Available online at www.sciencedirect.com
SCIENCE DIRECT®
Neurocomputing 67 (2005) 443–449
NEUROCOMPUTING
www.elsevier.com/locate/neucom

Letter

Supervised kernel locality preserving projections for face recognition

Jian Cheng[a], Qingshan Liu[a], Hanqing Lu[a,*], Yen-Wei Chen[b]

[a]*National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, P.O. Box 2728, Beijing 100080, China*

[b]*College of Information Science and Engineering, Ritsumeikan University, 525-8577, Japan*

Received 24 June 2004; received in revised form 9 August 2004; accepted 18 August 2004
Communicated by R.W. Newcomb
Available online 24 February 2005

Abstract

Subspace analysis is an effective approach for face recognition. Finding a suitable low-dimensional subspace is a key step of subspace analysis, for it has a direct effect on recognition performance. In this paper, a novel subspace method, named supervised kernel locality preserving projections (SKLPP), is proposed for face recognition, in which geometric relations are preserved according to prior class-label information and complex nonlinear variations of real face images are represented by nonlinear kernel mapping. SKLPP cannot only gain a perfect approximation of face manifold, but also enhance local within-class relations. Experimental results show that the proposed method can improve face recognition performance.

Keywords: Kernel trick; Subspace analysis; Locality preserving projection; Face recognition

1. Introduction

Recent studies show that subspace analysis is an effective approach for face recognition [1,3,7,10]. Specially learning a compact manifold (subspace) that can

*Corresponding author. Tel.: +86 1082 614 465; fax: +86 1062 551 993.
E-mail address: luhq@nlpr.ia.ac.cn (H. Lu).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2004.08.006

preserve local structure of face images has attracted a great deal of attention in the past few years. There are three popular manifold learning methods, i.e., locally linear embedding (LLE) [4], Isomap [6], Laplacian Eigenmap [2], but these methods are not suitable for face recognition, because they cannot give an explicit subspace mapping for a new test sample. In order to overcome this drawback, He et al. proposed a method, named locality preserving projections (LPP) [9], to approximate the eigenfunctions of the Laplace Beltrami operator on the face manifold, and new face images can be easily mapped to the learned low-dimensional face subspace. Although LPP is successful in many circumstances, it often fails to deliver good performance when face images are subject to complex nonlinear changes due to large pose, expression or illumination variations, for it is a linear method in nature.

In this paper, a novel subspace analysis method named supervised kernel locality preserving projections (SKLPP) is proposed for face recognition. Firstly, we use nonlinear kernel mapping to map the data into an implicit feature space F, which is successfully used in support vector machine (SVM). Then we seek a linear transformation that can preserve within-class geometric structures in F. Thus, we can gain a nonlinear subspace that can approximate the intrinsic geometric structure of the face manifold. Though He et al. mentioned that LPP could be generalized into a reproducing kernel Hilbert space through a nonlinear mapping, it was not further discussed [8]. Moreover, LPP seeks to preserve local structure defined by the nearest neighbors. So it fails to preserve within-class local structure, which is very important for object recognition, because the nearest neighbors may belong to different classes due to influence of complex variations, such as lighting, expression, pose, and so on.

The rest of this paper is organized as follows: LPP is introduced briefly in Section 2. SKLPP is proposed in Section 3. Experimental results are reported in Section 4, and followed by the conclusions in Section 5.

2. Locality preserving projections

LPP is a linear approximation of Laplacian Eigenmap [2]. It seeks a transformation P to project high-dimensional input data $X = [x_1, x_2, \ldots, x_n]$ into a low-dimensional subspace Y in which the local structure of the input data can be preserved. The linear transformation P can be obtained by minimizing an objective function as follows [9]:

$$\min_P \sum_{i,j=1}^{n} \|y_i - y_j\|^2 S(i,j), \tag{1}$$

where $y_i = P^{\mathrm{T}} x_i$, the weight matrix S (called heat kernel) is constructed through the nearest-neighbor graph. If x_i is among the l nearest neighbors of x_j or x_j is among the l nearest neighbors of x_i, then

$$S(i,j) = \mathrm{e}^{-\frac{\|x_i - x_j\|^2}{t}}, \tag{2}$$

where parameter t is a suitable constant. Otherwise, $S(i,j) = 0$. Alternatively, the weight matrix can be simply set as: $S(i,j) = 1$ when x_i and x_j are the nearest neighbors, otherwise $S(i,j) = 0$. The minimization problem can be converted to solving a generalized eigenvalue problem as follows:

$$XLX^{\mathrm{T}}P = \lambda XDX^{\mathrm{T}}P, \tag{3}$$

where $D_{ii} = \sum_j S(i,j)$ is a diagonal matrix, and $L = D - S$.

3. Supervised kernel locality preserving projections

LPP is a linear method in nature, and it is inadequate to represent the nonlinear face space. Moreover, LPP seeks to preserve local structure defined by the nearest neighbors. It often fails to preserve within-class local structure, which is very important for object recognition, because the nearest neighbors may belong to different classes due to influence of complex variations, such as lighting, expression, and pose. In this paper, we propose a novel subspace method for face recognition, i.e., SKLPP. First, the nonlinear kernel mapping is used to map the data into an implicit feature space F, which is successfully used in SVM, and then a linear transformation is performed to preserve within-class geometric structures in F. Thus, we can gain a nonlinear subspace that can approximate the intrinsic geometric structure of the face manifold.

Assuming a set of face images $X = [x_1, x_2, \ldots, x_n]$, 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: $\phi(X) = [\phi(x_1), \phi(x_2), \ldots, \phi(x_n)]$. Then in feature space F, we seek a projecting transformation P_ϕ that can preserve the within-class geometric structure of the data $\phi(X)$ by minimizing the sum of the weighted distance of samples. The minimization problem can be expressed as

$$\min_{P_\phi} \sum_{i,j=1}^{n} \|z_i - z_j\|^2 W(i,j), \tag{4}$$

where $z_i = P_\phi^{\mathrm{T}}\phi(x_i)$ is the projection of $\phi(x_i)$ onto P_ϕ, and the weight $W(i,j)$ represents the relations of x_i and x_j. The objective function (4) can be simplified as

$$\begin{aligned}\sum_{i,j=1}^{n} \|z_i - z_j\|^2 W(i,j) &= \sum_{i,j=1}^{n} \|P_\phi^{\mathrm{T}}\phi(x_i) - P_\phi^{\mathrm{T}}\phi(x_j)\|^2 W(i,j) \\ &= 2P_\phi^{\mathrm{T}}\phi(X)(D - W)\phi(X)^{\mathrm{T}}P_\phi,\end{aligned} \tag{5}$$

where $D_{ii} = \sum_j W(i,j)$ is a diagonal matrix. Because the linear transformation P_ϕ should lie in the span of $\phi(x_1), \phi(x_2), \ldots, \phi(x_n)$, there exists a coefficient vector $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_n]^{\mathrm{T}}$ such that

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

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

$$\sum_{i,j=1}^{n} \|z_i - z_j\|^2 W(i,j) = 2\alpha^{\mathrm{T}} K(D-W)K\alpha, \tag{7}$$

where the matrix $K(i,j) = \phi(x_i)\cdot\phi(x_j)$, is a positive definite and symmetric matrix. According to the kernel trick, the dot produce of two vectors in F is calculated by a kernel function $k(x,y) = \phi(x)\cdot\phi(y)$ without knowing the nonlinear mapping ϕ explicitly.

Thus, this minimization problem can be converted to a generalized eigenvalue problem with a constraint condition $\alpha^{\mathrm{T}}KDK\alpha = 1$. The eigenvectors corresponding to the smallest eigenvalues are the solution

$$K(D-W)K\alpha = \lambda KDK\alpha. \tag{8}$$

Up to now, the weight matrix W is still unknown. In [8,9], the weight matrix W is just defined by the nearest-neighbor relations. Here, with prior class label information, we define the W using a supervised approach. In fact, each entry of the weight matrix W can be regarded as the similarity metric of a pair of samples. The dot product between two samples is in a sense a similarity measure. So we define the weight matrix W as follows:

$$W(i,j) = \begin{cases} \phi(x_i)\cdot\phi(x_j) & \text{if } x_i \text{ and } x_j \text{ belong to the same class,} \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

It means that the within-class geometric information is emphasized, and we set the similarity between two samples to zero, if they belong to different classes. From (9), we find that the matrix W and K are unified into a consistent dot product form except that the matrix W has a strong constraint.

4. Experimental results

To verify the proposed method, SKLPP is applied to face recognition compared with LPP [8,9], PCA [7], LDA [1], KPCA [10], and KLDA [3]. The experiments are performed on two publicly available databases: Yale [1] and ORL [5]. The Yale database contains 165 grayscale face images from 15 persons. All face images are cropped into 80×90. Both expression and lighting variations exist in the Yale database. The ORL database contains 40 persons, and each person has 10 different grayscale face images that include variations in pose and scale. The size of face images is 92×112. The gray values of all images are rescaled to [0,1] and the norm of each image vector is normalized to 1 (Fig. 1).

It is well known that the kernel selection is still an open problem till now. In this paper, two kernels are discussed, i.e., polynomial kernel

$$k_{\mathrm{P}}(x,y) = \phi(x)\cdot\phi(y) = (a(x\cdot y))^d \tag{10}$$

and Gaussian kernel

$$k_G(x, y) = \phi(x) \cdot \phi(y) = \exp(-\|x - y\|^2/(2\sigma^2)). \tag{11}$$

As can be seen, heat kernel (Eq. (2)) can be regarded as a special case of Gaussian kernel. For subsequent analysis, we denote the three kernel-based methods as SKLPP_P, KPCA_P and KLDA_P when the polynomial kernel is used, and as SKLPP_G, KPCA_G and KLDA_G when the Gaussian kernel is adopted. The parameters of the polynomial kernel are empirically set as: $a = 0.1$, $d = 2$, and $\sigma = 1$ for Gaussian kernel.

All experiments were conducted using the *Leave-One-Out* strategy. For simplicity, the nearest-neighbor classifier based on the Cosine distance metric is used.

$$d(z_i, z_j) = 1 - \frac{z_i^{\mathrm{T}} \cdot z_j}{\|z_i\| \cdot \|z_j\|}. \tag{12}$$

The best recognition rates are shown in Table 1. SKLPP_P achieves the best recognition rate 99.39% with 40 dimensions while LPP only gets 85.45% with 60 dimensions on the YALE database. On the ORL database, SKLPP_P achieves 98.75% with 20 dimensions while LPP gets 92.75% with 100 dimensions. Because the maximum dimensions of LDA and KLDA are no more than $c - 1$ (c is the number of classes), the best recognition rates of LDA and KLDA are obtained with

Fig. 1. Upper are samples from the Yale database, bottom are samples from the ORL database.

Table 1
Best recognition rates among 6 algorithms 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
SKLPP_P	40	99.39	20	98.75
SKLPP_G	80	98.78	30	98.50

14 dimensions on YALE and 39 dimensions on ORL, respectively. It is shown that the polynomial kernel is slightly better than the Gaussian kernel in our experiments. Experimental results suggest that SKLPP also outperforms the other methods. It demonstrates that the performance is significantly improved because SKLPP takes into account the nonlinear within-class structure in kernel feature space.

5. Conclusions

In this paper, we present a novel subspace analysis method, named SKLPP, which attempts to preserve local geometric relations of the within-class samples in nonlinear kernel feature space. Extensive experiments on two benchmarks show that the proposed method has an encouraging performance.

Acknowledgements

We would like to thank Prof. Xiaoou Tang for his proofreading and the anonymous reviewers for their useful comments. We also would like to acknowledge support from Natural Sciences Foundation of China under Grant no. 60475010, 60405005 and 60121302, the joint fund of NSFC-RGC under Grant no. 60318003.

References

[1] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.

[2] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, in: Proceedings of Advances in Neural Information Processing System 14, Vancouver, Canada, December 2001.

[3] Q. Liu, R. Huang, H. Lu, S. Ma, Face recognition using kernel based fisher discriminant analysis, in: Proceedings of the International Conference on Automatic Face and Gesture Recognition, Washington, May 2002, pp. 197–201.

[4] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.

[5] F. Samaria, A. Harter, Parameterisation of a stochastic model for human face identification, in: Proceedings of the Second IEEE Workshop on Applications of Computer Vision, Sarasota, USA, December 1994.

[6] J. Tenenbaum, V. de Silva, J. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319–2323.

[7] M. Turk, Pentland, Face recognition using eigenfaces, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, June 1991, pp. 586–591.

[8] Xiaofei He, Partha Niyogi, Locality preserving projections, in: Proceedings of Advances in Neural Information Processing Systems 16, Vancouver, Canada, December 2003.

[9] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Hong-Jiang Zhang, Learning a locality preserving subspace for visual recognition, in: Proceedings of Ninth International Conference on Computer Vision, France, October 2003, pp. 385–392.

[10] M.H. Yang, N. Ahuja, D. Kriegman, Face recognition using kernel eigenfaces, in: Proceedings of International Conference on Image Processing, Vancouver, Canada, September 2000, pp. 37–40.

Jian Cheng was born in 1977. He received a B.S. degree and a M.S. degree in Mathematics from Wuhan University, China in 1998 and 2001, respectively. Currently, he is a Ph.D. candidate in Institute of Automation, Chinese Academy of Sciences, Beijing. His research interests include face recognition, machine learning, image retrieval, etc.

Qingshan Liu was born in 1975. He received his B.E. degree from Chinese Textile University (now DongHua University) in 1997, M.S. degree from South East University in 2000. He received his Ph.D. degree in 2003 from the National Laboratory of Pattern Recognition of Institute of Automation, Chinese Academy of Science, where now he is an assistant professor. His current research interests are image and video processing, object tracking and recognition, and machine learning.

Hanqing Lu Professor received his B.E. degree in 1982 and his M.E. degree in 1985 from Harbin Institute of Technology, and Ph.D. degree in 1992 from Huazhong University of Sciences and Technology. Currently, he is a deputy director of National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His research interests include image and video analysis, medical image processing, object recognition, etc. He has published more than 100 papers in these fields.

Yen-Wei Chen received a B.E. degree in 1985 from Kobe University, a M.E. degree in 1987 and a D.E. degree in 1990, both from Osaka University, Japan. He is currently a professor in college of Information Science and Engineering, Ritsumeikan University. He is an Overseas Assessor of Chinese Academy of Sciences. His research interests include intelligent signal and image processing, radiological imaging and soft computing. He has published more than 100 research papers in these fields.