

LETTER

Face Recognition Based on Nonlinear DCT Discriminant Feature Extraction Using Improved Kernel DCV

Sheng LI[†], Student Member, Yong-fang YAO[†], Xiao-yuan JING^{†a)}, Heng CHANG[†], Shi-qiang GAO[†], David ZHANG^{††}, and Jing-yu YANG^{†††}, Nonmembers

SUMMARY This letter proposes a nonlinear DCT discriminant feature extraction approach for face recognition. The proposed approach first selects appropriate DCT frequency bands according to their levels of nonlinear discrimination. Then, this approach extracts nonlinear discriminant features from the selected DCT bands by presenting a new kernel discriminant method, i.e. the improved kernel discriminative common vector (KDCV) method. Experiments on the public FERET database show that this new approach is more effective than several related methods.

key words: DCT frequency bands selection, the improved KDCV, nonlinear DCT feature extraction, face recognition

1. Introduction

Discrete cosine transform (DCT) is a widely used image processing technique [1], and discriminant analysis is an effective image feature extraction and recognition technique [2]. To date, many linear discriminant methods have been put forward, such as linear discriminant analysis (LDA) [3], direct LDA (DLDA) [4], and discriminative common vector (DCV) [5]. In our previous work, we combined the DCT and linear discriminant techniques and presented a DCT-LDA face recognition method which outperforms some conventional linear discriminant methods [6]. As an extension of linear discriminant technique, the kernel-based nonlinear discriminant analysis technique has now been widely applied to the field of pattern recognition. Baudat et al. developed a commonly used generalized discriminant analysis (GDA) method for nonlinear discrimination [7]. Jing et al. put forward a kernel DCV (KDCV) method [8]. Shen et al. combined Gabor wavelets and GDA for face identification and verification [9]. In this paper, we develop DCT-LDA and propose a nonlinear DCT discriminant feature extraction approach for face recognition. First, we provide the representation of DCT frequency bands and select appropriate bands. Second, we extract the nonlinear discriminant features from the selected bands by presenting a new kernel discriminant method, that is, the improved KDCV method which takes advantages of both

kernel discriminative common vectors and different vectors. The nearest neighbor classifier is adopted in classifying the extracted features. We employ a large public face database, the FERET database, as the test data. Experiments demonstrate the effectiveness of the proposed new approach.

2. Recognition Approach

Suppose that each gray image sample in the sample set X is sized $C \times D$ where $C \geq D$. We perform the two-dimensional DCT on each image and obtain its transformed image [1]. Figure 1 shows the face demonstration images of DCT transform: (a) an original image; (b) its DCT transformed image and frequency bands representation. According to Fig. 1 (b), most information or energy of face image is concentrated in the left-up corner, that is, in the low frequency bands. We use a half square ring to express the k^{th} frequency band [6], where $1 \leq k \leq C$. If we select the k^{th} band, then we keep the corresponding frequency band values of DCT transformed image, otherwise we set the band values to change to zero.

The nonlinear discriminant analysis technique generally uses the kernel-based method to realize a conceptual nonlinear transformation from an input feature space into a high-dimensional space. With respect to a given nonlinear mapping function Φ , the input data space R^n can be mapped into the kernel space F : $\Phi: R^n \rightarrow F, x \mapsto \Phi(x)$. Suppose that there are c known pattern classes in the sample set X , l_i is the training sample number of the i^{th} class, and there will be a total of $M = \sum_{i=1}^c l_i$ training samples. Let S_B^Φ and S_W^Φ represent the between-class scatter matrix and the within-class scatter matrix in F , respectively. They are defined as:

$$S_B^\Phi = \frac{1}{M} \sum_{i=1}^c l_i (\mu_i^\Phi - \mu^\Phi)(\mu_i^\Phi - \mu^\Phi)^T, \quad (1)$$

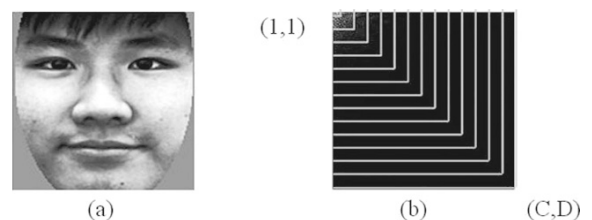


Fig. 1 Face demonstration images of DCT transform and filtering.

Manuscript received May 8, 2009.

Manuscript revised July 24, 2009.

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

^{††}The author is with the Dept. of Computing, Hong Kong Polytechnic University, Hong Kong.

^{†††}The author is with the Nanjing University of Science and Technology, Nanjing, China.

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

DOI: 10.1587/transinf.E92.D.2527

$$S_W^\Phi = \frac{1}{M} \sum_{i=1}^c \sum_{m=1}^{l_i} (\Phi(x_m^i) - \mu_i^\Phi)(\Phi(x_m^i) - \mu_i^\Phi)^T, \quad (2)$$

where $\Phi(x_m^i)$ is the m^{th} training sample of the i^{th} class, μ_i^Φ is the mean value of the i^{th} class samples, and μ^Φ is the mean value of all training samples. Define the Fisher discriminant criterion in F as:

$$J(\varphi) = \frac{\varphi^T S_B^\Phi \varphi}{\varphi^T S_W^\Phi \varphi}. \quad (3)$$

The optimal discriminant vector φ can be expressed by a linear combination of the observations in F , we have $\varphi = \sum_{j=1}^M a_j \Phi(x_j) = H\alpha$, where $H = [\Phi(x_1), \Phi(x_2), \dots, \Phi(x_M)]$ and $\alpha = (a_1, a_2, \dots, a_M)^T$. We thus obtain [7]:

$$J^\Phi(\alpha) = \frac{\alpha^T (KUK) \alpha}{\alpha^T (K(I_N - U)K) \alpha}, \quad (4)$$

where K is an $M \times M$ kernel symmetric matrix, $K = (K_{ij})_{i,j=1,2,\dots,M}$, $K_{ij} = \Phi(x_i)^T \Phi(x_j)$; I_N is an $M \times M$ identity matrix; and U is an $M \times M$ block diagonal matrix. $U = \text{diag}(U_1, \dots, U_c)$ where U_i ($i = 1, 2, \dots, c$) is a $l_i \times l_i$ matrix, the elements of which are equal to $1/l_i$.

Let

$$S'_B = KUK \quad \text{and} \quad S'_W = K(I_N - U)K, \quad (5)$$

the nonlinear discriminant capability of the band can be calculated by:

$$F_k = \frac{\text{tr}(S'_B)}{\text{tr}(S'_W)}, \quad (6)$$

where $\text{tr}()$ indicates the trace of matrix, S'_B and S'_W separately represent the corresponding between-class scatter matrix and the within-class scatter matrix in the kernel space [8]. In the experiments, we calculate the nonlinear discriminant capabilities of all frequency bands, and then we rank them in a descending order and select a part of bands with the largest discriminant capabilities.

Next, we present an improved KDCV method to extract nonlinear discriminant features from the selected DCT bands as follows:

Step1. Compute the common vectors and different vectors:

The principle of KDCV is to acquire the nonlinear projection transform in the null space of S'_W :

$$\begin{aligned} J(W) &= \arg \max_{|W^T S'_W W|=0} |W^T S'_B W| \\ &= \arg \max_{|W^T S'_W W|=0} |W^T S^\Phi_{com} W|, \end{aligned} \quad (7)$$

where S^Φ_{com} is the common vector scatter matrix, S'_W is an $M \times M$ symmetric matrix, and M is the total number of training samples. Let V be the non-null space of S'_W , and V^\perp be the null space of S'_W . For any sample $\Phi(x_m^i)$ of sample set X in the kernel space, we have [8]:

$$\Phi(x_m^i) = \Phi(x_{com}^i) + \Phi(x_{m_dif}^i), \quad (8)$$

where $\Phi(x_{com}^i) \in V^\perp$, $\Phi(x_{m_dif}^i) \in V$, $\Phi(x_{com}^i)$ and $\Phi(x_{m_dif}^i)$ separately represent the common vector and different vector parts of $\Phi(x_m^i)$. Φ indicates a given nonlinear mapping function. In this paper, Φ is expressed by the Gaussian kernel function $k(c_1, c_2) = \exp(-\|c_1 - c_2\|^2 / 2\delta^2)$. Note that for all the samples of the i^{th} class, their common vector parts are same [10].

Step2. Compute the optimal projection transforms W_{com} and W_{dif} : According to KDCV [8], we use the common vectors to construct the scatter matrix S^Φ_{com} , and then get the projection transform W_{com} composed of the eigenvectors of S^Φ_{com} with nonzero eigenvalues. We obtain the following kernel discriminative common vectors:

$$y_{com}^i = W_{com}^T \Phi(x_{com}^i), \quad i = 1, 2, \dots, c, \quad (9)$$

where y_{com}^i is identical for the i^{th} class and the feature dimension of y_{com}^i is $c - 1$.

In this paper, we use the different vectors to calculate W_{dif} . Take $\Phi(x_{m_dif}^i)$ of each training sample to construct the corresponding within-class scatter matrix $S^\Phi_{W_dif}$ and between-class scatter matrix $S^\Phi_{B_dif}$. W_{dif} is designed to satisfy the Fisher discriminant criterion:

$$J(W_{dif}) = \frac{|W_{dif}^T S^\Phi_{B_dif} W_{dif}|}{|W_{dif}^T S^\Phi_{W_dif} W_{dif}|}. \quad (10)$$

W_{dif} is composed of the eigenvectors corresponding to the nonzero eigenvalues of $(S^\Phi_{W_dif})^{-1} S^\Phi_{B_dif}$. We obtain the following kernel discriminative different vectors:

$$y_{m_dif}^i = W_{dif}^T \Phi(x_{m_dif}^i), \quad i = 1, 2, \dots, c, \quad (11)$$

where the feature dimension of $y_{m_dif}^i$ is $c - 1$.

Step3. Construct the synthesized optimal projection transform W for the improved KDCV: W is constructed by serially combining W_{com} and W_{dif} :

$$W = [W_{com}^T \quad W_{dif}^T]. \quad (12)$$

The whole recognition procedure of proposed approach is implemented as follows:

- (i) Perform DCT on each image sample of X . Select appropriate DCT frequency bands for the transformed images and express the selected bands in the form of feature vector. We thus obtain a one-dimensional sample set X' .
- (ii) Use the improved KDCV to compute the optimal projection transform W in the kernel space. We acquire the kernel discriminative common vector y_{com} and kernel discriminative different vector y_{dif} :

$$\begin{aligned} W \cdot \Phi(x) &= \begin{bmatrix} W_{com} & W_{dif} \end{bmatrix} \cdot \Phi(x) \\ &= \begin{bmatrix} y_{com} & y_{dif} \end{bmatrix} \end{aligned} \quad (13)$$

Normalize y_{com} and y_{dif} as

$$y = \begin{bmatrix} \frac{y_{com}}{\|y_{com}\|} & \frac{y_{dif}}{\|y_{dif}\|} \end{bmatrix}, \quad (14)$$

where $\|\cdot\|$ represents the 2-norm of vector. We obtain a new sample set Y corresponding to X' ;

- (iii) Take the nearest neighbor classifier with the cosine distance measure to classify Y . The distance $d(\cdot)$ between two arbitrary samples, y_1 and y_2 , is defined by

$$d(y_1, y_2) = -\frac{y_1^T y_2}{\|y_1\| \|y_2\|}. \quad (15)$$

3. Experimental Results

We use a public large face image database, the FERET database [11], as the test data. In the experiment it contains 600 frontal facial images corresponding to 200 individuals with each person contributing 3 images. The images in this database were captured under various illuminations and facial expressions. Each image is 384×256 with 256 gray levels. Since many images in this database include the background and the body chest region, we adopt the FERET protocol preprocessing software to automatically crop every image sample. That is, the facial region of each image is cropped off by using the centers of eyes and mouth, and the cropped images are resized to 128×128 . Figure 2 shows the demonstration face images of five subjects.

In the test, two images of each subject are randomly chosen for training, and the remaining one is used for testing. We perform the two-dimensional DCT on the images. Figure 3 shows the nonlinear discriminability values of all DCT bands by using Formula (6).

According to the nonlinear discriminability values shown in Fig. 3, we rank all frequency bands in a descending order and in turn choose a part of them with the largest nonlinear discriminability values. Figure 4 shows the recognition rates of the selected DCT frequency bands. While using the selected bands from No. 1 to No. 23, we get the highest recognition rate of 99%.

We compare the proposed approach with five representative discriminant methods which include the DCT-LDA method [6], the generalized discriminant analysis (GDA) method [7], the DCT-GDA method, the DCT-KDCV method, and the Gabor-GDA method [9]. DCT-GDA and DCT-KDCV take the same method of selecting frequency bands as the proposed approach. In the test, the numbers of selected frequency bands are 22, 25, and 21 for DCT-LDA, DCT-GDA, and DCT-KDCV, respectively. All compared methods use the same classifier, i.e. the nearest neighbor classifier. Table 1 shows the recognition rates of all compared methods. The proposed approach performs better than other methods. From Table 1, the proposed approach improves the recognition rates at least by 1.5% (=



Fig. 2 Demonstration face images of five subjects.

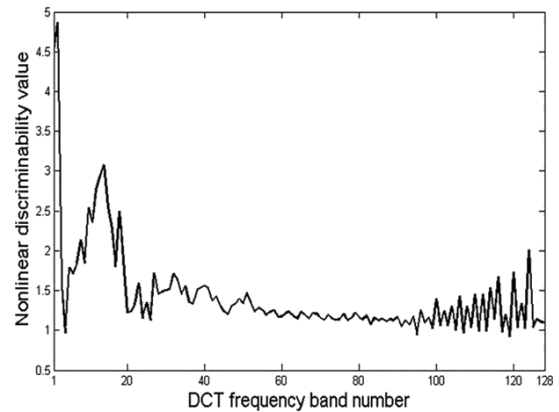


Fig. 3 Nonlinear discriminability values of all DCT bands.

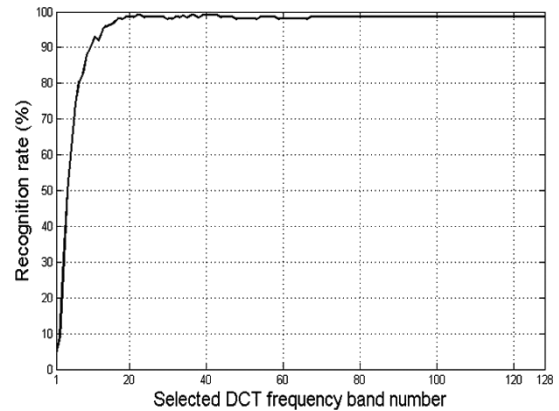


Fig. 4 Recognition rates of selected DCT frequency bands.

Table 1 Recognition rates of compared methods.

Methods	Recognition rates (%)
Proposed Approach	99%
DCT-LDA	93%
GDA	90%
DCT-GDA	96%
DCT-KDCV	97.5%
Gabor-GDA	97.5%

99% – 97.5%). Besides, compared with the commonly used nonlinear discriminant method GDA, the proposed approach saves 75.64% (= $(65.76 - 16.02) / 65.76 \times 100\%$) computing time (second), where the computing time indicates the time of achieving discriminant features.

4. Conclusions

The proposed nonlinear DCT discriminant approach is obviously superior to DCT-LDA, and it can extract more effective nonlinear discriminant features than the two representative methods of GDA and KDCV. And it outperforms the Gabor-GDA method. By selecting the DCT frequency bands, our approach also improves the computing speed of nonlinear discriminant methods.

Acknowledgements

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

References

- [1] Z.M. Hafed and M.D. Levine, "Face recognition using the discrete cosine transform," *Int. J. Comput. Vis.*, vol.43, no.3, pp.167–188, 2001.
- [2] D. Zhang, X.Y. Jing, and J. Yang, "Biometric image discrimination (BID) technologies," IGP/INFOSCI/IRM Press, 2006.
- [3] A.K. Jain, R.P.W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.22, no.1, pp.4–37, 2000.
- [4] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data with application to face recognition," *Pattern Recognit.*, vol.34, no.12, pp.2067–2070, 2001.
- [5] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.27, no.1, pp.4–13, 2005.
- [6] X.Y. Jing and D. Zhang, "A face and palmprint recognition approach based on discriminant DCT feature extraction," *IEEE Trans. Syst., Man Cybern. B*, vol.34, no.6, pp.2405–2415, 2004.
- [7] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Comput.*, vol.12, no.10, pp.2385–2404, 2000.
- [8] X.Y. Jing, D. Zhang, J.Y. Yang, Y.F. Yao, and M. Li, "Face and palmprint pixel level fusion and kernel DCV-RBF classifier for small sample biometrics recognition," *Pattern Recognit.*, vol.40, no.11, pp.3209–3224, 2007.
- [9] L.L. Shen, L. Bai, and M. Fairhurst, "Gabor wavelets and general discriminant analysis for face identification and verification," *Image Vis. Comput.*, vol.25, no.5, pp.553–563, 2007.
- [10] M.B. Gulmezoglu, V. Dzhaferov, and A. Barkana, "The common vector approach and its relation to principal component analysis," *IEEE Trans. Speech Audio Process.*, vol.9, no.6, pp.655–662, 2001.
- [11] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.22, no.10, pp.1090–1104, 2000.