A New Kernel Orthogonal Projection Analysis Approach for Face Recognition

Xiaoyuan Jing1,2, Min Li1*, Yongfang Yao1, Songhao Zhu1, Sheng Li1
1College of Automation, Nanjing University of Posts and Telecommunications, Nanjing, China
2State 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 and among feature coordinates extracted in the same class and the nearest neighbor class. Further, the proposed approach makes every projective vector satisfy orthogonal constraint. On the basis of KODV, 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 orthogonal constraint. Zheng et al. [14] put forward a kernel Foley-Sammon optimal discriminant vectors (KODV) method, which subject to 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 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 orthogonal constraint. Besides, 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 \( \phi: \mathbb{R}^d \rightarrow F \) denote a nonlinear mapping. The original sample set \( X = \{X_1, X_2, \ldots, X_n\} \) is injected into the kernel space \( F \) by \( \phi: X_i \rightarrow \phi(x_i), \) and we obtain a set of mapped samples \( \Psi = \{X^\phi_1, X^\phi_2, \ldots, X^\phi_n\}. \)

B. Get adjacent classes of each class

First, calculate the Euclidean distance between any two classes \( X_i^\phi \) and \( X_j^\phi \) in the kernel space as:
\[ d \left( X_\alpha^\beta, X_\gamma^\beta \right) = \left\| m_\alpha^\beta - m_\gamma^\beta \right\|, \]

where \( \| \cdot \| \) represents the 2-norm operator, \( m_\alpha^\beta \) and \( m_\gamma^\beta \) are the mean vectors of \( X_\alpha^\beta \) and \( X_\gamma^\beta \), respectively. After obtaining the Euclidean distance between any two classes, the following distance matrix can be constructed:

\[ G(i, j) = d(X_i^\beta, X_j^\beta). \]

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 \( K_i \) for each class.

C. Construct new scatter matrices with adjacent classes

We reconstructed the between-class scatter matrix \( S^\beta_\alpha \) and the total scatter matrix \( \tilde{S}^\beta_\alpha \) of the \( i^{th} \) class as follows:

\[ S^\beta_\alpha = \left( m_\alpha^\beta - m_\beta^\alpha \right) \left( m_\gamma^\beta - m_\beta^\gamma \right)^T, \]

\[ \tilde{S}^\beta_\alpha = \frac{1}{n_i} \sum_{j=1}^{n_i} (\varphi(x_j) - \bar{m}_\alpha^\beta)(\varphi(x_j) - \bar{m}_\alpha^\beta)^T, \]

\[ + \frac{1}{K_i} \sum_{q=1}^{K_i} 1_{1 \leq q \leq K_i} \sum_{j=1}^{n_q} \theta_q (\varphi(x_j) - \bar{m}_q^\beta)(\varphi(x_j) - \bar{m}_q^\beta)^T, \]

where \( m_\alpha^\beta = \frac{1}{n_i} \sum_{j=1}^{n_i} \varphi(x_j) \), \( \bar{m}_\gamma^\beta = \frac{1}{2} (m_\alpha^\beta + m_\gamma^\beta) \), and the coefficient \( \theta_q \) is defined as

\[ \theta_q = \begin{cases} 1 & \text{if class } q \text{ is adjacent to class } i \\ 0 & \text{otherwise} \end{cases}. \]

We rewrite \( S^\beta_\alpha \) and \( \tilde{S}^\beta_\alpha \) as \( \tilde{S}_K^\alpha = K'K' \) and \( \tilde{S}_K^\alpha = K'W'K' \), where \( K' \) is an \( m \times n \) kernel matrix calculated by using the \( i^{th} \) class and its adjacent classes, \( n_i = n + \sum_{q=1}^{K_i} \theta_q n_q \), \( W' = \text{diag}(w_1', w_2') \), \( w_i' \) is an \( n \times n \) matrix with all terms equal to \( 1/n \), while \( w_2' \) 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 \( \beta_i \) of the first class with the Fisher criterion as follows:

\[ J(\beta_i) = \begin{bmatrix} \beta_i \tilde{S}_K^\alpha \beta_i \\ \beta_i \tilde{S}_K^\alpha \beta_i \end{bmatrix}. \]

According to (3), the rank of \( \tilde{S}_K^\alpha \) is 1. Therefore, \( \beta_i \) is the eigenvector of \( (\tilde{S}_K^\alpha)^{-1} \tilde{S}_K^\alpha \) corresponding to the nonzero eigenvalue.

E. Construct locally orthogonal constraints

Assume that we have obtained the first \( i - 1 \) projective vectors \( (\beta_1, \beta_2, \ldots, \beta_{i-1}) \), \( \beta_i \) is the optimal projective vector of the \( i^{th} \) class. For the \( i^{th} \) class, we select \( K_i \) obtained optimal projective vectors \( (\beta_1, \beta_2, \ldots, \beta_{K_i}) \) subject to locally orthogonal constraints:

\[ \beta_i \beta_m = 0, \quad m = 1, 2, \ldots, K_i \quad \text{and} \quad \beta_i \beta_i = 1, \]

where \( \beta_m \) 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 \( \beta_1 \), we calculate the \( i^{th} \) optimal projective vector \( \beta_i \) with the Fisher criterion and locally orthogonal constraints as following theorem:

**Theorem.** The \( i^{th} \) optimal projective vector \( \beta_i \) \((i \geq 2)\) is the eigenvector corresponding to the nonzero eigenvalue of \( (\tilde{S}_K^\alpha)^{-1} P_i \tilde{S}_K^\alpha \), where \( P_i = I - \frac{1}{n_i} \sum_{q=1}^{K_i} \theta_q \sum_{j=1}^{n_q} (\varphi(x_j) - \bar{m}_q^\beta)(\varphi(x_j) - \bar{m}_q^\beta)^T \),

\[ D_i = \left[ \beta_1, \beta_2, \ldots, \beta_{K_i} \right]^T, \quad \text{and} \quad I = \text{diag}(1, 1, \ldots, 1) \]

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

\[ L(\beta_i) = \beta_i^T \tilde{S}_K^\alpha \beta_i - \lambda (\beta_i^T \tilde{S}_K^\alpha \beta_i - h - \sum_{m=1}^{K_i} \mu_m \beta_i \beta_m), \]

where \( \lambda \) and \( \mu_m (m = 1, \ldots, K_i) \) are Lagrange multipliers.

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

\[ \partial (L(\beta_i))/\partial (\beta_i) = 0. \]

So we have:

\[ 2\tilde{S}_K^\alpha \beta_i - 2\lambda \tilde{S}_K^\alpha \beta_i - \sum_{m=1}^{K_i} \mu_m \beta_i \beta_m = 0. \]

Left multiplying (10) by \( \tilde{S}_K^\alpha (\tilde{S}_K^\alpha)^{-1} \) \((s = 1, 2, \ldots, K_i)\), we obtain:

\[ 2\beta_i^T (\tilde{S}_K^\alpha)^{-1} \tilde{S}_K^\alpha \beta_i - \sum_{m=1}^{K_i} \mu_m \beta_i^T (\tilde{S}_K^\alpha)^{-1} \beta_m = 0. \]

where \( s = 1, 2, \ldots, K_i \).
Let \( U_i = [\mu_1, \mu_2, \cdots, \mu_{k_i}]^\top \),
(12)
\[
D_i = \begin{bmatrix} \beta_1, \beta_2, \cdots, \beta_{k_i} \end{bmatrix}^\top .
\]
(13)

(11) can be represented as follows:
\[
D_i (\tilde{S}_i^\top)^{-1} D_i^\top U_i = 2D_i (\tilde{S}_i^\top)^{-1} \tilde{S}_i \beta_i .
\]
(14)
Thus, we obtain:
\[
U_i = 2(D_i (\tilde{S}_i^\top)^{-1} D_i^\top)^{-1} D_i (\tilde{S}_i^\top)^{-1} \tilde{S}_i \beta_i .
\]
(15)

(10) can be written as:
\[
2\tilde{S}_i \beta_i - 2\lambda \tilde{S}_i \beta_i - D_i^\top U_i = 0 .
\]
(16)
Substituting (15) into (16), we have:
\[
2\tilde{S}_i \beta_i - 2\lambda \tilde{S}_i \beta_i - D_i^\top [2(D_i (\tilde{S}_i^\top)^{-1} D_i^\top)^{-1} D_i (\tilde{S}_i^\top)^{-1} \tilde{S}_i \beta_i] = 0 .
\]
(17)
Hence, we obtain \( P_i \tilde{S}_i \beta_i = \lambda \tilde{S}_i \beta_i \), where \( \beta_i \) is the eigenvector corresponding to the nonzero eigenvalue of \((\tilde{S}_i^\top)^{-1} P_i \tilde{S}_i \), where
\[
P_i = I - D_i D_i^\top (D_i (\tilde{S}_i^\top)^{-1} D_i^\top)^{-1} D_i (\tilde{S}_i^\top)^{-1} .
\]
(18)

G. Feature extraction and classification

A low-dimensional sample set \( Y = (Y^1, Y^2, \cdots, Y^n)^\top \) can be described as
\[
Y^i = \beta_i ^\top \hat{K}_i ,
\]
(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

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 \times 26) samples. All color images are transformed into grey images and each image was scaled to 60\times60 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\times48. 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
\[
k(x, y) = \exp(-||x - y||^2/(2\delta^2))
\]
for the compared kernel methods, and set the parameter \( \delta_i = i \times \delta \), \( i = 1, \cdots, 20 \), where \( \delta \) 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>
<thead>
<tr>
<th>Method</th>
<th>Recognition Rates (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KDA</td>
<td>82.25</td>
</tr>
<tr>
<td>CSKDA</td>
<td>83.07</td>
</tr>
<tr>
<td>KLFDA</td>
<td>82.61</td>
</tr>
<tr>
<td>LPP</td>
<td>77.50</td>
</tr>
<tr>
<td>KUDV</td>
<td>84.60</td>
</tr>
<tr>
<td>KODV</td>
<td>84.34</td>
</tr>
<tr>
<td>Our Approach</td>
<td>86.33</td>
</tr>
</tbody>
</table>

TABLE I. AVERAGE RECOGNITION RATES (%) OF COMPARED METHODS AND OUR PROPOSED APPROACH ON TWO FACE DATABASES.
Our Approach

In the experiment, $K_1$ is set as 35 on AR face database. $K_1$ is determined by the following strategy: set $K_1$ 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 $K_1 = 5$. 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 1.73% (=86.33%-84.60%) on AR 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 2.02% (=92.26%-90.24%) on CAS-PEAL face database, whose images of each subject were captured under variable illumination.

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.

REFERENCES


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.