A SUPERVISED NONLINEAR LOCAL EMBEDDING FOR FACE RECOGNITION

Jian Cheng[a], Qingshan Liu[a], Hanqing Lu[a] and Yen-Wei Chen[b]
[a] National Laboratory of Pattern Recognition, Institute of Automation, CAS, Beijing 100080,China
[b] College of Information Science and Engineering, Ritsumeikan University, 525-8577, Japan.

ABSTRACT

Many recent works demonstrated that subspace analysis is a good method for face recognition. How to find the subspace is a key issue. In this paper, a Supervised Nonlinear Local Embedding (SNLE) method is proposed to construct a subspace for face recognition, in which we combine the idea of nonlinear kernel mapping and preserving local geometric relations of the samples belonging to same class. SNLE can not only gain a perfect approximation of the nonlinear face manifold, but also enhance within-class local information. Moreover, it is also equivalent to solving a generalized eigenvalue problem in mathematics. Our experiments are performed on two benchmarks, and experimental results show that the proposed method has an encouraging performance.

1. INTRODUCTION

Face recognition has attracted much attention in last two decades because of its potential applications, such as biometrics, human-computer interface, information security, and so on. Numerous algorithms have been proposed and many works show that subspace analysis is a perfect method for face recognition. However, how to find the subspace is a key issue, because different subspace structure will produce different recognition performance. Eigenface [1] uses Principal Component Analysis (PCA) to construct a linear subspace that spanned by a set of the leading eigenvectors of covariance matrix of the training samples. Linear Discriminant Analysis or Fisher face [2] is an efficient discriminating subspace method that seeks a projection transformation maximizing the ratio between the between-class scatter and within-class scatter matrix.

Because there exist complex nonlinear variations of real face images, such as illumination, pose and expression, face images should reside on a nonlinear subspace, i.e., nonlinear face manifold. Some algorithms have been proposed to learn a compact representation of the manifold. Such as Locally Linear Embedding (LLE) [3], Isomap [4], Laplace Eigenmap [5]. However, these nonlinear methods offer embeddings without mapping for new test. To overcome this drawback, He et al [6] proposed a method named Locality Preserving Projection (LPP) to approximate the eigenfunctions of the Laplace Beltrami operator on the face manifold and new test can be explicitly mapped to the learned low-dimensional face submanifold.

However, LPP is a linear method in nature, so it is inadequate to represent the nonlinear face manifold. Although He et al have given a proof in form that LPP can be generalized to reproducing kernel Hilbert space through a nonlinear mapping, it will yield the same results as Laplacian Eigenmaps on the training points [7]. Moreover, LPP uses the nearest-neighbor graph to seek the nearest neighbors, which will fail on database containing complex variations, because the nearest neighbor samples may belong to different classes in databases containing complex variations, such as lighting, expression, pose, and so on.

In this paper, motivated by LPP and the idea of nonlinear kernel mapping trick, we proposed a Supervised Nonlinear Local Embedding (SNLE) approach to recover the intrinsic geometric structure of face space for face recognition. First, nonlinear kernel mapping is used to map the input samples into an implicit feature space *F*. Then we seek a transformation that tries to reduce dimensionality while the within-class geometric structures are preserved in feature space. The experiments are carried out on the Yale and ORL database. Comparison with related works are also performed. The experimental results are impressive.

2. LOCALITY PRESERVING PROJECTION

LPP seeks a linear transformation P to project high-dimensional data into a low-dimensional submanifold that preserves the local structure of the data. The linear transformation P can be obtained by minimizing an objective function as follows [6]:

$$P = \arg\min_{P^T X D X^T P = 1} P^T X (D - W) X^T P \qquad (1)$$

where $X = [x_1, x_2, \cdots, x_n]$ is training samples, W is the weight matrix and $D = \sum_j W(i,j)$. $L = D - W$ is called Laplacian matrix. The minimization problem can be converted to solving a generalized eigenvalue problem as follows:

0-7803-8554-3/04/$20.00 ©2004 IEEE.

$$XLX^TP = \lambda XDX^TP \quad (2)$$

Actually, the weight is used to measure the similarity between a pair of samples. In [6], the weight matrix W is constructed through the nearest-neighbor graph. If x_i is among l nearest neighbors of x_j or x_j is among l nearest neighbors of x_i, then

$$W(i,j) = e^{-\frac{\|x_i - x_j\|^2}{t}} \quad (3)$$

Otherwise, $W(i,j)=0$. Alternatively, the weight matrix can be simply set: $W(i,j)=1$ when x_j and x_j are nearest neighbors; otherwise $W(i,j)=0$.

3. THE PROPOSED METHOD

LPP is a linear method in nature, so it is inadequate to represent the nonlinear face space. Moreover, LPP gives the weight matrix through the nearest-neighbor graph, but the nearest face images may belong to different classes due to pose and lighting variations. In this paper, we proposed a Supervised Nonlinear Local Embedding (SNLE) method to recover the intrinsic nonlinear geometrical structure of face space for face recognition. Based on class labels information, our approach gets a nonlinear face submanifold in kernel feature space. Assuming a set of face images $X=[x_1, x_2, \cdots, x_n]$, x_i is a N-dimensional face image. Firstly, we use a nonlinear function ϕ to map the data into a high-dimensional feature space F : $\phi(X)=[\phi(x_1),\phi(x_2),\cdots,\phi(x_n)]$. Then in feature space F, similar to LPP, a projecting transformation P_ϕ that can preserve within-class geometric structure of the data $\phi(X)$ is learned by minimized the sum of the weighted distance of samples. The minimization problem can be expressed as:

$$\min \sum_{i,j=1}^{n} (y_i - y_j)^2 W(i,j) \quad (4)$$

where $y_i = P_\phi^T \phi(x_i)$ is the projected sample in feature space, $W(i,j)$ is the associated weight representing the geometric relation of x_i and x_j. The objective function (4) can be simplified as:

$$\begin{aligned} &\sum_{i,j=1}^{n} (y_i - y_j)^2 W(i,j) \\ &= \sum_{i,j=1}^{n} (P_\phi^T \phi(x_i) - P_\phi^T \phi(x_j))^2 W(i,j) \\ &= 2P_\phi^T \phi(X)(D-W)\phi(X)^T P_\phi^T \end{aligned} \quad (5)$$

where $D_{ii} = \sum_j W(i,j)$ is a diagonal matrix. Note that the projecting transformation P_ϕ should lie in the span of $\phi(x_1),\phi(x_2),\cdots,\phi(x_n)$. This means existing a coefficient vector $\alpha=[\alpha_1,\alpha_2,\cdots,\alpha_n]^T$ such that

$$P_\phi = \sum_{i=1}^{n} \alpha_i \phi(x_i) = \phi(X)\alpha \,. \quad (6)$$

Substituting (6) into (5), we can obtain:

$$\begin{aligned} &\sum_{i,j=1}^{n} (y_i - y_j)^2 W(i,j) \\ &= 2\alpha^T K(D-W)K\alpha \end{aligned} \quad (7)$$

where $K(i,j) = <\phi(x_i),\phi(x_j)>$ is a positive definite and symmetric matrix.

In order to normalized the solutions, we put a constraint on the solution:

$$\alpha^T KDK\alpha = 1 \quad (8)$$

The minimization problem can be converted to a generalized eigenvalue problem. The eigenvectors corresponding to the least eigenvalues are the solution:

$$K(D-W)K\alpha = \lambda KDK\alpha \quad (9)$$

where the kernel matrix K is a dot product in kernel feature space. W is still unknown. Next, we will give the definition of the weight matrix W containing the class label information.

In fact, every entry of weight matrix W can be regarded as the similarity metric of a pair of samples. In this paper, we take the dot product between two samples to measure their similarity, and define the weight matrix W as follows: $W(i,j) = <\phi(x_i),\phi(x_j)>$ when x_i and x_j belong to same class, otherwise, $W(i,j)=0$. We can see that the form of W is consistent with that of the matrix K except a supervised constraint. Here, we do not yet give an explicit definition of nonlinear mapping ϕ. According to the idea of kernel mapping trick, it is unnecessary to know the nonlinear mapping ϕ explicitly, and we are just demanded to calculate the dot product of two samples with a kernel function $k(x,y) = <\phi(x),\phi(y)>$. In our experiments, two kernel functions are discussed, i.e. polynomial kernel:

$$k(x_i, x_j) = (a(x_i \cdot x_j) + b)^d \quad (10)$$

Gaussian kernel:

$$k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / (2\sigma^2)) \quad (11)$$

According to above description, the SNLE learning algorithm can be summarized as follows:

Given training face images $X = [x_1, x_2, \cdots, x_n]$ and kernel function k

Step 1: Compute kernel matrix K: $K(i,j) = k(x_i, x_j)$.

Step 2: Compute weight matrix W:

if x_i and x_j belong to same class, $W(i,j) = k(i,j)$;

otherwise, $W(i,j) = 0$.

Step 3: Compute diagonal matrix D: $D_{ii} = \sum_j W(i,j)$

Step 4: To solve generalized eigenvalue problem:

$$K(D-W)K\alpha = \lambda KDK\alpha$$

Retain eigenvectors corresponding to the least eigenvalues. For simplicity, we denote them α.

Step 5: The new features of training images is $Y = \alpha^T K$

For a new test images x, can be represented as:

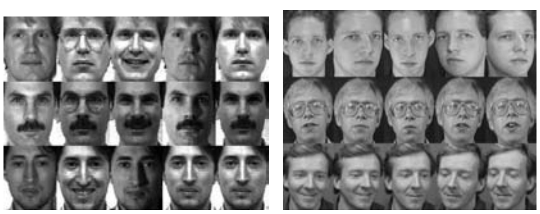
$$y = \alpha^T \phi(X)^T \phi(x)$$

where $\phi(X)^T \phi(x) = [\phi(x_1)^T \cdot \phi(x), \cdots, \phi(x_n)^T \cdot \phi(x)]^T$

Figure 1. SNLE learning algorithm.

4. EXPERIMENTAL RESULTS

We test the proposed method (SNLE) against LPP and the other four popular subspace methods: PCA, LDA (i.e. Fisherface), KPCA [8] and KLDA [9] on two publicly available database: Yale and ORL (AT&T) (Fig.2). Among the six algorithms, PCA, LPP and LDA are linear methods while SNLE, KPCA and KLDA are nonlinear methods. PCA, LPP and KPCA are unsupervised approaches, whereas SNLE, LDA and KLDA are supervised approaches.



(a) (b)

Figure 2. (a) samples from Yale database, (b) samples from ORL database.

The Yale database used in our experiments contains 165 grayscale face images from 15 persons. All face images are cropped into 80×90. Both expression and lighting variation exist in Yale images. The ORL database contains 40 persons, 10 grayscale face images each person, 400 images in total, whose images include variations in pose and scale. The size of face images is 92×112. The gray values of all original images are rescaled to [0,1] and the norm of each image vector is normalized to 1.

All experiments were performed using the *Leave-One-Out* strategy: one face image is selected as test from the database and the remained are training data. The recognition rate is the average of all possible selection. The feature representation after embedded into the nonlinear submanifold can be calculated as the procedure described in Fig.1. Cosine similarity metric is used to measure similarity between a pair of feature representations.

$$d(y_i, y_j) = \frac{y_i^T \cdot y_j}{\| y_i \| \cdot \| y_j \|} \qquad (12)$$

It is well known that the kernel selection is still an open problem till now. In this paper, two kernels are discussed without theoretical consideration, i.e., polynomial kernel (10) and Gaussian kernel (11). For simplicity, we denote them SNLE_P, KPCA_P and KLDA_P when polynomial kernel is used, SNLE_G, KPCA_G and KLDA_G when Gaussian kernel is adopted. The parameters of polynomial kernel are empirically set as: a = 0.1, b = 0, d = 2, and we set parameter $\sigma = 1$ for Gaussian kernel.

We investigate recognition rate varying with number of components. The experimental results on the Yale and ORL database are shown in left column of Fig.3 and Fig.4, respectively. Because the maximum components of LDA and KLDA are no more than $c-1$ (c is the number of class), the results of LDA and KLDA are not plotted in Fig.3 and Fig.4. It is shown that SNLE outperforms the other five algorithms. The comparison experiments between polynomial and Gaussian kernel are reported in right column of Fig.3 and Fig.4. The polynomial kernel behaves slight better than Gaussian kernel. A summary of the best recognition rate is provided in Table 1. SNLE_P achieves 99.39% on the YALE database and 98.75% on the ORL database. The consistent good performances achieved by SNLE demonstrate that SNLE is reasonable.

5. CONCLUSIONS

In this paper, we present a Supervised Nonlinear Local Embedding (SNLE) to learn the embedding from a high-dimensional face images space into the intrinsic low-dimensional space. Our algorithm attempts to contain the geometric relation between within-class samples. The experimental results are impressive.

Acknowledge

This work is partly supported by Natural Sciences Foundation of China under grant No. 60135020 and 60121302.

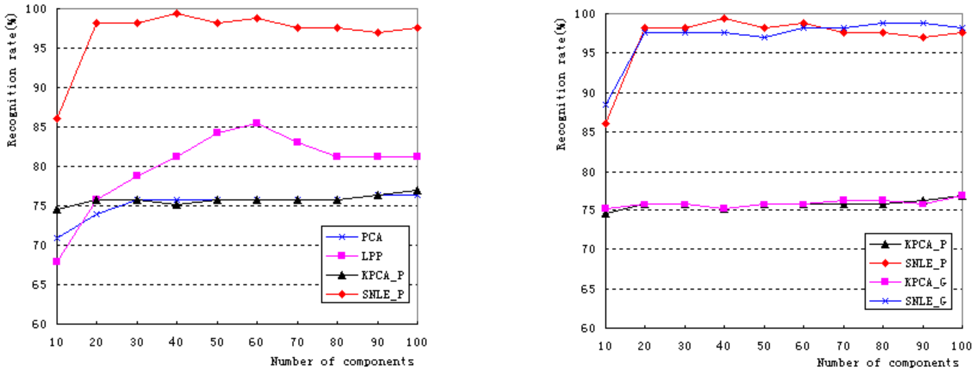


Figure 3. Experimental results on the Yale database. Left: different methods; Right: different kernels.

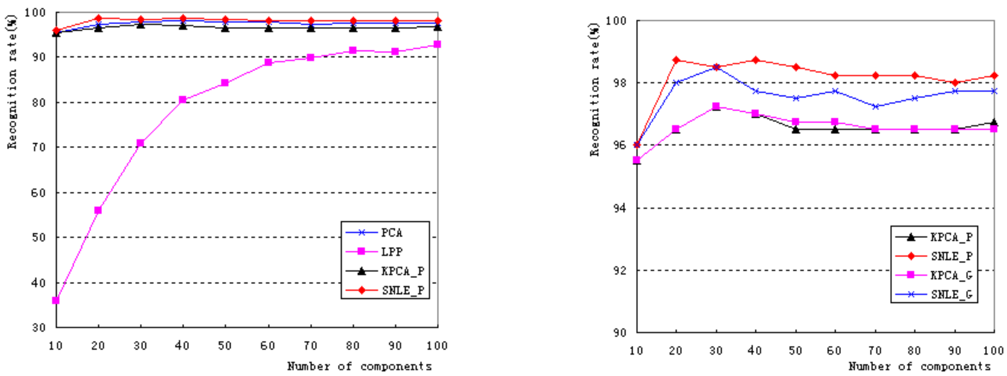


Figure 4. Experimental results on the ORL database. Left: different methods; Right: different kernels.

Table 1. Comparison results among 6 methods and 2 kernel functions

Method	Dims	Recognition rate (YALE)	Dims	Recognition rate (ORL)
PCA	90	76.36%	40	98.00%
LPP	60	85.45%	100	92.75%
LDA	14	96.96%	39	94.75%
KPCA_P	100	76.96%	30	97.25%
KPCA_G	100	76.96%	30	97.25%
KLDA_P	14	98.78%	39	98.50%
KLDA_G	14	98.18%	39	97.75%
SNLE__P	40	99.39%	20	98.75%
SNLE_G	80	98.78%	30	98.50%

11. REFERENCES

[1] M. Turk and Pentland, "Face recognition using eigenfaces", *IEEE Conference on Computer Vision and Pattern Recognition,* Maui, Hawaii, 1991.

[2] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection", *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol.19, No.7, pp.711-720, 1997.

[3] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding", *Science,* Vol.290, pp.2323-2326, 2000.

[4] J. Tenenbaum, V. de Silva and J. Langford, "A global geometric framework for nonlinear dimensionality reduction", *Science*, Vol.290, pp.2319-2323, 2000.

[5] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering", *Advances in Neural Information Processing System* 15, Vancouver, Canada, 2001.

[6] Xiaofei He, Shuicheng Yan, Yuxiao Hu and Hong-Jiang Zhang, "Learning a locality preserving subspace for visual recognition", *ICCV 2003*, pp.385-392.

[7] Xiaofei He and Partha Niyogi, "Locality Preserving Projections," *Advances in Neural Information Processing Systems 16*, Vancouver, Canada, 2003.

[8] Yang M.H., Ahuja N. and Kriegman D., "Face recognition using kernel Eigenfaces". *In proc. of ICIP*, pp.37-40. Vancouver, Canada, 2000.

[9] Qingshan Liu, Rui Huang, Hanqing Lu and Songde Ma, "Kernel-based nonlinear discriminant analysis for face recognition", *J. Comput. Sci. & Technol.,*Vol.18, No.6, pp.788-795. 2003.