Face recognition using improved-LDA with facial combined feature

Dake Zhou (周大可), Xin Yang (杨 新), and Ningsong Peng (彭宁嵩)

Institute of Image Processing & Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200030

Received August 9, 2004

Face recognition subjected to various conditions is a challenging task. This paper presents a combined feature improved Fisher classifier method for face recognition. Both of the facial holistic information and local information are used for face representation. In addition, the improved linear discriminant analysis (I-LDA) is employed for good generalization capability. Experiments show that the method is not only robust to moderate changes of illumination, pose and facial expression but also superior to the traditional methods, such as eigenfaces and Fisherfaces.

OCIS codes: 100.5010, 100.2000, 100.2960.

A significant problem encountered in face recognition (FR) community is that many FR methods perform poorly under various conditions, such as the changes of illumination, pose and facial expression. A solution to this problem may be to use facial holistic as well as local information for face representation, which is inspired by the fact that both holistic information and local information are necessary for human recognition of faces. In addition, a good (robust) FR algorithm should also consider classification problem[1]. Linear discriminant analysis (LDA), derived from the Fisher linear classifier to seek the projection which maximizes the ratio of the between- and within-class scatters, is widely used in face recognition community[2−4]. However, there are two main limitations in the LDA based FR methods: 1) Fisher criterion is nonoptimal with respect to classification rate[4], and 2) the degenerated generalization ability caused by the "small sample size" (SSS) problem[3,4]. In Ref. [4], a novel LDA extension technique called improved-LDA (I-LDA), which can effectively deal with the above two problems, is proposed for face recognition.

The main objective of this research is to improve the accuracy of face recognition subjected to various conditions. In this paper, a combined feature improved Fisher classifier (CFIFC) method is proposed for face recognition. Experimental results on ORL and Yale face databases show that the proposed method is more robust than the traditional FR methods, such as eigenface[6] and Fisherfaces[2].

In the CFIFC framework, a face image is first divided into smaller sub-images and then the discrete cosine transform (DCT) technique is applied on the whole face image and some sub-images to extract facial holistic and local information. After concatenating these DCT based facial holistic and local feature vectors to a combined feature vector, the I-LDA is employed to obtain a low-dimensional feature vector with enhanced discrimination power. Finally, the nearest neighbor (to the mean) rule with Euclidean distance measure is used for classification.

We use DCT to extract facial holistic information, i.e., a low-to-mid frequency square subset of the two-dimensional (2D) DCT coefficients of a face image is extracted as the facial holistic feature, which is similar to that used in Ref. [5]. The size of this subset is chosen such that it can sufficiently represent a face, but it can in fact be quite small, as will be shown in our experiments. The similar technique is used to extract facial local information. We first divide the whole face image roughly into four small overlapping regions: eyes, nose, mouth and the remainder (First, detect the centers of eyes, and then the nose tip. Finally, the segmentation is completed by using these location information and the priori knowledge of face structure). Obviously, the regions of eyes, nose and mouth are the most salient regions for face recognition[1]. However, since the mouth shape is very sensitive to changes of facial expression, the mouth region is discarded and only the eyes and nose regions are used in this paper. DCT is then used on the two sub-images to extract local information, respectively. Figure 1 shows a face image and its local regions of eyes and nose.

Let $\mathbf{X}_{\rm h}$, $\mathbf{X}_{\rm le}$, $\mathbf{X}_{\rm ln}$ denote the facial holistic feature vector, the eyes and nose regions feature vectors, respectively. $\mathbf{X}_{\rm h}$, $\mathbf{X}_{\rm le}$, $\mathbf{X}_{\rm ln}$ can be defined as follows:

$$\mathbf{X}_{\rm h} = {\rm Re\ shape}[\Phi(f), n_{\rm h}], \quad (1)$$

$$\mathbf{X}_{\rm le} = {\rm Re\ shape}[\Phi(f_{\rm le}), n_{\rm e}], \quad (2)$$

$$\mathbf{X}_{\rm ln} = {\rm Re\ shape}[\Phi(f_{\rm n}), n_{\rm n}], \quad (3)$$

where $\Phi(\cdot)$ denotes the 2D DCT function, f, $f_{\rm le}$, and $f_{\rm n}$ denote the face image, eyes and nose sub-images, respectively, Re shape(A, n) is a function that extracts the

Fig. 1. A face image (a) and its local regions of the eyes and nose (b).

1671-7694/2005/060330-03 http://www.col.org.cn

top-left $n \times n$ square matrix from the matrix A and then transforms the square matrix into a n^2-dimensional column vector.

A new feature vector $\widetilde{\mathbf{Y}}$ is then defined as the concatenation of $\mathbf{X}_\mathrm{h}$, $\mathbf{X}_\mathrm{le}$, $\mathbf{X}_\mathrm{n}$: $\widetilde{\mathbf{Y}} = (X_\mathrm{h}^t, X_\mathrm{le}^t, X_\mathrm{ln}^t)^t$, and t denotes the transpose operator. The corresponding facial combined feature vector can be derived from $\widetilde{\mathbf{Y}}$

$$\mathbf{Y} = \frac{\widetilde{\mathbf{Y}} - \mathbf{u}}{\sigma}, \tag{4}$$

where $\mathbf{u}$ is the mean vector, σ consists of σ_j $(j = 1, \cdots, k)$, σ_j is the j-th component of the standard deviation, and k is the dimensionality of vector $\mathbf{Y}$.

I-LDA (The first step of I-LDA is to seek the subspace containing the most significant discrimination information for dimension reduction. But in this paper, since the dimensionality of combined feature is small, this step is discarded.) can effectively deal with the two problems encountered in traditional LDA-based FR methods[4]: 1) Fisher criterion is not directly related to classification rate, and 2) the degenerated generalization ability caused by the SSS problem. To obtain a modified criterion that it is more closely related to classification error, I-LDA introduces weighted schemes to reconstruct the between-class scatter matrix

$$S_\mathrm{b} = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} P_i P_j w(d_{ij})(\mathbf{M}_i - \mathbf{M}_j)(\mathbf{M}_i - \mathbf{M}_j)^t, \tag{5}$$

where $\mathbf{M}_i$ and P_i are the mean vector and priori probability of the i-th class, and d_{ij} is the Mahanalobis distance between the i-th class and j-th class. The weighted function $w(d_{ij})$ is a monotonically decreasing function of the distance d_{ij}, with the constraint that it should drop faster than the square of d_{ij}

$$w(d_{ij}) = d_{ij}^{-4} \Big/ \sum d_{ij}^{-4}. \tag{6}$$

To improve its generalization capability, I-LDA decomposes the traditional LDA procedure into a simultaneous diagonalization of the weighted between- and within-class scatter matrices. The simultaneous diagonalization is stepwisely equivalent to two operations: whitening the within-class scatter matrix and applying principal component analysis (PCA) on the weighted between-class scatter matrix using the transformed data. I-LDA first whitens the within-class scatter matrix S_w

$$S_\mathrm{w}\tilde{\Xi} = \tilde{\Xi}\tilde{\Gamma}, \quad \tilde{\Gamma}^{-\frac{1}{2}}\tilde{\Xi}^t S_w' \tilde{\Xi}\tilde{\Gamma}^{-\frac{1}{2}} = I, \tag{7}$$

where $\tilde{\Xi} = (e_1, e_2, \cdots, e_k) \in R^{k \times k}$ is the eigenvector matrix of S_w, I is the unitary matrix and $\tilde{\Gamma} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_k) \in R^{k \times k}$ is the diagonal eigenvalue matrix of S_w with diagonal elements in decreasing order. During the whitening step, I-LDA keeps a good tradeoff between the need for adequate signal representation and generalization performance by selecting proper principal components. Suppose eigenvalue set $\{\lambda_i\}_{i=1}^m$ $(m < k)$ is selected, thus m can be determined as the largest integer that satisfies

$$\Big(\sum_{j=1}^{m} \lambda_j \Big/ \sum_{i=1}^{k} \lambda_i\Big) \geq \alpha \quad \text{and} \quad \lambda_m \geq \varepsilon, \tag{8}$$

where α and ε are the preset thresholds, respectively.

Then we have matrices $\Xi = (e_1, \cdots, e_m) \in R^{k \times m}$ and $\Gamma = \mathrm{diag}(\lambda_1, \cdots, \lambda_m) \in R^{m \times m}$. The new between-class scatter matrix is defined by

$$\tilde{S}_\mathrm{b} = \Gamma^{-\frac{1}{2}}\Xi^t S_\mathrm{b} \Xi \Gamma^{-\frac{1}{2}}. \tag{9}$$

Now diagonalize the new between-class scatter matrix to obtain its eigenvector matrix Ψ. Finally, the overall transformation matrix of the I-LDA is defined as

$$T = \Xi\Gamma^{-\frac{1}{2}}\Psi. \tag{10}$$

The publicly available ORL and Yale face databases are used to test the proposed method. The ORL database contains 400 face images of 40 distinct subjects and there are variations in pose, facial expression and details. The Yale database consists of 165 face images of 15 subjects and there are variations in illumination, facial expression and details. Note that for the images from the Yale database, before they are used in our experiment, they are normalized to the size of 100×100.

The first series of experiments on the ORL database investigate that how many DCT coefficients should be used for the facial holistic representation. We employ different number of coefficients to identify face with the "leave-one" strategy, and the results are shown in Table 1. To keep a good tradeoff between the recognition accuracy and subsequent computational complexity, we use 49 coefficients to represent facial holistic information. After performing similar experiments, we employ 16 coefficients to represent the each local information.

We next studied the comparative performance of four methods on the ORL database, including the proposed method, the well-known Fisherfaces[2] and eigenfaces[6] methods, and the enhanced Fisher model (EFM) method[3]. For a fair comparison, the same

Table 1. Classification Rate Versus Number of DCT Coefficients (ORL)

Number of Coefficients	16	25	36	49	64	81	100
Recognition Accuracy (%)	97.1	98.2	98.8	99.0	99.2	98.4	97.9

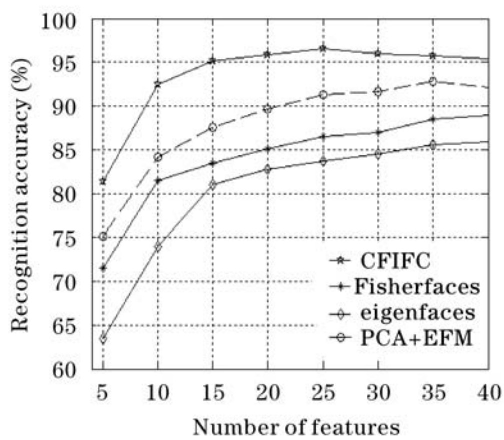

Fig. 2. Classification rate versus feature dimensionality (ORL).

classification rule and similarity measure, i.e., the nearest neighbor (to the mean) rule and Euclidean distance, are used for the four methods. Figure 2 shows the classification rate curves of the four methods with respect to the dimensionality of features where 5 face images per person are selected randomly for training. It is clear that the proposed method performs better than the other three methods. Note that our method achieves 96.4% recognition accuracy while only 25 features are used. The classification rate curves of the four methods are also shown in Fig. 3 as functions of the number of training samples per person (the feature dimensionalities of the four methods are 39, 39, 40 and 39, respectively). The figure shows that our method is with the best performance among the four methods. Because of the SSS problem, the eigenfaces performs better than the other two methods when there are only two training samples per person. One can also learn from the figure that the proposed method has better generalization capability compared to the other two LDA based methods.

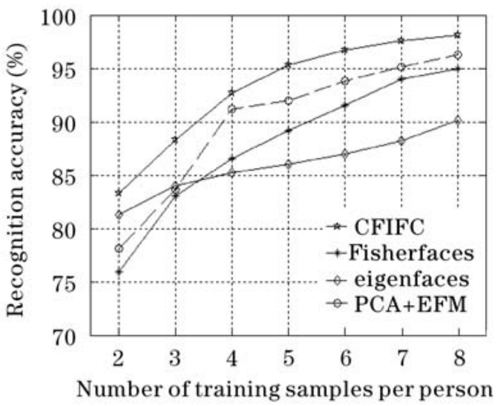

Fig. 3. Classification rate versus number of training samples per person (ORL).

Table 2. Comparative Recognition Performance of the Four Methods (Yale)

Results/Method	CFIFC	PCA+EFM	Fisherfaces	Eigenfaces
Accuracy	96.7%	93.9%	92.1%	87.8%
Features	14	14	14	15

To evaluate the recognition performance under varying lighting conditions, we performed the last series of experiments on the Yale database with 5 training samples per person. The proposed method also performs well compared to the other three methods (see Table 2).

In conclusion, this paper introduces a combined feature improved Fisher classifier (CFIFC) for face recognition. The key to the method lies in the fact that it uses both facial global and local information for robust face representation while at the same time employs the I-LDA for good g eneralization. Experiments show that the method is not only robust to moderate changes of illumination, pose and facial expression but also superior to the traditional FR methods. Another advantage of our method is its lower computational complexity during the training stage. Our method uses DCT for facial combined feature extraction, which means that the method can be computed more efficiently than the entirely statistics based methods, especially while running in a large face database.

This work was partially supported by National Natural Science Foundation of China (No. 30170264), and the Foundation of the Science and Technology Committee of Shanghai, China (No. 025115010). D. Zhou's e-mail address is normanzhou@sjtu.edu.cn.

References

1. W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, ACM Computing Surveys **35,** 399 (2003).
2. P. Belhumeur, J. Hespanha, and D. Kriegman, IEEE Trans. PAMI **19,** 711 (1997).
3. C. Liu and H. Wechsler, IEEE Trans. Image Processing **11,** 467 (2002).
4. D. Zhou and X. Yang, in *Lecture Notes in Computer Science 3212* A. Campilho and M. Kamel (eds.) (Springer-Verlag, Berlin Heidelberg, 2004) P692.
5. Z. M. Hafed and M. D. Levine, International J. Computer Vision **43,** 167 (2001).
6. M. A. Turk and A. P. Pentland, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 586 (1991).