

LETTER

Facial Image Recognition Based on a Statistical Uncorrelated Near Class Discriminant Approach

Sheng LI[†], *Student Member*, Xiao-Yuan JING^{†a)}, Lu-Sha BIAN[†], Shi-Qiang GAO[†], Qian LIU[†],
and Yong-Fang YAO[†], *Nonmembers*

SUMMARY In this letter, a statistical uncorrelated near class discriminant (SUNCD) approach is proposed for face recognition. The optimal discriminant vector obtained by this approach can differentiate one class and its near classes, i.e., its nearest neighbor classes, by constructing the specific between-class and within-class scatter matrices and using the Fisher criterion. In this manner, SUNCD acquires all discriminant vectors class by class. Furthermore, SUNCD makes every discriminant vector satisfy locally statistical uncorrelated constraints by using the corresponding class and part of its most neighboring classes. Experiments on the public AR face database demonstrate that the proposed approach outperforms several representative discriminant methods.

key words: *near classes, locally statistical uncorrelated constraints, statistical uncorrelated near class discriminant (SUNCD), face recognition*

1. Introduction

Feature extraction is an important research topic in the field of pattern recognition. Linear discriminant analysis (LDA) is a widely-used feature extraction method which obtains discriminant vectors by using the Fisher criterion [1]. However, LDA can not extract the most discriminative features of a specific class because every discriminant vector achieved by LDA 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 linear discriminant analysis (CSLDA) [3] and class-specific kernel discriminant analysis (CSKDA) [4] apply this idea to discriminant feature extraction. For each specific class, they acquire a group 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. Then, CSLDA puts the discriminant vectors of all classes together and constructs a discriminant transform for classification, while CSKDA separately uses the discriminant vectors of each class for face verification. 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 pre-

serving projections (LPP) [5] and local Fisher discriminant analysis (LFDA) [6]. LPP finds a linear map that preserves local neighborhood information of each sample, and its criterion is to minimize the local scatter of mapped samples. LPP is an unsupervised method, so it has no direct connection to classification. LFDA uses local neighborhood information to construct weighted between-class and within-class scatter matrices and then performs discriminant analysis.

In this letter, we first propose a near class discriminant (NCD) approach. Unlike LPP, NCD preserves local neighborhood information of each class, not each sample. The optimal discriminant vector obtained by NCD can differentiate one class and its near classes, i.e., its nearest neighbor classes, by constructing the corresponding between-class and within-class scatter matrices and using the Fisher criterion. In this manner, NCD acquires all optimal discriminant vectors class by class. Different from class-specific discriminant methods (CSLDA and CSKDA), NCD calculates the between-class scatter matrix using a small number of samples that belong to the near classes of a specific class. And different from LDA, LPP and LFDA, NCD extracts discriminative features class by class.

In many applications, it is desirable to eliminate the redundancy among discriminant vectors. Uncorrelated optimal discriminant vectors (UODV) method can realize this aim since it makes each discriminant vector satisfy statistical uncorrelated constraints [7]. Enlightened by UODV, we further propose an statistical uncorrelated NCD (SUNCD) approach. SUNCD makes every discriminant vector satisfy locally statistical uncorrelated constraints by using the corresponding class and part of its most neighboring classes, and gets an optimal discriminant transform.

The rest of this paper is organized as follows: In Sect. 2, we outline LDA and describe NCD. In Sect. 3, we outline UODV and describe the SUNCD approach. In Sect. 4, experiments on the public AR face database are performed. Finally, we offer the conclusions in Sect. 5.

2. Near Class Discriminant (NCD) Approach

2.1 LDA

Assume that X is a sample set having N training samples and c classes $\{X_1, X_2, \dots, X_c\}$. In LDA, the between-class scatter matrix S_B , the within-class scatter matrix S_W and the total-

Manuscript received September 7, 2009.

Manuscript revised December 16, 2009.

[†]The authors are with the Nanjing University of Posts and Telecommunications, Nanjing, 210003 China.

a) E-mail: jingxy@njupt.edu.cn

DOI: 10.1587/transinf.E93.D.934

scatter matrix S_T are defined as:

$$S_B = \sum_{i=1}^c (m_i - \bar{m})(m_i - \bar{m})^T, \quad (1)$$

$$S_W = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - m_i)(x_{ij} - m_i)^T, \quad (2)$$

$$S_T = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \bar{m})(x_{ij} - \bar{m})^T. \quad (3)$$

where n_i is the number of samples of the i^{th} class, x_{ij} is a sample of X , m_i is the mean of the i^{th} class and \bar{m} is the mean of all training samples.

LDA uses the Fisher criterion to find a discriminant transform W and maximizes the following function:

$$J(W) = \frac{|W^T S_B W|}{|W^T S_W W|}. \quad (4)$$

In this paper, we use an equivalent form of Eq. (4) [1]:

$$J(W) = \frac{|W^T S_B W|}{|W^T S_T W|}. \quad (5)$$

Generally, W is composed of the eigenvectors of $S_T^{-1} S_B$ corresponding to the non-zero eigenvalues.

2.2 NCD

We realize NCD by following four steps:

Step1. Get near classes.

We first compute the Euclidean distance between any two classes X_i and X_j as:

$$d(X_i, X_j) = \|m_i - m_j\|, \quad (6)$$

where $\|\cdot\|$ represents the 2-norm operator, m_i and m_j are the mean vectors of X_i and X_j , respectively. We construct a distance matrix G , where $G(i, j) = d(X_i, X_j)$.

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 $d(X_i, X_j)$. These classes are regarded as the near classes of the i^{th} class. In this paper, we set the number of near classes as the same value K_1 for every class.

Step2. Construct the scatter matrices.

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

$$S_b^i = (m_i - m_i^k)(m_i - m_i^k)^T, \quad (7)$$

$$S_t^i = \frac{1}{n_i + \sum_{q=1}^c w_{iq} n_q} \left(\sum_{j=1}^{n_i} (x_{ij} - \bar{m}_i)(x_{ij} - \bar{m}_i)^T + \sum_{q=1}^c \sum_{l=1}^{n_q} w_{iq} (x_{ql} - \bar{m}_i)(x_{ql} - \bar{m}_i)^T \right), \quad (8)$$

where m_i is the mean of the i^{th} class, $m_i^k =$

$\frac{1}{K_1} \sum_{q=1}^c \frac{1}{n_q} \sum_{l=1}^{n_q} w_{iq} x_{ql}$, $\bar{m}_i = \frac{1}{K_1+1} (m_i + K_1 m_i^k)$, and the coefficient w_{iq} is defined as:

$$w_{iq} = \begin{cases} 1 & \text{if class } q \text{ is neighboring to class } i \\ 0 & \text{otherwise} \end{cases}$$

Step3. Calculate the discriminant vector of class i .

For the i^{th} class, we calculate the discriminant vector ϕ_i which maximizes the following function:

$$J(\phi_i) = \frac{|\phi_i^T S_b^i \phi_i|}{|\phi_i^T S_t^i \phi_i|}. \quad (9)$$

According to Eq. (7), the rank of S_b^i is 1. Therefore, ϕ_i is the eigenvector of $(S_t^i)^{-1} S_b^i$ corresponding to the nonzero eigenvalue.

Step4. Obtain all discriminant vectors class by class.

We repeat Steps 1-3 and obtain c discriminant vectors class by class. The discriminant transform W of NCD is composed of these vectors, that is, $W = [\phi_1, \phi_2, \dots, \phi_c]$.

3. Statistical Uncorrelated NCD (SUNCD) Approach

3.1 UODV

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

$$\phi_i^T S_T \phi_j = 0, \quad 1 \leq j \leq (i-1), \quad (10)$$

where S_T defined in Eq. (3) is the total scatter matrix of sample set.

According to improved UODV algorithm [7], the first optimal discriminant vector ϕ_i is obtained by maximizing Eq. (5). Then, UODV gives the following theorem:

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

$$PS_B \phi_i = \lambda S_T \phi_i, \quad (11)$$

where $P = I - S_T D^T (D S_T D^T)^{-1} D$, $D = [\phi_1, \phi_2, \dots, \phi_{i-1}]^T$ and $I = \text{diag}(1, 1, \dots, 1)$.

3.2 SUNCD Approach

We realize the statistical uncorrelated NCD (SUNCD) approach by following two steps:

Step1. Construct locally statistical uncorrelated constraints.

Assume that the first $i-1$ optimal discriminant vectors $(\phi_1, \phi_2, \dots, \phi_{i-1})$ of SUNCD have been obtained, and ϕ_i is the optimal discriminant vector of the i^{th} class. For the i^{th} class, SUNCD selects K_2 obtained optimal discriminant vectors $(\phi_{j1}, \phi_{j2}, \dots, \phi_{jK2})$ to satisfy locally statistical

uncorrelated constraints:

$$\phi_i^T S_i^i \phi_{jm} = 0, \quad m = 1, 2, \dots, K_2, \quad (12)$$

$$\text{and } \phi_i^T S_i^i \phi_i = a, \quad (13)$$

where ϕ_{jm} corresponds to one of most neighboring classes of the i^{th} class, and a 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 near classes of each class to construct locally statistical uncorrelated constraints. Therefore, the constraints of SUNCD are different from those of UODV, since ϕ_i of SUNCD does not need to be statistically uncorrelated with every obtained $\phi_j (1 \leq j \leq i-1)$, and SUNCD uses S_i^i defined in Eq. (8) to replace S_T defined in Eq. (3).

Step2. Calculate optimal discriminant vectors.

The first discriminant vector ϕ_1 of SUNCD is same as that of NCD. ϕ_1 is the eigenvector of $(S_i^1)^{-1} S_b^1$ corresponding to the nonzero eigenvalue. Then, SUNCD calculates optimal discriminant vectors using the following theorem:

Theorem 1: The i^{th} optimal discriminant vector $\phi_i (i \geq 2)$ is the eigenvector corresponding to the nonzero eigenvalue of $(S_i^i)^{-1} P_i S_b^i$, where

$$P = I - S_i^i D_i^T (D_i S_i^i D_i^T)^{-1} D_i, \quad D_i = [\phi_1, \phi_2, \dots, \phi_{K_2}]^T \\ \text{and } I = \text{diag}(1, 1, \dots, 1). \quad (14)$$

Proof. Using the Lagrange multipliers and locally statistical uncorrelated constraints in Eqs. (12–13) to transform Eq. (9), we have:

$$L(\phi_i) = \phi_i^T S_b^i \phi_i - \lambda (\phi_i^T S_i^i \phi_i - a) \\ - \sum_{m=1}^{K_2} \mu_m \phi_i^T S_i^i \phi_{jm}, \quad (15)$$

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

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

$$\frac{\partial (L(\phi_i))}{\partial (\phi_i)} = 0. \quad (16)$$

So we have:

$$2S_b^i \phi_i - 2\lambda S_i^i \phi_i - \sum_{m=1}^{K_2} \mu_m S_i^i \phi_{jm} = 0. \quad (17)$$

Multiplying Eq. (16) by $\phi_{js}^T (s = 1, 2, \dots, K_2)$, we obtain K_2 equations:

$$2\phi_{js}^T S_b^i \phi_i - \sum_{m=1}^{K_2} \mu_m \phi_{js}^T S_i^i \phi_{jm} = 0, \quad s = 1, \dots, K_2. \quad (18)$$

Let $U_i = [\mu_1, \mu_2, \dots, \mu_{K_2}]^T$, $D_i = [\phi_1, \phi_2, \dots, \phi_{K_2}]^T$. The above equations can be represented in the form of matrix:

$$D_i S_i^i D_i^T U_i = 2D_i S_b^i \phi_i. \quad (19)$$

Thus, we obtain:

$$U_i = 2(D_i S_i^i D_i^T)^{-1} D_i S_b^i \phi_i, \quad (20)$$

Eq. (17) can be written as:

$$2S_b^i \phi_i - 2\lambda S_i^i \phi_i - S_i^i D_i^T U_i = 0. \quad (21)$$

Substituting (20) into (21), we have:

$$2S_b^i \phi_i - 2\lambda S_i^i \phi_i - S_i^i D_i^T [2(D_i S_i^i D_i^T)^{-1} D_i S_b^i \phi_i] = 0. \quad (22)$$

Hence, we obtain $P_i S_b^i \phi_i = \lambda S_i^i \phi_i$, that is, ϕ_i is the eigenvector corresponding to the nonzero eigenvalue of $(S_i^i)^{-1} P_i S_b^i$, where P_i is defined in Eq. (14). Proof is over.

Theorem 1 and Lemma 1 show that the realization of SUNCD and UODV are different: (i) SUNCD constructs specific total scatter matrix S_i^i and between-class scatter matrix S_b^i 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 SUNCD is different from the matrix P constructed by UODV.

4. Experimental Results

In the experiment, we use the public AR face database. This database contains 119 individuals, each 26 images with size 60×60 [8]. All image samples of one subject are shown in Fig. 1. The major differences between them are the expression, illumination, position, pose and sampling time. In order to effectively evaluate the impact of different variations to the recognition results, we in turn choose following 1–10 representative images of every subject as the training samples: (1), (14), (2), (5), (8), (11), (17), (19), (23) and (25). And the remainders are chosen as the testing samples.

In the experiment, the value of K_1 is selected such that the best classification performance is obtained. On the AR face database, K_1 is set as 30. K_2 is determined by using the following strategy: set K_2 as the number of most neighboring classes of each class, where the optimal discriminant vectors of these classes have been acquired; and if the number is more than 10, then set $K_2 = 10$. Since the image samples are high-dimensional, we first use the PCA transform to reduce inputted feature dimension, and then apply various methods to extracting discriminative features.

Figure 2 shows the recognition rates of SUNCD, NCD and five related methods including LDA, UODV, CSLDA, LPP and LFDA on the AR face database. SUNCD and NCD perform better than other compared methods in all cases. Furthermore, SUNCD outperforms NCD.

Table 1 shows the average recognition rates of all compared methods. Compared with NCD, LDA, CSLDA, LPP



Fig. 1 Demo images of one subject from the AR database.

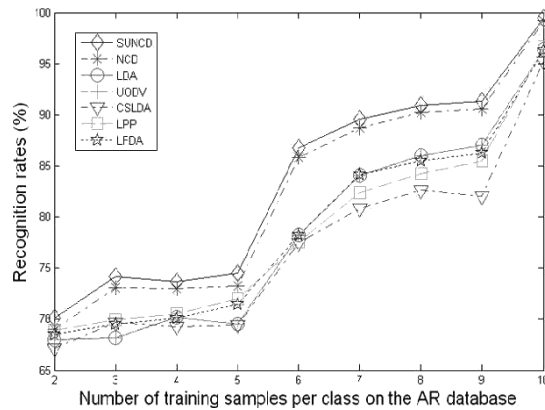


Fig. 2 Recognition rates of all compared methods.

Table 1 Average recognition rates of compared methods.

Methods	Average recognition rates(%)
SUNCD	83.369%
NCD	82.281%
LDA	78.620%
UODV	78.620%
CSLDA	77.079%
LPP	78.625%
LFDA	78.903%

and LFDA, SUNCD separately improves average recognition rates by 1.088% ($= 83.369\% - 82.281\%$), 4.749% ($= 83.369\% - 78.620\%$), 6.290% ($= 83.369\% - 77.079\%$), 4.744% ($= 83.369\% - 78.625\%$), and 4.466% ($= 83.369\% - 78.903\%$). Here, UODV obtains the same recognition results as LDA on these two databases. The reason is that if the nonzero Fisher discriminant values are mutually unequal, then UODV is equivalent to LDA. This has been proved in Ref. [7].

5. Conclusions

In this letter, we propose an SUNCD approach for facial feature extraction and recognition. SUNCD obtains a group of optimal discriminant vectors, which can differentiate one class and its near classes by virtue of constructing the

specific scatter matrices class by class. Moreover, SUNCD makes the achieved discriminant vectors satisfy locally statistical uncorrelation, which is demonstrated to be a favorable theoretical property. Experimental results on the public AR face database demonstrate that SUNCD outperforms several representative discriminant methods including LDA, UODV, CSLDA, LPP and LFDA, and improves the average recognition rates at least by 4.466%.

Acknowledgements

The work described in this paper was fully supported by the National Natural Science Foundation of China (NSFC) under Project No.60772059, the Qin-Lan Engineering Academic Leader of Jiangsu Province, the Natural Science Research Foundation of Jiangsu Province Universities under Project No.07KJB520081, 09KJB510011, and the Research Foundation of Nanjing University of Posts and Telecommunications under Project No.NY207027, Project No.NY208051.

References

- [1] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, New York, 1990.
- [2] P. Baggenstoss, "Class-specific feature sets in classification," IEEE Trans. Signal Process., vol.47, no.12, pp.3428–3432, 1999.
- [3] X.H. Chen and S.C. Chen, "Class-specific linear discriminant analysis," J. Chinese Computer Systems, vol.25, no.9, pp.894–897, 2008.
- [4] G. Goudelis, S. Zafeiriou, A. Tefas, and I. Pitas, "Class-specific kernel discriminant analysis for face verification," IEEE Trans. Inf. Forensics Security, vol.2, no.3, pp.570–587, 2007.
- [5] X.F. He, S.C. Yan, Y.X. Hu, P. Niyogi, and H.J. Zhang, "Face recognition using laplacianfaces," IEEE Trans. Pattern Anal. Mach. Intell., vol.27, no.3, pp.328–340, 2005.
- [6] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," J. Mach. Learn. Res., vol.8, pp.1027–1061, 2007.
- [7] X.Y. Jing, D. Zhang, and Z. Jin, "UODV improved algorithm and generalized theory," Pattern Recognit., vol.36, no.11, pp.2593–2602, 2003.
- [8] X.Y. Jing, Y.F. Yao, D. Zhang, J.Y. Yang, and M. Li, "Face and palmprint pixel level fusion and kernel DCV-RBF classifier for small sample biometrics recognition," Pattern Recognit., vol.40, no.12, pp.3209–3224, 2007.