Kernel Uncorrelated Adjacent-class Discriminant Analysis

Xiaoyuan Jing[1,*], Sheng Li[1], Yongfang Yao[1], Lusha Bian[1], Jingyu Yang[2]
[1]School of Automation, Nanjing University of Posts and Telecommunications
[2]School of Computer Science, Nanjing University of Science and Technology
[*]Jingxy@njupt.edu.cn

Abstract—**In this paper, a kernel uncorrelated adjacent-class discriminant analysis (KUADA) approach is proposed for image recognition. The optimal nonlinear discriminant vector obtained by this approach can differentiate one class and its adjacent classes, i.e., its nearest neighbor classes, by constructing the specific between-class and within-class scatter matrices in kernel space using the Fisher criterion. In this manner, KUADA acquires all discriminant vectors class by class. Furthermore, KUADA makes every discriminant vector satisfy locally statistical uncorrelated 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 discriminant methods.**

Keywords-adjacent classes; locally statistical uncorrelated constraints; kernel uncorrelated adjacent-class discriminant analysis (KUADA)

I. INTRODUCTION

Discriminant analysis is an important research topic in the field of pattern recognition. In order to solve the nonlinear classification problems, many kernel discrimination algorithms have been presented, such as Kernel discriminant analysis (KDA) [1]. However, KDA can not extract the most discriminative features of a specific class because every achieved discriminant vector extracts discriminative information from the whole sample set. To solve this problem, P. Baggenstoss proposed a class-specific idea that each class has its own feature sets and designed the probabilistic classifiers [2]. Class-specific kernel discriminant analysis (CSKDA) [3] applies this idea to face verification. For each specific class, it acquires a set of discriminant 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, with respect to a specific class, we think that it is unnecessary to use all samples of the sample set to construct the between-class scatter matrix of this class.

Some feature extraction methods considering the local structure of data have been proposed, such as locality preserving projections (LPP) [4] and kernel local Fisher discriminant analysis (KLFDA) [5]. LPP finds a linear map that preserves local neighborhood information of each sample. It is an unsupervised method, so it has no direct connection to classification. KLFDA takes local structure of the data into account so the multimodal data can be embedded appropriately.

In this paper, we first propose a kernel adjacent-class discriminant analysis (KADA) approach. Unlike LPP, KADA preserves local neighborhood information of each class, not each sample. The optimal discriminant vector obtained by KADA can differentiate one class and its adjacent classes by constructing the corresponding between-class and within-class scatter matrices in kernel space and using the Fisher criterion. In this manner, KADA acquires all optimal discriminant vectors class by class. Different from CSKDA, KADA calculates the between-class scatter matrix using a small number of samples that belong to the adjacent classes of a specific class. And different from KDA and KLFDA, KADA extracts discriminative features class by class.

In many applications, it is desirable to eliminate the redundancy among discriminant vectors. Uncorrelated optimal discriminant vectors (UODV) can realize this aim since it makes each discriminant vector satisfy statistical uncorrelated constraints [6]. Kernel uncorrelated discriminant analysis (KUDA) [7] and GSVD-based KUDA [8] methods were proposed to realize UODV in the kernel space. Enlightened by UODV, we propose a kernel uncorrelated adjacent-class discriminant analysis (KUADA) approach. KUADA makes every discriminant vector satisfy locally statistical uncorrelated constraints by using the corresponding class and part of its most adjacent classes, and gets an optimal discriminant transform in nonlinear space.

II. KERNEL ADJACENT-CLASS DISCRIMINANT ANALYSIS (KADA)

A. KDA

Assume that the original sample set $X=\{x_1,x_2,\cdots,x_N\}$ is composed of c classes, and there are n_i training samples in the i^{th} class. For a given nonlinear mapping function φ, the original samples can be mapped into the kernel space F, $\varphi: x \to \varphi(x)$.Suppose that S_b^{φ} and S_t^{ϕ} are the between-class scatter and total scatter matrices defined in F, KDA finds a projection transform α to maximize the following function:

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.178

$$J(\alpha)=\frac{\left|\alpha^T S_b^{\varphi}\alpha\right|}{\left|\alpha^T S_t^{\varphi}\alpha\right|}. \quad (1)$$

where $S_t'^{\varphi}=KK$ and $S_b'^{\varphi}=K\mathrm{W}K$, K is an $N\times N$ kernel matrix, $\mathrm{W}=diag(\mathrm{w}_1,\mathrm{w}_2,...,\mathrm{w}_c)$, w_i is an $n_i\times n_i$ matrix with all terms equal to $1/n_i$.

B. KADA

We realize KADA by the following four steps:

1). Map to the kernel space and Get adjacent classes.

Let $\phi:R^d\to F$ denotes 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 $\Psi=\{X_1^{\varphi},X_2^{\varphi},\cdots,X_c^{\varphi}\}$. Firstly, we compute the Euclidean distance between any two classes X_i^{φ} and X_j^{φ} in the kernel space as:

$$d\left(X_i^{\varphi},X_j^{\varphi}\right)=\left\|m_i^{\varphi}-m_j^{\varphi}\right\|, \quad (2)$$

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. We construct a distance matrix $G(i,j)=d(X_i^{\varphi},X_j^{\varphi})$.

Then, we sort G in the ascending order. For the i^{th} class, we can get its nearest neighbor classes with the smallest between-class distances. These classes are regarded as the adjacent classes of the i^{th} class. In this paper, we set the number of adjacent classes as the same value K_1 for every class.

2). Construct the new scatter matrices in kernel space.

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

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

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

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

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

3). Calculate the discriminant vector of class i.

According to Eq. (1), for the i^{th} class, we calculate the discriminant vector α_i which maximizes the following function:

$$J(\alpha_i)=\frac{\left|\alpha_i^T\tilde{S}_b^{i\varphi}\alpha_i\right|}{\left|\alpha_i^T\tilde{S}_t^{i\varphi}\alpha_i\right|}, \quad (5)$$

where $\tilde{S}_t^{i\varphi}=K^iK^i$, $\tilde{S}_b^{i\varphi}=K^i\mathrm{W}^iK^i$, 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 an $m_i\times m_i$ 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 $1/(m_i-n_i)$.

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

4). Obtain all the discriminant vector class by class.

We repeat Steps *1-3* and finally obtain c discriminant vectors class by class. Thus, the new training samples $Y_{KADA}=\left(Y_{KADA}^1,Y_{KADA}^2,...,Y_{KADA}^c\right)^T$ can be described as

$$Y_{KADA}^i=\alpha_i^T\hat{K}^{iT}, \quad (6)$$

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.

III. KERNEL UNCORRELATED ADA (KUADA)

A. UODV

UODV achieves a group of optimal discriminant vectors which can satisfy both the Fisher criterion and the following statistical uncorrelated constraints:

$$w_i^TS_tw_j=0,\quad 1\le j\le(i-1), \quad (7)$$

where S_t is the total scatter matrix of sample set.

According to the improved UODV algorithm [7], the first optimal discriminant vector w_1 is obtained by maximizing Eq. (1). Then, UODV gives the following theorem:

Lemma 1. The i^{th} optimal discriminant vector w_i ($i\ge 2$) is the eigenvector corresponding to the maximal eigenvalue of the equation:

$$PS_bw_i=\lambda S_tw_i, \quad (8)$$

where $P=I-S_tD^T(DS_tD^T)^{-1}D$, $D=[w_1,w_2,\cdots,w_{i-1}]^T$ and $I=diag(1,1,\cdots,1)$.

B. KUADA Approach

We realize KUADA by following two steps:

1). Construct locally statistical uncorrelated constraints.

Assume that the first $i-1$ eigenvectors $(\alpha_1, \alpha_2, \cdots, \alpha_{i-1})$ of KUADA have been obtained, α_i is the optimal discriminant vector of the i^{th} class. For the i^{th} class, KUADA selects K_2 obtained optimal discriminant vectors $(\alpha_{j1}, \alpha_{j2}, \cdots, \alpha_{jK_2})$ to satisfy locally statistical uncorrelated constraints:

$\alpha_i^T \tilde{S}_t^{i\varphi} \alpha_{jm} = 0, m = 1, 2, \cdots, K_2$ and $\alpha_i^T \tilde{S}_t^{i\varphi} \alpha_i = b$, (9)

where α_{jm} corresponds to one of most adjacent classes of the i^{th} class, and b is a constant.

In the experiment, the value of K_2 is set to be smaller than the value of K_1, i.e., $K_2 < K_1$. We only use part of its adjacent classes of each class to construct locally uncorrelated constraints. Therefore, the constraints of KUADA are different from those of UODV, since α_i of KUADA does not need to be statistically uncorrelated with every obtained $\alpha_j (1 \le j \le i-1)$, and KUADA uses $\tilde{S}_t^{i\varphi}$ to replace S_t defined in Eq. (8).

2). Calculate optimal discriminant vectors.

The first discriminant vector α_1 of KUADA is same as that of KADA. Then, KUADA calculates discriminant vectors using the following theorem:

Theorem 1. The i^{th} optimal discriminant vector α_i ($i \ge 2$) is the eigenvector corresponding to the nonzero eigenvalue of $(\tilde{S}_t^{i\varphi})^{-1} P_i \tilde{S}_b^{i\varphi}$, where

$P_i = I - S_t^{i\varphi} D_i^T (D_i S_t^{i\varphi} D_i^T)^{-1} D_i$, $D_i = \left[\alpha_1, \alpha_2, \cdots, \alpha_{K_2}\right]^T$,

and $I = diag\left(1, 1, \cdots, 1\right)$. (10)

Proof. Use the Lagrange multipliers method to express Eq. (5) including all the locally statistical uncorrelated constraints in Eq. (9), we have:

$$L(\alpha_i) = \alpha_i^T \tilde{S}_b^{i\varphi} \alpha_i - \lambda(\alpha_i^T \tilde{S}_t^{i\varphi} \alpha_i - b) - \sum_{m=1}^{K2} \mu_m \alpha_i^T \tilde{S}_t^{i\varphi} \alpha_{jm} , \quad (11)$$

where λ and $\mu_m (m = 1, ..., K_2)$ are Lagrange multipliers.

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

$$\partial\left(L(\alpha_i)\right) / \partial\left(\alpha_i\right) = 0 \cdot \quad (12)$$

So we have:

$$2\tilde{S}_b^{i\varphi} \alpha_i - 2\lambda \tilde{S}_t^{i\varphi} \alpha_i - \sum_{m=1}^{K_2} \mu_m \tilde{S}_t^{i\varphi} \alpha_{jm} = 0 \cdot \quad (13)$$

Left multiplying Eq. (13) by $\alpha_{js}^T (s = 1, 2, \cdots, K_2)$, we obtain K_2 equations:

$$2\alpha_{js}^T \tilde{S}_b^{i\varphi} \alpha_i - \sum_{m=1}^{K2} \mu_m \alpha_{js}^T \tilde{S}_t^{i\varphi} \alpha_{jm} = 0, \quad s = 1, 2, \cdots, K_2 \cdot \quad (14)$$

Let $U_i = [\mu_1, \mu_2, \cdots, \mu_{K_2}]^T$, $D_i = \left[\alpha_1, \alpha_2, \cdots, \alpha_{K_2}\right]^T$. The above equations can be represented in the form of matrix:

$$D_i \tilde{S}_t^{i\varphi} D_i^T U_i = 2 D_i \tilde{S}_b^{i\varphi} \alpha_i . \quad (15)$$

Thus, we obtain:

$$U_i = 2(D_i \tilde{S}_t^{i\varphi} D_i^T)^{-1} D_i \tilde{S}_b^{i\varphi} \alpha_i , \quad (16)$$

Eq. (16) can be written as:

$$2\tilde{S}_b^{i\varphi} \alpha_i - 2\lambda \tilde{S}_t^{i\varphi} \alpha_i - \tilde{S}_t^{i\varphi} D_i^T U_i = 0 . \quad (17)$$

Substituting (16) into (17), we have:

$$2\tilde{S}_b^{i\varphi} \alpha_i - 2\lambda \tilde{S}_t^{i\varphi} \alpha_i - \tilde{S}_t^{i\varphi} D_i^T [2(D_i \tilde{S}_t^{i\varphi} D_i^T)^{-1} D_i \tilde{S}_b^{i\varphi} \alpha_i] = 0 . \quad (18)$$

Hence, we obtain $P_i \tilde{S}_b^{i\varphi} \alpha_i = \lambda \tilde{S}_t^{i\varphi} \alpha_i$, where α_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 is defined in Eq. (10). Proof is over.

Theorem 1 and Lemma 1 show that the realization of KUADA and UODV are different: (i) KUADA constructs specific total scatter matrix $\tilde{S}_t^{i\varphi}$ and between-class scatter matrix $\tilde{S}_b^{i\varphi}$ for every class, while UODV uses identical total scatter matrix S_t and between-class scatter matrix S_b for all classes; (ii) The matrix P_i constructed by KUADA is different from the matrix P constructed by UODV.

IV. EXPERIMENTAL RESULTS

In the experiment, we use the AR and CAS-PEAL face databases. The AR database contains 119 individuals, each 26 images with size 60×60 [3]. All image samples of one subject are shown in Figure 1. We in turn choose following 1-8 representative images of every subject as the training samples: (1), (14), (2), (5), (8), (11), (17) and (19). The remainders are chosen as the testing samples. The CAS-PEAL database [9] 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. Figure 2 shows 10 images of an individual of the CAS-PEAL face database. We use the first 2-6 images of each person as training samples. The remainder are regarded as testing samples.

Figure 1. Demo images of one subject from AR database

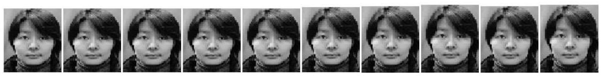

Figure 2. Demo images of one subject from CAS-PEAL database

In the experiment, K_1 is set as 45 for the AR database and 35 for the CAS-PEAL database, respectively. K_2 is determined by the following strategy: set K_2 as the number of most adjacent classes of each class, where the optimal discriminant vectors of these classes have been acquired; if the number is more than 5, then set $K_2 = 5$.

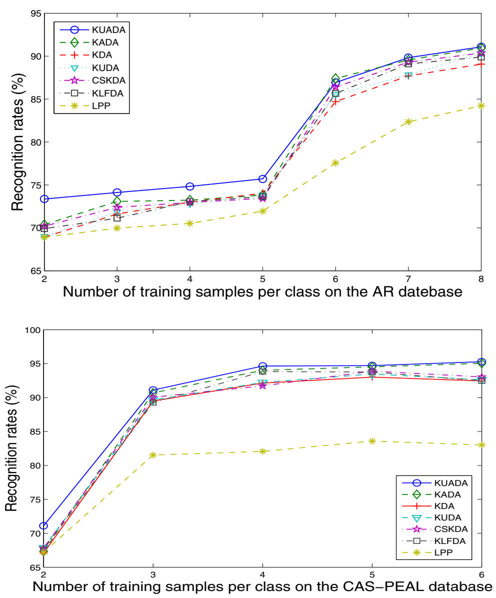

Figure 3. Recognition rates of all compared methods

Table 1. Comparison of Average Recognition Rates

Methods	Average recognition rates (%)	
	AR	CAS-PEAL
KUADA	80.84	89.37
KADA	79.77	88.33
KDA	78.43	86.84
KUDA	78.75	87.17
CSKDA	79.30	87.30
KLFDA	78.91	87.40
LPP	75.07	78.49

Figure 3 shows the recognition rates of KUADA, KADA and five related methods including KDA, KUDA, CSKDA, KLFDA and LPP on the AR and CAS-PEAL databases, respectively. KUADA and KADA perform better than the compared methods in all cases. Table 1 shows that KUADA improves the average recognition rate at least by 1.54% (=80.84%-79.30%) in contrast with other methods. Furthermore, KUADA outperforms KADA.

V. CONCLUSION

In this paper, we propose a KUADA approach for nonlinear facial feature extraction and recognition. Experimental results on AR and CAS-PEAL face databases demonstrate that KUADA outperforms several representative kernel discriminant methods.

ACKNOWLEDGEMENTS

The work described in this paper was fully supported by the NSFC under Project No.60772059, the Doctoral Foundation of Ministry of Education under Project No. 20093223110001, the Qin-Lan Engineering Academic Leader of Jiangsu Province, the Foundation of Jiangsu Province Universities under Project No.09KJB510011, the Foundation of NUPT under Project No. NY207027, Project No.NY208051.

REFERENCES

[1] S. Mika, et al. Fisher Discriminant Analysis with Kernels. *Proceeding of IEEE Neural Networks for Signal Processing Workshop*, pp. 41-48, 1999.

[2] P. Baggenstoss. Class-Specific Feature Sets in Classification. *IEEE Trans. Signal Processing*, 47(12): 3428-3432, 1999.

[3] G. Goudelis, et al. Class-Specific Kernel Discriminant Analysis for Face Verification. *IEEE Trans. Inf. Forensics Security*, 2(3):570–587, 2007.

[4] X.F. He, et al. Face Recognition Using Laplacianfaces. *IEEE Trans. PAMI*, 27(3):328-340, 2005.

[5] M. Sugiyama. Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis. *J. of Mach. Learn. Res.*, 8:1027-1061, 2007.

[6] X.Y. Jing, D. Zhang and Z. Jin. UODV Improved Algorithm and Generalized Theory. *Pattern Recognition*, 36(11):2593–2602, 2003.

[7] Z. Liang and P. Shi. Uncorrelated discriminant vectors using a kernel method. *Pattern Recognition*, 38(2):307-310, 2005.

[8] X. Tao and J. Ye. Kernel Uncorrelated and Orthogonal Discriminant Analysis: A Unified Approach. *IEEE Conference on CVPR*, pp. 125-131, 2006.

[9] W. Gao, et al. The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations. *IEEE Trans. SMC-A*, 38(1):149-161, 2008.