The Analysis of Parameters t and k of LPP on Several Famous Face Databases

Sujing Wang, Na Zhang, Mingfang Sun, and Chunguang Zhou[*]

College of Computer Science and Technology,
Jilin University, Changchun 130012, China
{wangsj08, nazhang08}@mails.jlu.edu.cn; cgzhou@jlu.edu.cn

Abstract. The subspace transformation plays an important role in the face recognition. LPP, which is so-called the *Laplacianfaces*, is a very popular manifold subspace transformation for face recognition, and it aims to preserve the local structure of the samples. Recently, many variants of LPP are proposed. LPP is a baseline in their experiments. LPP uses the adjacent graph to preserve the local structure of the samples. In the original version of LPP, the local structure is determined by the parameters t (the heat kernel) and k (k-nearest neighbors) and directly influences on the performance of LPP. To the best of our knowledge, there is no report on the relation between the performance and these two parameters. The objective of this paper is to reveal this relation on several famous face databases, i.e. ORL, Yale and YaleB.

Keywords: locality preserving projection; the adjacent graph; the nearest neighbors; heat kernel; parameters set

1 Introduction

As one of the most important biometric techniques, face recognition has gained lots of attentions in pattern recognition and machine learning areas. The subspace transformation plays an important role in the face recognition. Feature extraction is one of the central issues for face recognition. Subspace transformation (ST) is often used as a feature extraction method. The idea of ST is to project the feature from the original high dimensional space to a low dimensional subspace, which is called *projective subspace*. In the projective subspace, the transformed feature is easier to be distinguished than the original one.

Principal Component Analysis (PCA)[12] is a widely used subspace transformation. It attempts to find the projective directions to maximize variance of the samples. To improve classification performance, LDA[1] encodes discriminant information by maximizing the ratio between the between-class and within-class scatters. LDA can be thought of as an extension with discriminant information of PCA. Both PCA and LDA focus on preserving the global structure of the samples. However, Seung[10] assumed that the high dimensional visual image information in the real world lies on or is close to a smooth low dimensional manifold.

---

[*] Corresponding author

Inspired by this idea, multiple manifold dimensionality reduction methods that preserve the local structure of samples have been proposed, such as ISOMAP[11], LLE[9], Laplacian Eigenmaps[2] etc. Locality Preserving Projections (LPP)[5] is a linear Laplacian Eigenmaps. Its performance is better than those of PCA and LDA for face recognition[6]. Recently, many variants[15][3][13][16][7][14] of LPP are proposed. LPP is a baseline in their experiments.

However, the performance of LPP depends mainly on its underlying adjacent graph whose construction suffers from the following points: (1)such adjacent graph is artificially constructed; (2) it is generally uneasy about assigning appropriate values for the neighborhood size k and heat kernel parameter t involved in graph construction. To the best of our knowledge, there is no report on the relation between the performance and these two parameters k and t. The objective of this paper is to reveal this relation on several famous face databases.

2 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)[5] 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

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

2. Heat kernel

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

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 $\mathbf{S}$. The criterion function of LPP is as follows:

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

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. Finally, the transformation

matrix consists of the eigenvectors associated with the smallest eigenvalues of the following generalized eigenvale problem:

$$\mathbf{XLX}^T\mathbf{w} = \lambda\mathbf{XDX}^T\mathbf{w} \tag{4}$$

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.

We define $\mathbf{S}_L = \mathbf{XLX}^T$ and $\mathbf{S}_D = \mathbf{XDX}^T$, and rewrite Eq. (4) as follows:

$$\mathbf{S}_L\mathbf{w} = \lambda\mathbf{S}_D\mathbf{w} \tag{5}$$

Theorem 1. *Let N and D be the dimension of the sample and the number of the samples,respectively .If $N > D$, 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 definition of the Laplacian matrix and the fact that the similarity matrix is symmetrical.

$$|\mathbf{L}| = \begin{vmatrix} \sum_j w_{1j} - w_{11} & -w_{12} & \cdots & -w_{1N} \\ -w_{12} & \sum_j w_{2j} - w_{22} & \cdots & -w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -w_{1N} & -w_{2N} & \cdots & \sum_j w_{Nj} - w_{NN} \end{vmatrix} \tag{6}$$

we add the 2rd, 3nd,... Nth rows into the 1st row, and obtain $|\mathbf{L}| = 0$. So, the rake 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 $rank(\mathbf{S}_L) \leq N$.

From Theorem 1, LPP like LDA also suffers from the SSS problem. Another problem is how to measure he local structure of the samples. LPP uses the similarity matrix $\mathbf{S}$. If every entries are the same, the local structure of the samples is not preserved. Without loss of generality, each entry in $\mathbf{S}$ is set as $1/N^2$, i.e., $\mathbf{L} = \frac{1}{N}\mathbf{I} - \frac{1}{N^2}\mathbf{ee}^T$, where $\mathbf{e}$ ia a vector, whose entries are 1. The matrix $\mathbf{S}_L$ is equivalent to the covariance matrix in PCA[6]. In this case, LPP degenerates into PCA. Obviously, the performance of LPP dependents on how construct the similarity matrix $\mathbf{S}$. In next section, the performance of LPP with respect to the neighborhood size k and heat kernel parameter t on several famous face databases will be reported.

3 Experiment

3.1 Database and experimental set

Three well-known face database ORL[1], Yale[2] and the Extended Yale Face Database B[4] (denoted by YaleB hereafter) were used in our experiments.

---

[1] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[2] http://cvc.yale.edu/projects/yalefaces/yalefaces.html

The ORL database collects images from 40 individuals, and 10 different images are captured for each individual. For each individual, the images with different facial expressions and details are obtained at different times. The face in the images may be rotated, scaled and be tilting in some degree. The sample images of one individual from the ORL database are shown in Figure 1.

Fig. 1. Sample images of one individual from the ORL database.

There are total of 165 gray scale images for 15 individuals where each individual has 11 images in Yale face database. 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 2.

Fig. 2. Sample images of one individual in the YALE database.

The YaleB contains 21888 images of 38 individuals under 9 poses and 64 illumination conditions. A subset containing 2414 frontal pose images of 38 individuals under different illuminations per individual is extracted. The sample images of one individual from the YaleB database are showed in Figure 3.

3.2 the analysis of the performance with respect to t and k

In our experiments, the similarity matrix $\mathbf{S}$ is governed by two parameters: the neighborhood size k and heat kernel parameter t. k is searched from $\{2, 3, \ldots, N-1\}$. Each image vector is normalized before calculating the similarity matrix $\mathbf{S}$. We randomly split the image samples so that p (for ORL and yale, $p = 2, 3, 4, 5, 6, 7, 8$; for YaleB, $p = 5, 10, 20, 30, 40, 50$) images for each individual are used as the training set and the rest are used as the testing set.

In yale database, p is set as 2 i.e. 30 samples in the training set, and t is searched from $\{1, 1.1, 1.2, \ldots, 1.9, 2, 3, \ldots, 9, 10, 20, \ldots, 90, 100\}$. This process is repeated 50 times. The results of three of them are plotted in Fig. 4. From the figure, little influence has been brought to the performances due to the variation of t value. Due to normalization of the image vectors, when $t > 2$, $exp(-\|x_i - x_j\|^2/t)$ approximates 1. From the figure, we can also see that in

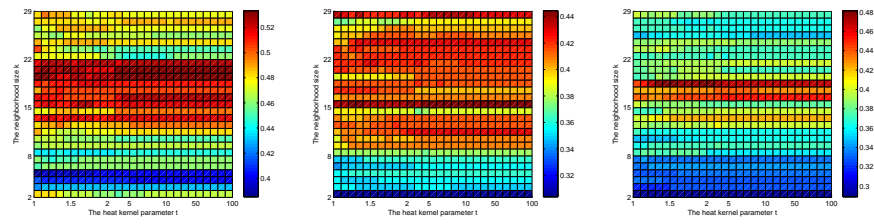**Fig. 3.** Sample images of one individual from the YaleB database.

Fig. 4. The performance of LPP vs. the two parameters k and t on Yale face database

the same the number of the training samples, the top performance does not also incur in the same neighborhood size.

The same experiments are conducted on the ORL, Yale, YaleB face databases. t is searched from $\{1, 1.1, 1.2, \ldots, 1.9, 2, 3, 4, 5\}$. The several results are plotted in Fig. 5. From the figure, the top performance incurs on when the neighborhood size k is greater than the half of the numbers of the samples. And we can also see the fact that the performance is sensitive to the parameter k. This stems from the fact that the essential manifold structure of samples. An alternative interpretation is that facial images lie on multi-manifolds instead of a single manifold. Recently, the efforts of research on multi-manifolds for face recognition are proposed[8]. In order to verify the validation of the assumption that the performance is insensitive to the heat kernel parameter t and the top performance incurs in the case that the neighbors size k is greater than the half of the number of the samples, 50-time cross-validations are performed on Yale database. The results are illustrated in Fig. 6.
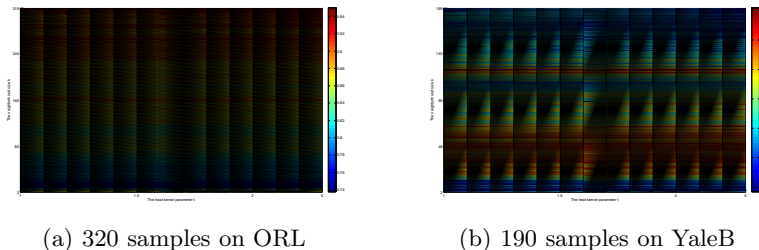

(a) 320 samples on ORL   (b) 190 samples on YaleB

Fig. 5. The performance of LPP vs. the two parameters k and t

4 Conclusion

LPP is a very popular subspace transformation method for face recognition. Recently, its many variants have been proposed. However, their performances mainly depend on how to construct the adjacent graph, which artificially constructs the local structure. To the best of our knowledge, there is no report on the relation between the performance of LPP vs. the nearest neighbor size k and the heat kernel parameter t. This issue is discussed in this paper. We find that the performance is insensitive to the heat kernel parameter t and the top performance incurs in the case that the neighbors size k is greater than the half of the number of the samples.

Our future researches will focus on the performance of the variants of LPP vs. the two parameters t and k. We also focus on the multi-manifold face recognition.
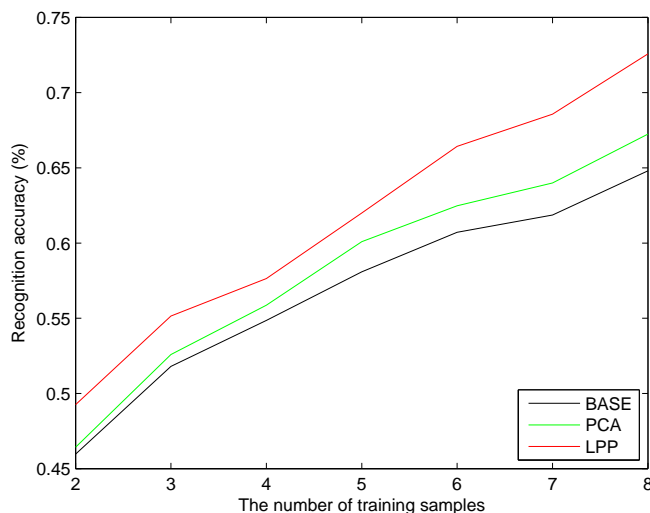
Fig. 6. the grid-search parameter result on yale database.

References

1. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 711–720 (Jul 1997)
2. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. Advances in neural information processing systems 1, 585–592 (2002)
3. Chen, S.B., Zhao, H.F., Kong, M., Luo, B.: 2D-LPP: a two-dimensional extension of locality preserving projections. Neurocomputing 70(4-6), 912–921 (2007)
4. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(6), 643–660 (2001)
5. He, X.F., Niyogi, P.: Locality preserving projections. In: Advances in neural information processing systems. vol. 16, pp. 153–160. The MIT Press (2004)
6. He, X.F., Yan, S.C., Hu, Y.X., Niyogi, P., Zhang, H.J.: Face recognition using laplacianfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(3), 328–340 (2005)
7. Liu, Y., Liu, Y., Chan, K.C.C.: Tensor distance based multilinear locality-preserved maximum information embedding. IEEE Transactions on Neural Networks 21(11), 1848–1854 (2010)
8. Park, S., Savvides, M.: An extension of multifactor analysis for face recognition based on submanifold learning. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. pp. 2645–2652. IEEE (2010)
9. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500), 2323 (2000)

10. Seung, H.S., Lee, D.D.: The manifold ways of perception. Science 290(5500), 2268–2269 (2000)
11. Tenenbaum, J., Silva, V., Langford, J.: A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319 (2000)
12. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (Jan 1991)
13. Wan, M.H., Lai, Z.H., Shao, J., Jin, Z.: Two-dimensional local graph embedding discriminant analysis (2DLGEDA) with its application to face and palm biometrics. Neurocomputing 73(1-3), 197–203 (Dec 2009)
14. Xu, Y., Zhong, A., Yang, J., Zhang, D.: LPP solution schemes for use with face recognition. Pattern Recognition (2010)
15. Yu, W.W., Teng, X.L., Liu, C.Q.: Face recognition using discriminant locality preserving projections. Image and Vision Computing 24(3), 239–248 (2006)
16. Yu, W.: Two-dimensional discriminant locality preserving projections for face recognition. Pattern Recognition Letters 30(15), 1378–1383 (Nov 2009)