A New Kernel Orthogonal Projection Analysis Approach for Face Recognition

Xiaoyuan Jing[1,2], Min Li[1]*, Yongfang Yao[1],Songhao Zhu[1],Sheng Li[1]
[1]College of Automation, Nanjing University of Posts and Telecommunications, Nanjing, China
[2]State Key Laboratory of Software Engineering, Wuhan University, Wuhan, China
* Email: lemonlee31@126.com

Abstract—**In the field of face recognition, how to extract effective nonlinear discriminative features is an important research topic. In this paper, we propose a new kernel orthogonal projection analysis approach. We obtain the optimal nonlinear projective vector which can differentiate one class and its adjacent classes, by using the Fisher criterion and constructing the specific between-class and within-class scatter matrices in kernel space. In addition, to eliminate the redundancy among projective vectors, our approach makes every projective vector satisfy locally orthogonal constraints by using the corresponding class and part of its most adjacent classes. Experimental results on the public AR and CAS-PEAL face databases demonstrate that the proposed approach outperforms several representative nonlinear projection analysis methods.**

Keywords-kernel orthogonal projection analysis; feature extraction; locally orthogonal constraints; face recognition

I. INTRODUCTION

How to extract effective nonlinear projective features is an important research topic in face recognition [1]. Though nonlinear methods do not always outperform linear ones [2], to solve the nonlinear classification problems, many kernel discrimination algorithms have been presented, such as kernel discriminant analysis (KDA) [3], kernel independent component analysis (KICA) [4]. Zeng et al. [5] employed the minimum squared errors (MSE) criterion into KDA, and presented KDA-MSE that can reduce the amount of calculation. Cai et al. [6] presented a spectral regression KDA (SRKDA) method, which casts discriminant analysis into a regression framework by using spectral graph analysis. However, one problem of these methods is that they target the global optimal discriminative power for all classes, but not for a specific class. To solve this problem, P. Baggenstoss proposed a class-specific idea that each class has its own feature sets and designed the probabilistic classifiers [7]. Class-specific kernel discriminant analysis (CSKDA) [8] applies this idea to face verification. For each specific class, it acquires a set of projective vectors by minimizing the within-class scatter, and maximizing the between-class scatter that is calculated using the mean of this specific class and the samples of all other classes. However, use all samples in the sample set to construct the between-class scatter matrix of a specific class is unnecessary.

Some feature extraction methods considering the local structure of data have been proposed, such as locality preserving projections (LPP) [9], which aims at seeking a low-dimensional embedded space to preserve the locality relationship of samples. LPP is an unsupervised method, so it has no direct connection to classification. Kernel local Fisher discriminant analysis (KLFDA) [10] uses local neighbourhood information to construct the weighted between-class and within-class scatter matrices, and then performs projection analysis.

In order to eliminate the redundancy among projective vectors and among feature coordinates extracted in the kernel subspace [11], Liang et al. [12] presented the kernel uncorrelated discriminant vectors (KUDV) which is subject to St-orthogonal constraint. Based on KUDV, Jing et al. [13] proposed a kernel uncorrelated adjacent-class discriminant analysis (KUADA) approach which considers discriminative power of each specific class and makes every projective vector satisfy St-orthogonal constraint. Zheng et al. [14] put forward a kernel Foley-Sammon optimal discriminant vectors (KODV) method, which subject to orthogonal constraint. On the basis of KODV, we proposed our approach in order to eliminate the correlation between projective vectors of the specific class and part of its most adjacent classes in according with orthogonal constraint.

In this paper, by using the Fisher criterion and constructing the specific between-class and within-class scatter matrices in kernel space, we obtain the optimal nonlinear projective vector which can differentiate one class and its nearest neighbor classes. Further, the proposed approach makes every projective vector satisfy orthogonal constraint between the corresponding class and part of its most adjacent classes, which can effectively eliminate the redundancy among projective vectors.

II. APPROACH DESCRIPTION

In this section, we describe the proposed kernel orthogonal projection analysis approach by the following steps:

A. Map original sample set to the kernel space

Let $\phi: R^d \to F$ denote a nonlinear mapping. The original sample set $X = \{X_1, X_2, \cdots, X_c\}$ is injected into the kernel space F by $\phi: x_i \to \phi(x_i)$, and we obtain a set of mapped samples $\Psi = \{X_1^\varphi, X_2^\varphi, \cdots, X_c^\varphi\}$.

B. Get adjacent classes of each class

First, calculate the Euclidean distance between any two classes X_i^φ and X_j^φ in the kernel space as:

$$d\left(X_i^{\varphi}, X_j^{\varphi}\right) = \left\|m_i^{\varphi} - m_j^{\varphi}\right\| , \quad (1)$$

where $\|\ \|$ represents the 2-norm operator, m_i^{φ} and m_j^{φ} are the mean vectors of X_i^{φ} and X_j^{φ}, respectively. After obtaining the Euclidean distance between any two classes, the following distance matrix can be constructed:

$$G(i, j) = d(X_i^{\varphi}, X_j^{\varphi}) . \quad (2)$$

Then, sort G in an ascending order. In this way, we can get the nearest neighbor classes with the smallest between-class distances of the i^{th} class, i.e., the so-called adjacent classes of the i^{th} class. In this paper, we set the number of adjacent classes as the same value K_1 for each class.

C. Construct new scatter matrices with adjacent classes

We reconstructed the between-class scatter matrix $S_b^{i\phi}$ and the total scatter matrix $S_t^{i\phi}$ of the i^{th} class as follows:

$$S_b^{i\varphi} = \left(m_i^{\varphi} - m_i^{k\varphi}\right)\left(m_i^{\varphi} - m_i^{k\varphi}\right)^T , \quad (3)$$

$$\begin{aligned} S_t^{i\varphi} = \frac{1}{2}\Bigg(& \frac{1}{n_i}\sum_{j=1}^{n_i}\left(\varphi\left(x_{ij}\right) - \bar{m}_i^{\varphi}\right)\left(\varphi\left(x_{ij}\right) - \bar{m}_i^{\varphi}\right)^T \\ & + \frac{1}{K_1}\sum_{q=1}^{c}\frac{1}{n_q}\sum_{l=1}^{n_q}\theta_{iq}\left(\varphi\left(x_{ql}\right) - \bar{m}_i^{\varphi}\right)\left(\varphi\left(x_{ql}\right) - \bar{m}_i^{\varphi}\right)^T \Bigg), \end{aligned} \quad (4)$$

where $m_i^{k\varphi} = \frac{1}{K_1}\sum_{q=1}^{c}\frac{1}{n_q}\sum_{l=1}^{n_q}\theta_{iq}\phi\left(x_{ql}\right)$, $\bar{m}_i^{\varphi} = \frac{1}{2}\left(m_i^{\varphi} + m_i^{k\varphi}\right)$, and the coefficient θ_{iq} is defined as

$$\theta_{iq} = \begin{cases} 1 & if\ class\ q\ is\ adjacent\ to\ class\ i \\ 0 & otherwise \end{cases} .$$

We rewrite $S_b^{i\phi}$ and $S_t^{i\phi}$ as $\tilde{S}_t^{i\varphi} = K^i K^i$ and $\tilde{S}_b^{i\varphi} = K^i \mathrm{W}^i K^i$, where K^i is an $m_i \times m_i$ kernel matrix calculated by using the i^{th} class and its adjacent classes, $m_i = n_i + \sum_{q=1}^{c}\theta_{iq} n_q$, $\mathrm{W}^i = diag(\mathrm{w}_1^i, \mathrm{w}_2^i)$, W^i is a $m_i \times m_i$ matrix, and w_1^i is an $n_i \times n_i$ matrix with all terms equal to $1/n_i$ while w_2^i is an $(m_i - n_i)\times(m_i - n_i)$ matrix with all terms equal to $1/(m_i - n_i)$.

D. Calculate the projective vector of the first class

We calculate the projective vector β_1 of the first class with the Fisher criterion as follows:

$$J(\beta_1) = \frac{\left|\beta_1^T \tilde{S}_b^{1\varphi}\beta_1\right|}{\left|\beta_1^T \tilde{S}_t^{1\varphi}\beta_1\right|} . \quad (5)$$

According to (3), the rank of $S_b^{i\varphi}$ is 1. Therefore, β_1 is the eigenvector of $(\tilde{S}_t^{1\varphi})^{-1}\tilde{S}_b^{1\varphi}$ corresponding to the nonzero eigenvalue.

E. Construct locally orthogonal constraints

Assume that we have obtained the first $i-1$ projective vectors $(\beta_1, \beta_2, \cdots, \beta_{i-1})$, β_i is the optimal projective vector of the i^{th} class. For the i^{th} class, we select K_3 obtained optimal projective vectors $(\beta_{j1}, \beta_{j2}, \cdots, \beta_{jK_3})$ subject to locally orthogonal constraints:

$$\beta_i^T \beta_{jm} = 0,\ m = 1, 2, \cdots, K_3 \ \text{and}\ \beta_i^T \beta_i = 1 , \quad (6)$$

where β_{jm} corresponds to one of most adjacent classes of the i^{th} class.

F. Calculate optimal projective vectors of remaining classes

Except the first projective vector β_1 , we calculate the i^{th} optimal projective vector β_i with the Fisher criterion and locally orthogonal constraints as following theorem:

Theorem. The i^{th} optimal projective vector β_i ($i \geq 2$) is the eigenvector corresponding to the nonzero eigenvalue of $(\tilde{S}_t^{i\varphi})^{-1} P_i \tilde{S}_b^{i\varphi}$, where $P_i = I - D_i^T (D_i (\tilde{S}_t^{i\varphi})^{-1} D_i^T)^{-1} D_i (\tilde{S}_t^{i\varphi})^{-1}$,

$$D_i = \left[\beta_1, \beta_2, \cdots, \beta_{K_3}\right]^T , \text{and } I = diag\left(1, 1, \cdots, 1\right) \quad (7)$$

Proof of the Theorem: Use the Lagrange multipliers method to express the Fisher criterion and the locally orthogonal constraints in (6), we have:

$$L(\beta_i) = \beta_i^T \tilde{S}_b^{i\varphi}\beta_i - \lambda(\beta_i^T \tilde{S}_t^{i\varphi}\beta_i - b) - \sum_{m=1}^{K2}\mu_m \beta_i^T \beta_{jm} , \quad (8)$$

where λ and $\mu_m (m = 1, ..., K_3)$ are Lagrange multipliers.

The optimization is performed by setting the partial derivative of $L(\varphi_i)$ to be equal to zero:

$$\partial\left(L(\beta_i)\right)/\partial\left(\beta_i\right) = 0 . \quad (9)$$

So we have:

$$2\tilde{S}_b^{i\varphi}\beta_i - 2\lambda\tilde{S}_t^{i\varphi}\beta_i - \sum_{m=1}^{K_3}\mu_m \beta_{jm} = 0 . \quad (10)$$

Left multiplying (10) by $\beta_{js}^T (\tilde{S}_t^{i\varphi})^{-1}$ $(s = 1, 2, \cdots, K_3)$, we obtain:

$$2\beta_{js}^T (\tilde{S}_t^{i\varphi})^{-1}\tilde{S}_b^{i\varphi}\beta_i - \sum_{m=1}^{K_3}\mu_m \beta_{js}^T (\tilde{S}_t^{i\varphi})^{-1}\beta_{jm} = 0 . \quad (11)$$

where $s = 1, 2, \cdots, K_3$.

Let $$U_i = [\mu_1, \mu_2, \cdots, \mu_{K_3}]^T , \quad (12)$$

$$D_i = \left[\beta_1, \beta_2, \cdots, \beta_{K_3}\right]^T . \quad (13)$$

(11) can be represented as follows:

$$D_i (\tilde{S}_t^{i^\varphi})^{-1} D_i^T U_i = 2 D_i (\tilde{S}_t^{i^\varphi})^{-1} \tilde{S}_b^{i^\varphi} \beta_i . \quad (14)$$

Thus, we obtain:

$$U_i = 2 (D_i (\tilde{S}_t^{i^\varphi})^{-1} D_i^T)^{-1} D_i (\tilde{S}_t^{i^\varphi})^{-1} \tilde{S}_b^{i^\varphi} \beta_i , \quad (15)$$

(10) can be written as:

$$2 \tilde{S}_b^{i^\varphi} \beta_i - 2\lambda \tilde{S}_t^{i^\varphi} \beta_i - D_i^T U_i = 0 . \quad (16)$$

Substituting (15) into (16), we have:

$$\begin{aligned} & 2 \tilde{S}_b^{i^\varphi} \beta_i - 2\lambda \tilde{S}_t^{i^\varphi} \beta_i - D_i^T [2 (D_i (\tilde{S}_t^{i^\varphi})^{-1} D_i^T)^{-1} \\ & D_i (\tilde{S}_t^{i^\varphi})^{-1} \tilde{S}_b^{i^\varphi} \beta_i] = 0 \end{aligned} . \quad (17)$$

Hence, we obtain $P_i \tilde{S}_b^{i^\varphi} \beta_i = \lambda \tilde{S}_t^{i^\varphi} \beta_i$, where β_i is the eigenvector corresponding to the nonzero eigenvalue of $(\tilde{S}_t^{i^\varphi})^{-1} P_i \tilde{S}_b^{i^\varphi}$, where

$$P_i = I - D_i^T (D_i (\tilde{S}_t^{i^\varphi})^{-1} D_i^T)^{-1} D_i (\tilde{S}_t^{i^\varphi})^{-1} . \quad (18)$$

G. Feature extraction and classification

A low-dimensional sample set $Y = (Y^1, Y^2, \cdots, Y^c)^T$ can be described as

$$Y^i = \beta_i^T \hat{K}^{iT} , \quad (19)$$

where $\hat{K}^i$ is an $N \times m_i$ matrix calculated by using all the training samples and the adjacent classes' samples of the i^{th} class. Then we use the nearest neighbor classifier with the cosine distance to classify Y .

III. Experiments and Analysis

In this section, we compare the proposed approach with several representative projection analysis methods on the public AR and CAS-PEAL face databases.

A. Database Introduction

Figure 1. Demo images of one subject from the AR face database.

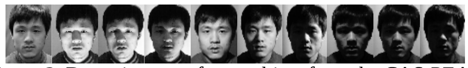

Figure 2. Demo images of one subject from the CAS-PEAL face database

The AR face database [15] contains over 4000 color face images of 126 people, including frontal views of faces with different facial expressions, under different lighting conditions and with various occlusions. We selected images from 119 individuals for use in our experiment for a total number of 3094 (=119 × 26) samples. All color images are transformed into grey images and each image was scaled to 60×60 with 256 grey levels. In order to evaluate the impact of different variations to the recognition results, we randomly choose 5 images of every subject as the training samples. And the remainder images are chosen as the testing samples.

The CAS-PEAL face database [16] we employed contains 1060 images of 106 individuals (10 images each person) with varying lighting. A frontal image of each subject was captured under variable illumination. In the experiment, each image was automatically cropped and scaled to 60×48. Fig. 2 shows 10 images of an individual of the CAS-PEAL face database. We randomly choose 5 images of each person as training samples. The remainder images are regarded as testing samples.

B. Experimental Settings

We compare the proposed approach with six representative methods that are KDA [3], CSKDA [8], KLFDA [10], LPP [9], KODV [14] and KUDV [12]. For all compared methods, we use the nearest neighbor classifier with the cosine distance to do classification. In this paper, we consider the Gaussian kernel $k(x, y) = \exp(-\|x - y\|^2 / 2\delta_i^2)$ for the compared kernel methods, and set the parameter $\delta_i = i \times \delta,\ i \in 1, \cdots, 20$, where δ is the standard deviation of training data set. For each compared kernel method, the parameter i was selected, so that the best classification performance was obtained.

TABLE I. Average recognition rates (%) of compared methods and our proposed approach on two face databases.

Method	**Recognition Rates (%)**		
	AR	***CAS-PEAL***	***Average***
KDA	82.25	89.26	85.75
CSKDA	83.07	89.48	86.27
KLFDA	82.61	87.86	85.23
LPP	77.50	84.51	81.00
KUDV	84.60	90.83	87.71
KODV	84.34	90.24	87.29
Our Approach	**86.33**	**92.26**	**89.29**

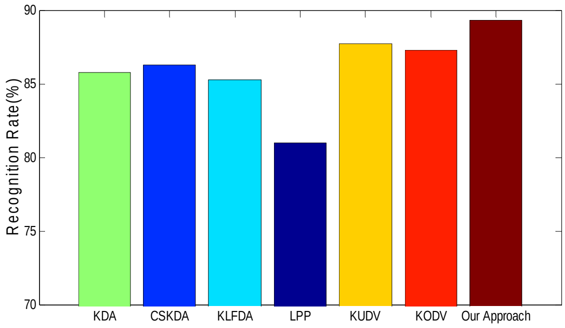

Figure3. Average recognition rates of compared methods and our proposed approach.

In the experiment, K_1 is set as 35 on AR face database. K_3 is determined by the following strategy: set K_3 as the number of most adjacent classes of each class, where the optimal projective vectors of these classes have been obtained; if the number is more than 5, then set $K_3 = 5$ .We set the parameter K_1 as 35 on CAS-PEAL face database and the parameters K_3 can be determined.

C. *Experimental Results and Analysis*

Table 1 shows the average recognition rates of all compared methods across 30 random tests.

In contrast with other related methods, our approach improves at least by 1.73% (=86.33%-84.60%) on AR face database, whose face images have different facial expressions and under different lighting conditions and with various occlusions. It also improves the recognition rate at least by 2.02% (=92.26%-90.24%) on CAS-PEAL face database, whose images of each subject were captured under variable illumination. In general, compared to the traditional KDA, our approach boosts the average recognition rates at least by 3.54% (=89.29%-85.75%), and at least by 3.02% (=89.29%-86.27%) for KDA taking specific class information into condition. Our approach also outperforms KUDV and KODV in recognition rate at least by 1.59% (=89.29%-87.87%), which simply eliminate the redundancy among projective vectors in kernel subspace. Fig.3 shows the average recognition rates of all compared methods on two face databases.

IV. CONCLUSION

In this paper, a new kernel orthogonal projection analysis approach is proposed. It calculates the optimal nonlinear projective vectors class by class with the Fisher criterion. It can make the projective vectors differentiate one class with its adjacent classes, and make them locally orthogonal. Experimental results on AR and CAS-PEAL face databases demonstrate that our approach outperforms several representative feature extraction methods.

ACKNOWLEDGMENT

The work described in this paper was fully supported by the NSFC under Project No. 61073113 and Project No. 61272273, the New Century Excellent Talents of Education Ministry under Project No. NCET-09-0162, the Doctoral Foundation of Education Ministry under Project No. 20093223110001.

REFERENCES

[1] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, K. R. Müller, “Fisher Discriminant Analysis with Kernels,” Proceeding of IEEE Neural Networks for Signal Processing Workshop, pp. 41-48, 1999.

[2] W. Huang and H.Yin, “On nonlinear dimensionality reduction for face recognition,” Image Vision Computing, vol. 30, pp. 355-366, 2012.

[3] G. Baudat, F. Anouar, “Generalized discriminant analysis using a kernel approach,” Neural Computation, vol. 12, pp. 2385-2404, 2000.

[4] Bach, F.R., Jordan, M.I., Kernel Independent Component Analysis. Journal of Machine Learning Research, 3, (2002), 1-48.

[5] W. J. Zeng, X. L. Li, X. D. Zhang, and E. Cheng, “Kernel-based nonlinear discriminant analysis using minimum squared errors criterion for multiclass and undersampled problems,” Signal Processing, vol. 90, pp. 2333-2343, 2010.

[6] D. Cai, X. F. He and J. W. Han, “Speed Up Kernel Discriminant Analysis,” Int. J. Very Large Data Bases, vol. 20, pp. 21-33, 2011.

[7] P. Baggenstoss, “Class-Specific Feature Sets in Classification,” IEEE Trans. Signal Processing, vol. 47, pp. 3428-3432, 1999.

[8] G. Goudelis, S. Zafeiriou, A. Tefas, I. Pitas, “Class-Specific Kernel Discriminant Analysis for Face Verification,” IEEE Trans. Information Forensics and Security, vol. 2, pp. 570–587, 2007.

[9] X. He, S. Yan, Y. Hu, P. Niyogi and H. Zhang, “Face Recognition Using Laplacianfaces,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, pp. 328-340, 2005.

[10] M. Sugiyama, “Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis,” Journal of Machine Learning Research, vol. 8, pp. 1027-1061, 2007.

[11] X. Tao, J. Ye, V. Cherkassky, “Kernel Uncorrelated and Orthogonal Discriminant Analysis: A Unified Approach,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 125-131, 2006.

[12] Z. Liang, P. Shi, “Uncorrelated Discriminant Vectors Using a Kernel Method,” Pattern Recognition, vol. 38, pp. 307-310, 2005.

[13] X. Y. Jing, S. Li, Y. F. Yao, L. S. Bian and J. Y. Yang, “Kernel Uncorrelated Adjacent-class Discriminant Analysis,” Int. Conf. Pattern Recognition, pp. 706-709, 2010.

[14] W. M. Zheng, L. Zhao, C. Zou, “Foley-Sammon optimal discriminant vectors using kernel approach,” IEEE Trans Neural Networks, vol. 16, pp. 1-9, 2005.

[15] A. M. Martinez, R. Benavente, “The AR Face Database,” CVC Technical Report, 1998.

[16] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, D. Zhao, “The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations,” IEEE Trans. System Man and Cybernetics, vol. A38, pp. 149-161, 2008.