text{nodes } i \text{ and } j \text{ are connected in } G \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

where t is a parameter that can be determined empirically. When t is large enough, $exp(-\|x_i - x_j\|^2/t) = 1$, heat kernel becomes 0-1 ways. Obviously, 0-1 ways is a special case of the heat kernel. In order to contain no any discriminant information, we do not use any label information to construct the similarity matrix $\mathbf{S}$. We hope that the criterion function incurs a heavy penalty if neighboring points $\mathbf{x}_i$ and $\mathbf{x}_j$ are mapped far apart. Therefore, minimizing it is an attempt to ensure that if $\mathbf{x}_i$ and $\mathbf{x}_j$ are close, then $\mathbf{y}_i$ and $\mathbf{y}_j$ are close, as well. That means to minimize:

$$\sum_{i,j} (\mathbf{y}_i - \mathbf{y}_j)^2 S_{ij} \quad (3)$$

A reasonable criterion function of LPP is as follows:

$$\min_{\mathbf{W}^T\mathbf{XDX}^T\mathbf{W}=\mathbf{I}} \mathbf{W}^T\mathbf{XLX}^T\mathbf{W} \quad (4)$$

where $\mathbf{D}$ is a diagonal matrix; its entries $D_{ii} = \sum_j S_{ij}$ measure the local density around $\mathbf{x}_i$. $\mathbf{L} = \mathbf{D} - \mathbf{S}$ is the Laplacian matrix. Finally, the transformation matrix consists of the eigenvectors associated with the smallest eigenvalues of the following generalized eigenvalue problem:

$$\mathbf{XLX}^T\mathbf{w} = \lambda\mathbf{XDX}^T\mathbf{w} \quad (5)$$

III. Matrix Exponential LPP

A. Matrix Exponential

In this section, the definition and properties of matrix exponential are introduced. Given an arbitrary $n \times n$ square matrix $\mathbf{A}$, its exponential is defined as follows:

$$\exp(\mathbf{A}) = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \cdots + \frac{\mathbf{A}^m}{m!} + \cdots \quad (6)$$

where I is a identity matrix with the size of $n \times n$. The properties of matrix exponential are listed as follows:

1) $\exp(\mathbf{A})$ is a finite matrix.
2) $\exp(\mathbf{A})$ is a full rank matrix
3) If matrix $\mathbf{A}$ commutes with $\mathbf{B}$, i.e.,$\mathbf{AB} = \mathbf{BA}$, then $\exp(\mathbf{A}+\mathbf{B}) = \exp(\mathbf{A})\exp(\mathbf{B})$.
4) If $\mathbf{B}$ is a nonsingular matrix, then $\exp(\mathbf{B}^{-1}\mathbf{AB}) = \mathbf{B}^{-1}\exp(\mathbf{A})\mathbf{B}$.
5) If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are eigenvectors of $\mathbf{A}$ that correspond to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are also eigenvectors of $\exp(\mathbf{A})$ that correspond to the eigenvalues $e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_n}$. It is also well known that the matrix is non-singular.

B. Matrix Exponential LPP

We define $\mathbf{S}_L = \mathbf{XLX}^T$ and $\mathbf{S}_D = \mathbf{XDX}^T$, and the eigen solution formulation of LPP (4) can be rewritten as follows:

$$\min_{\mathbf{W}^T\mathbf{S}_D\mathbf{W}=\mathbf{I}} \mathbf{W}^T\mathbf{S}_L\mathbf{W} \quad (7)$$

Theorem 1: Let D and N be the dimension of the sample and the number of the samples,respectively .If $D > N$, then the rank of $\mathbf{S}_L$ is at most $N-1$ and the rank of $\mathbf{S}_D$ is at most N.

Proof: According to the properties of the Laplacian matrix, it is easy-known that the determinant of $\mathbf{L}$ is 0. So, the rank of $\mathbf{L}$ 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. Hence, $rank(\mathbf{S}_L) = rank(\mathbf{XLX}^T) \leq N-1$. Similarly, we have $rank(\mathbf{S}_D) \leq N$. ■

From Theorem 1, LPP also suffers from the SSS problem, due to the fact that the matrix $\mathbf{S}_L$ is singular when the SSS problem incurs. We denote the eigenvectors of $\mathbf{S}_L$ as $\mathbf{V}_L = [\mathbf{v}_{L1}, \mathbf{v}_{L2}, \ldots, \mathbf{v}_{Ln}]$ that correspond to the eigenvalues $\mathbf{\Lambda}_L = diag(\lambda_{L1}, \lambda_{L2}, \ldots, \lambda_{Ln})$. Similarly, the eigenvectors of $\mathbf{S}_D$ are denoted as $\mathbf{V}_D = [\mathbf{v}_{D1}, \mathbf{v}_{D2}, \ldots, \mathbf{v}_{Dn}]$ that correspond to the eigenvalues $\mathbf{\Lambda}_D = diag(\lambda_{D1}, \lambda_{D2}, \ldots, \lambda_{Dn})$. The Eq. (7) can be rewritten as follows:

$$\min_{\mathbf{W}^T(\mathbf{V}_D\mathbf{\Lambda}_D\mathbf{V}_D^T)\mathbf{W}=\mathbf{I}} \mathbf{W}^T(\mathbf{V}_L\mathbf{\Lambda}_L\mathbf{V}_L^T)\mathbf{W} \quad (8)$$

The matrix $\mathbf{S}_L$ is not a singular, when the SSS problem incurs. In order to address the problem, the PCA is adopted to reduce the dimension of the feature space to $N-1$, before applying the standard LPP defined be Eq. (8). Unfortunately, the valuable information for LPP in the null space of $\mathbf{S}_L$ may also be discarded in the PCA step.To extract this kind of valuable information for LPP, we replace λ_{Li}, i.e., the eigenvalues of $\mathbf{S}_L$, by $\exp(\lambda_{Li})$ and λ_{Di}, i.e., the eigenvalues of $\mathbf{S}_D$, by $\exp(\lambda_{Di})$. Then, Eq. (8) is transformed into

$$\begin{aligned} &\min_{\mathbf{W}^T(\mathbf{V}_D\exp(\mathbf{\Lambda}_D)\mathbf{V}_D^T)\mathbf{W}=\mathbf{I}} \mathbf{W}^T(\mathbf{V}_L\exp(\mathbf{\Lambda}_L)\mathbf{V}_L^T)\mathbf{W} \\ &= \min_{\mathbf{W}^T\exp(\mathbf{S}_D)\mathbf{W}=\mathbf{I}} \mathbf{W}^T\exp(\mathbf{S}_L)\mathbf{W} \end{aligned} \quad (9)$$

The above equation is the criterion function of Matrix Exponential LPP. According to the properties of the matrix exponential, the $\exp(\mathbf{S}_L)$ is nonsingular. The valuable information for LPP in the null space of $\mathbf{S}_L$ can be extracted by Eq. (9).

IV. Experiments

A. Database and experimental set

We conducted the experiments on two well-known face databases Yale[1] and Georgia Tech face databases [2].

There are total of 165 gray scale images for 15 individuals where each individual has 11 images in Yale face database.

[1]http://cvc.yale.edu/projects/yalefaces/yalefaces.html
[2]http://www.anefian.com/research/face_reco.htm

Fig. 1. Sample images of one individual in the YALE database.

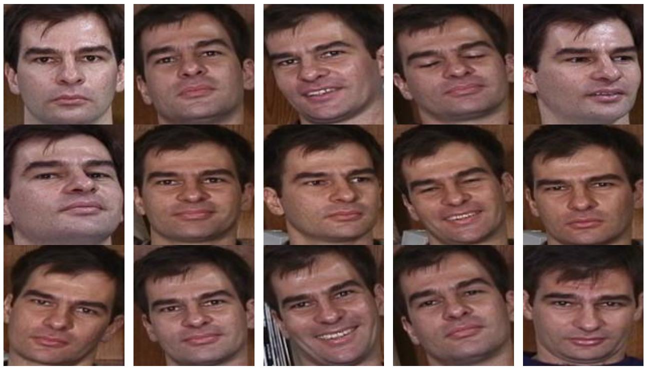


Fig. 2. Sample images of one individual from the Georgia Tech database (non-aligned head images).

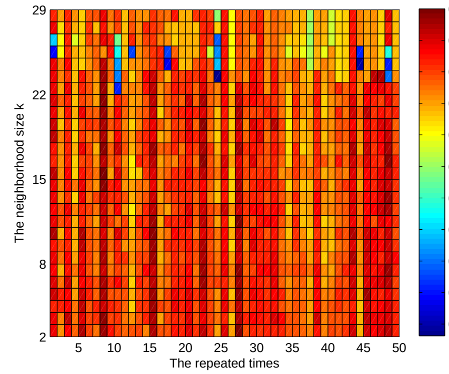
The images demonstrate variations in lighting condition, facial expression (normal, happy, sad, sleepy, surprised, and wink). The sample images of one individual from the Yale database are showed in Figure 1.

Georgia Tech face database contains images of 50 individuals taken in two or three sessions at different times. Each individual in the database is represented by 15 color JPEG images with cluttered background taken at resolution 640×480 pixels. The average size of the faces in these images is 150×150 pixels. The pictures show frontal and/or tilted faces with different facial expressions, lighting conditions and scale. Each image is manually cropped and resized to 32×32 pixels. The sample images for one individual of the Georgia Tech database are showed in Fig. 2.

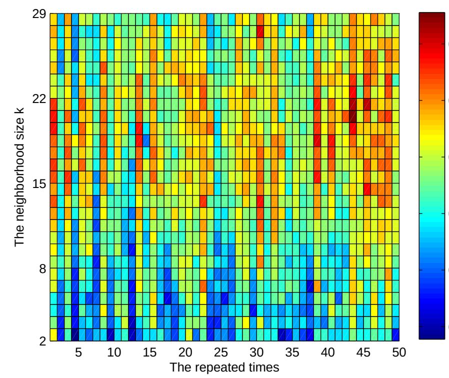
B. Experiments and results on the Yale database

The experiments are conducted on the Yale database. The similarity matrix $\mathbf{S}$ is defined by the heat kernel function. Empirically, the parameter t is set as the mean norm of the training set. The neighbors parameter k is searched from $\{2, 3, \ldots, N-1\}$. We randomly split the image samples so that p ($p = 2, 3, 4, 5, 6, 7, 8$) images for each individual are used as the training set and the rest are used as the testing set. This process is repeated 50 times. Fig. 3 plots the relationship between the performances of two algorithms and the neighborhood size k, when $p = 2$. The warmer color represents the better performance in the figure. Comparing the responding columns of Fig. 3(a) and Fig. 3(b), there's very little color difference in each column of Fig. 3(a). This means that the neighborhood size k is less sensitive to the performance of the proposed algorithm than that of LPP.

In order to investigate the performance of the proposed algorithm, we implement the proposed algorithm and LPP on Yale database. The results are also illustrated in Fig. 4. The solid lines denote that the neighborhood size k is searched from $\{2, 3, \ldots, N-1\}$. The dot-dash lines denote that the k



(a) Matrix Exponential LPP



(b) LPP

Fig. 3. The performances of two algorithms vs. the neighborhood size k on the Yale database.

is equal to 2. As is shown in Fig. 4, the performances of the proposed algorithm are better than these of LPP in two ranges of k. That stems from that the PCA before implementing LPP discards the valuable information for LPP. Whereas, there is not the PCA step in the proposed algorithm. It is interesting that when the training sample size is small, the performance of the proposed algorithm with $k = 2$ is better than that of LPP with $k \in \{2, 3, \ldots, N-1\}$. That also illustrates that the proposed algorithm is more effective than LPP for SSS problem. In Fig. 4, we also find that the margin between two green lines is much wider than that of red ones. This also proves the neighborhood size k is less sensitive to the performance of the proposed algorithm than that of LPP.

C. Experiments and results on the Georgia Tech face database

Georgia Tech face database is more complex than Yale database, because it contains various pose faces with different expressions on cluttered background. In this experiment, We randomly split the image samples so that p ($p = 2, 4, 6, 8, 10, 12$) images for each individual are used as the training set and the rest are used as the testing set. This process is repeated 30 times. Other setting is the same with Yale database. We plot the relationship between the performances and k in Fig. 5 ($p = 2$). In the figure, we not only see

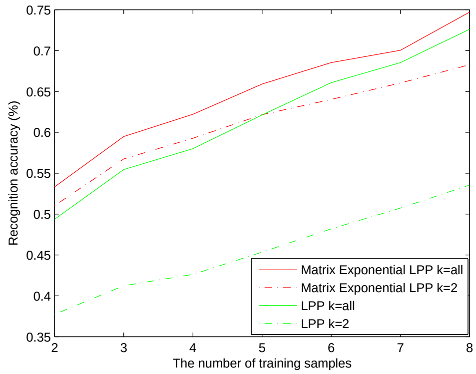


Fig. 4. The performances of two algorithms on the Yale database.

the properties in Fig. 3 but also see the fact of that the better performance of the proposed algorithm occur when k is small. The experimental results about the performances also illustrated in Fig. 6. As is shown in the figure, the properties in Fig. 4 can be seen. Moreover, we also find that the performance of the proposed algorithm with $k = 2$ is better than that of LPP with $k \in \{2, 3, \ldots, N-1\}$. Matrix Exponential LPP shows outstanding performance on complex face databases.

V. Conclusion

We have presented a new reduced dimensionality technique, which is named as Matrix Exponential LPP. It addressed the two problems of LPP: (1) Small Sample Size problem; (2)the performance is sensitive to the neighborhood size k. Matrix Exponential LPP avoids the singular of the matrices and obtains more valuable information for LPP. The experimental results prove the performances of Matrix Exponential LPP was better than that of LPP on two public face databases: Yale and Georgia Tech.

Acknowledgment

This paper is supported by (1)the National Natural Science Foundation of China under Grant No. 60873146, 60973092, 60903097, (2)project of science and technology innovation platform of computing and software science (985 engineering), (3)the Key Laboratory for Symbol Computation and Knowledge Engineering of the National Education Ministry of China.

References

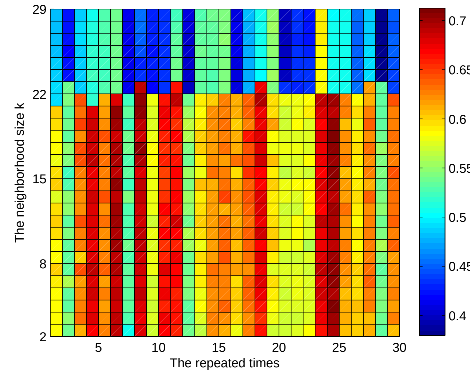
[1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Jan. 1991.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

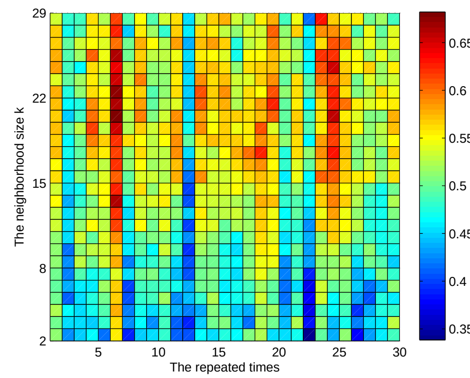
[3] X. F. He, S. C. Yan, Y. X. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.

[4] X. F. He and P. Niyogi, "Locality preserving projections," in *Advances in neural information processing systems*, vol. 16. The MIT Press, 2004, pp. 153–160.

[5] B. Yang and S. Chen, "Sample-dependent graph construction with application to dimensionality reduction," *Neurocomputing*, 2010.



(a) Matrix Exponential LPP



(b) LPP

Fig. 5. The performances of two algorithms vs. the neighborhood size k on the the Georgia Tech database.

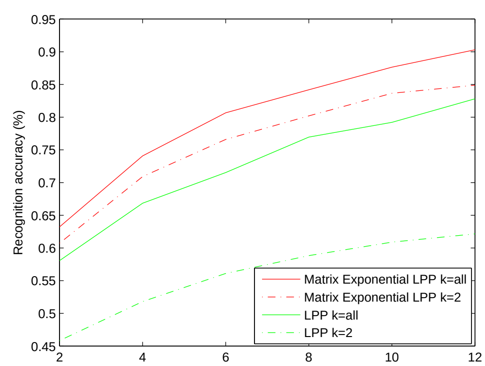


Fig. 6. The performances of two algorithms on the Georgia Tech database.

[6] H. Zhao, S. Sun, Z. Jing, and J. Yang, "Local structure based supervised feature extraction," *Pattern recognition*, vol. 39, no. 8, pp. 1546–1550, 2006.

[7] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, no. 1, pp. 331–341, 2010.

[8] G. Feng, D. Hu, and Z. Zhou, "A Direct Locality Preserving Projections (DLPP) Algorithm for Image Recognition," *Neural Processing Letters*, vol. 27, no. 3, pp. 247–255, 2008.

[9] Y. Xu, A. Zhong, J. Yang, and D. Zhang, "LPP solution schemes for use with face recognition," *Pattern Recognition*, 2010.

[10] S. Wang, N. Zhang, M. Sun, and C. Zhou, "The analysis of parameters t and k of lpp on several famous face databases," *Advances in Swarm Intelligence*, pp. 333–339, 2011.