An Improved LDA Approach

Xiao-Yuan Jing, David Zhang, and Yuan-Yan Tang, *Fellow, IEEE*

***Abstract*—Linear discrimination analysis (LDA) technique is an important and well-developed area of image recognition and to date many linear discrimination methods have been put forward. Despite these efforts, there persist in LDA at least three areas of weakness. The first weakness is that not all the discrimination vectors that are obtained are useful in pattern classification. Second, it remains computationally expensive to make the discrimination vectors completely satisfy statistical uncorrelation. The third weakness is that it is necessary to select the appropriate principal components. In this paper, we propose to improve discrimination technique in these three areas and to that end present an improved LDA (ILDA) approach which synthesizes these improvements. Experimental results on different image databases demonstrate that our improvements on LDA are efficient, and that ILDA outperforms other state-of-the-art linear discrimination methods.**

***Index Terms*—Discrimination vectors selection, Fisherface method, image recognition, improved linear discrimination analysis (ILDA) approach, statistical uncorrelation, principal components selection.**

I. Introduction

IN THIS section, we first analyze the advantages and disadvantages of the major linear discrimination methods and then suggest some ways they might be improved.

A. Major Linear Discrimination Methods

In the field of pattern recognition, and especially in image recognition, image data are always high-dimensional and require considerable computing time for classification. Linear discrimination analysis (LDA) technique is thus important in extracting effective discriminative features and reducing dimensionality, and costing little computing time. It has been shown in many applications of image recognition, that LDA technique can satisfy these requirements [1]–[4]. So far many linear discrimination methods have been proposed for use in image recognition. Two of the most well-known are the Eigenface and Fisherface methods.

Based on principal component analysis (PCA) [5], the Eigenface method [6] uses the total covariance (or scatter) matrix S_t, as the production matrix to perform the Karhunen–Loeve (KL) transform. It cannot, however, make full use of pattern separability information like the Fisher criterion, and its recognition effect is not ideal when the size of the sample set is large [7].

The Fisherface method [8] combines PCA and the Fisher criterion [9] to extract the information that discriminates between the classes of a sample set. It is a most representative method of LDA. Nevertheless, Martinez *et al.* demonstrated that when the training data set is small, the Eigenface method outperforms the Fisherface method [7]. Should the latter be outperformed by the former? This provoked a variety of explanations. Liu *et al.* thought that it might have been because the Fisherface method uses all the principal components, but the components with the small eigenvalues correspond to high-frequency components and usually encode noise [10], leading to recognition results that are less than ideal. In line with this theory, they presented two enhanced Fisher linear discrimination (FLD) models (EFMs) [10] and an enhanced Fisher classifier [11] for face recognition. Their experiential explanation lacks sufficient theoretical demonstration, however, and EFM does not provide an automatic strategy for selecting the components.

Chen *et al.* proved that the null space of the within-class scatter matrix S_w contains the most discriminative information when a small sample size problem takes place [12]. Their method is also inadequate, however, as it does not use any of the information outside the null space. In [13], Yu *et al.* propose a direct LDA (DLDA) approach to solve this problem. It simultaneously diagonalizes both the between-class scatter matrix S_b (or S_t) and S_w. Let $W^T S_w W = D_w$, and let $W^T S_b W = I$ or $W^T S_t W = I$. According to the theory, DLDA should discard some of the eigenvectors of D_w that correspond to the higher eigenvalues, and keep the remainders, especially those eigenvectors that correspond to the zero eigenvalues. This approach, however, has a number of limitations. First, it does not demonstrate how to select its eigenvectors. Second, the related demonstration is rather difficult. Third, in the application of DLDA, there is a contradiction between the theory and the experiment. The theory requires that the eigenvectors of D_w corresponding to the higher eigenvalues be discarded, but the experiment obtains the improved recognition results by employing all of the eigenvectors of D_w.

Optimal discrimination vectors (ODV) is a special kind of LDA method, which has been applied to a wide range of applications in pattern classification [16]–[18]. It requires that every discrimination vector satisfy the Fisher criterion and the obtained Fisher discrimination vectors are necessary to satisfy the orthogonality constraint [14], but as a result its solution is more complicated than other LDA methods. Jin *et al.* proposed an uncorrelated optimal discrimination vectors (UODV) method that used the constraint of statistical uncorrelation [19]. UODV produces better results than ODV on the same handwritten data,

Manuscript received July 10, 2003; revised January 19, 2004. The work was partially supported by the UGC/CRC fund from the HKSAR Government, the central fund from the Hong Kong Polytechnic University, and the NSFC fund under Contract 60332010. This paper was recommended by Associate Editor Q. Zhu.

X.-Y Jing is with the Bio-Computing Research Center and Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, Guangdong Province, China.

D. Zhang is with the Biometrics Research Center and Department of Computing, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (e-mail: csdzhang@comp.polyu.edu.hk).

Y.-Y. Tang is with the Pattern Recognition Laboratory and Department of Computer Science, Hong Kong Baptist University, Kowloon Tong, Hong Kong.

Digital Object Identifier 10.1109/TSMCB.2004.831770

where the only difference lies in their respective constraints [20]. Jing *et al.* subsequently presented a more rational UODV method and a generalized theorem for UODV [21], [22].

Many others methods have been proposed. Zhang *et al.* presented a face recognition system based on hybrid neural and dual eigenfaces methods [23]. Jing *et al.* put forward a classifier combination method for face recognition [24]. In [25], [26], several new discrimination principles based on the Fisher criterion were proposed. Yang used Kernel PCA for facial feature extraction and recognition [27], while Bartlett *et al.* applied independent component analysis (ICA) in face recognition [28]. However, M. H. Yang showed that both ICA and Kernel PCA need much more computing time than PCA. In addition, when the Euclidean distance is used, there is no significant difference in the classification performance of PCA and ICA [28]. Yang *et al.* presented an IMGPCA method for face recognition [29], which is a variant form of PCA. In this paper, we do not analyze and compare these extended discrimination methods [23]–[29], because they do not use the original Fisher criterion or the basic form of the PCA transform. And we confine ourselves to a comparison of major linear discrimination methods including the Eigenface method, the Fisherface method, DLDA, and UODV.

B. Necessary Improvements in LDA

The linear discrimination technique should be improved in three ways.

1) Discrimination vectors should be selected. Not all discrimination vectors are useful in pattern classification. Thus, vectors with the larger Fisher discrimination values should be chosen, since they possess more between-class than within-class scatter information.

2) Discrimination vectors should be made to satisfy the statistical uncorrelation, a favorable classification property. Although UODV satisfies this requirement, it also uses more computing time than the Fisherface method, since it respectively calculates every discrimination vector satisfying the constraint of uncorrelation. Our improvement should provide a measure that satisfies the requirement while saving a maximum of computing time. Therefore, this improvement will take advantages of both the Fisherface method and UODV. In other words, it is theoretically superior to UODV presented in [21], [22].

3) An automatic strategy for selecting principal components should be established. This would effectively improve classification performance and further reduce feature dimension. In [30], Jing *et al.* presented an elementary method for selecting the components. In this paper, we will perform a deep theoretical analysis and then provide a more logical selecting strategy.

We will now propose an improved LDA (ILDA) approach that synthesizes the foregoing suggestions. The rest of this paper is organized as follows. In Section II, we present three improvements to LDA and the related theoretical analysis. In Section III, we describe the ILDA approach, and in Section IV, we provide the experimental results on face and palmprint databases. Finally, we offer our conclusions in Section V.

II. Description of LDA Improvements

In this section, we first briefly describe two representative forms of the Fisherface method. Then, we present three improvements in LDA: improvements in the selection of discrimination vectors, in their statistical uncorrelation, and in the selection of principal components.

A. Two Representative Forms of the Fisherface Method

Generally, the image data is a two-dimensional (2-D) matrix $(A \times B)$, which can be transformed into a vector with H dimension, where $H = A \times B$. Thus, we can obtain a H-dimensional sample set X from the image database. Assuming there are c known pattern classes and N training samples in X. The original form of the Fisherface method is to maximize the following function [8]:

$$\begin{aligned} F(W_{\text{opt}}) &= \frac{\left|W_{\text{fld}}^T W_{\text{pca}}^T S_b W_{\text{pca}} W_{\text{fld}}\right|}{\left|W_{\text{fld}}^T W_{\text{pca}}^T S_w W_{\text{pca}} W_{\text{fld}}\right|} \\ W_{\text{opt}} &= W_{\text{pca}} W_{\text{fld}}. \end{aligned} \tag{1}$$

In order to avoid the complication of a singular S_w, the Fisherface method discards the smallest c principal components. This is because the rank of S_w is at most $N - c$ [8]. Nevertheless, when the rank of S_w is less than $N - c$, this method is incapable of completely ensuring that S_w is nonsingular in theory [16]. In other words, it cannot completely overcome the small sample size problem [12]. Here, an equivalent form of the Fisherface method is used

$$\begin{aligned} F(W_{\text{opt}}) &= \frac{\left|W_{\text{fld}}^T W_{\text{pca}}^T S_b W_{\text{pca}} W_{\text{fld}}\right|}{\left|W_{\text{fld}}^T W_{\text{pca}}^T S_t W_{\text{pca}} W_{\text{fld}}\right|} \\ W_{\text{opt}} &= W_{\text{pca}} W_{\text{fld}}. \end{aligned} \tag{2}$$

In [15] and [17], the equivalence of (1) and (2) has been proven. When S_w is nonsingular, the same linear discrimination transform can be obtained from these two equations. However, when S_w is singular (the small sample size problem arises), (2) can perform the linear discrimination transform, whereas (1) cannot do so. Consequently, (2) is also a complete solution of the small sample size problem. Note that the following proposed improvements and ILDA approach are based on (2), and that when we compare the classification performance of different methods, we still use (1) to represent the original Fisherface method.

B. Improving the Selection of Discrimination Vectors

In (2), to simplify the expression, we use S_b to represent $W_{\text{pca}}^T S_b W_{\text{pca}}$ and S_t to represent $W_{\text{pca}}^T S_t W_{\text{pca}}$. Suppose that $W_{\text{opt}} = [\phi_1, \phi_2, \ldots, \phi_r]$, where r is the number of discrimination vectors. For $\forall \phi_i (i = 1, \ldots, r)$, we have

$$\phi_i^T S_t \phi_i = \phi_i^T S_b \phi_i + \phi_i^T S_w \phi_i. \tag{3}$$

If $\phi_i^T S_b \phi_i > \phi_i^T S_w \phi_i$, then

$$F(\phi_i) = \frac{\phi_i^T S_b \phi_i}{\phi_i^T S_t \phi_i} > 0.5. \tag{4}$$

In this situation, according to the Fisher criterion, there is more between-class separable information than within-class scatter information. So, we choose those discrimination vectors whose Fisher discrimination values are more than 0.5, and discard the others. This improvement allows efficient linear discrimination information to be kept and nonuseful information to be discarded. Such a selection of the effective discrimination vectors is important to the recognition effect, especially where the number of vectors is larger, which often happens when the number of pattern classes is large. The experiment will demonstrate the importance of this.

C. Improving the Statistical Uncorrelation of Discrimination Vectors

In Section I, we observed that the statistical uncorrelation of discrimination vectors is a favorable property, useful in pattern classification [19]–[22]. The unique difference between the Fisherface method and Jing's UODV method [22] is that the discrimination vectors obtained from UODV satisfy the constraint of statistical uncorrelation. It is a simple matter to prove that the Eigenface method [6] satisfies the statistical uncorrelation. This characteristic of the Eigenface method provides an explanation for its relatively insensitivity to different training data sets, compared with the Fisherface method [7]. Now, we introduce a corollary provided in [22].

Lemma 1 [22]: Suppose that the discrimination vectors obtained from UODV (refer to Jing's method) are $(\varphi_1, \varphi_2, \ldots, \varphi_r)$, where r is the rank of $S_t^{-1}S_b$, and the nonzero eigenvalues of $S_t^{-1}S_b$ are represented in descending order as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0$, and the kth eigenvector ϕ_k of $S_t^{-1}S_b$ correspond to $\lambda_k (1 \leq k \leq r)$. If $(\lambda_1, \lambda_2, \ldots, \lambda_r)$ are mutually unequal, that is

$$\lambda_1 > \lambda_2 > \cdots > \lambda_r > 0 \tag{5}$$

then φ_k can be represented by ϕ_k.

Lemma 1 shows that when the nonzero Fisher discrimination values are mutually unequal, the discrimination vectors generated from the Fisherface method can satisfy the statistical uncorrelation. That is, in this situation, the Fisherface method and UODV obtain identical discrimination vectors with nonzero discrimination values. Therefore, Lemma 1 reveals the essential relationship between these two methods.

Although UODV satisfies the statistical uncorrelation completely, it requires more computational time than the Fisherface method. Furthermore, it is not necessary to use UODV if the nonzero Fisher discrimination values are mutually unequal, because the Fisherface method can take the place of UODV. In the application of the Fisherface method, we find that only a small number of the Fisher values are equal respectively, and the others are unequal mutually. How, then, can computational time be reduced, while simultaneously guaranteeing the statistical uncorrelation for the discrimination approach? Here, we propose an improvement on the Fisherface method. Using the assumption in Lemma 1, our measure is as follows.

Step 1) Use the Fisherface method to obtain the discrimination vectors $(\phi_1, \phi_2, \ldots, \phi_r)$. If the corresponding Fisher values $(\lambda_1, \lambda_2, \ldots, \lambda_r)$ are unequal mutually, over; else go to the next step.

Step 2) For $2 \leq k \leq r$, if $\lambda_k \neq \lambda_{k-1}$, then keep ϕ_k, else replace ϕ_k by φ_k from UODV.

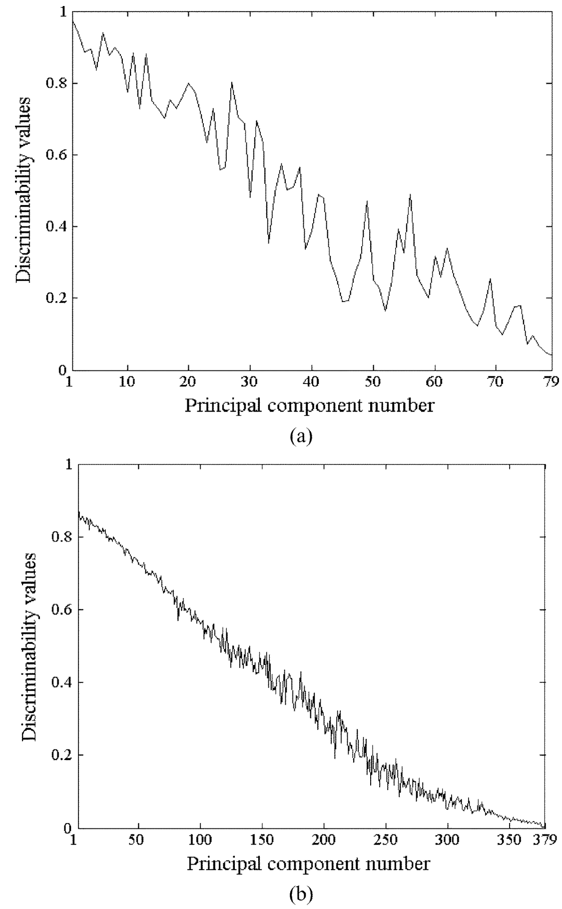


Fig. 1. Fisher discriminative values of the principal components obtained from: (a) ORL face database and (b) palmprint database.

Obviously, the proposal not only satisfies the statistical uncorrelation, it reduces the computing time. This will be further demonstrated by our experiments.

D. Improving the Selection of Principal Components

Assume that W_{pca} in (1) and (2) is represented by p eigenvectors (principal components) of S_t with nonzero eigenvalues, i.e., $W_{\text{pca}} = (\beta_1, \beta_2, \ldots, \beta_p)$. The Fisher discriminability of a principal component $\beta_i (1 \leq i \leq p)$ is evaluated as follows:

$$J_i = \frac{\beta_i^T S_b \beta_i}{\beta_i^T S_t \beta_i}, \quad (1 \leq i \leq p). \tag{6}$$

Obviously, this quantitative evaluation is rational because it is in accordance with the Fisher criterion. Fig. 1 shows the Fisher discriminative values of the principal components obtained from: 1) the ORL face database and 2) the palmprint database, where $p = 79$ and $p = 379$, respectively. From Fig. 1, two experimental rules can be obtained.

Rule 1: There is no completely direct proportional relationship between the discriminability value of a component and its eigenvalue;

TABLE I
A COMPARISON OF RECOGNITION RATES OF THE FISHERFACE METHOD A CHANGED FISHERFACE METHOD USING ALL THE COMPONENTS

Different data		ORL face database			Palmprint database		
Number of training samples per class		2	3	4	2	3	4
Recognition rates (%)	Fisherface method [8]	80.94	86.43	88.33	81.35	89.11	90.44
	A changed Fisherface method	47.5	49.28	53.33	48.83	55.75	56.05

Rule 2: Components with smaller eigenvalues generally have weaker discriminability values.

Rule 1 indicates that the selection method in EFM [10], which uses the components with the larger eigenvalues, is not completely reasonable, while *Rule 2* provides a quantitative explanation for why we can select the components with the larger eigenvalues for EFM. This is significant in Fig. 1(b), where the number of components (the training sample set) is large. We will give an automatic and more reasonable strategy for selecting the components than using EFM. The following theorem demonstrates that the total discriminability of LDA equals the sum of the discriminability of each component:

Theorem 1: Let tr represent the trace of the matrix. We have

$$\mathrm{tr}\left(\left(W_{\mathrm{pca}}^T S_t W_{\mathrm{pca}}\right)^{-1}\left(W_{\mathrm{pca}}^T S_b W_{\mathrm{pca}}\right)\right) = \sum_{i=1}^{p} J_i. \quad (7)$$

The proof is in the Appendix. Theorem 1 implies that in order to obtain the maximal total Fisher discriminability, we should use all the components. Nevertheless, some experiments in previous works [10], [11] have shown that the ideal recognition results may be obtainable by discarding those components with the smaller values.

Here, we also provide some experimental results. We use the Fisherface method but do a little change on it, that is not discarding the smallest c principal components and using all the components. Table I indicates a comparison of recognition rates of the Fisherface method and a changed Fisherface method using all the components on the ORL face database and the palmprint database, where the first two, three, and four samples per class are respectively taken as the training ones. We observe that the results of the changed Fisherface method are quite badly. However, the total Fisher discriminability obtained from this changed method is maximal according to Theorem 1. Thus, we have to face a contradiction between satisfying the maximal total discriminability and choosing as the discrimination vectors those with favorable characteristics. To solve this contradiction, it may be possible to make a tradeoff, i.e., the fundamental Fisher discriminability should be kept and some of components with the smaller Fisher values should be discarded. The following is our strategy.

Step 1) In accordance with *Rule 2*, discard the smallest c components, as in the Fisherface method. This helps to reduce the computing time.

Step 2) Compute the Fisher discrimination values J_i of the remainder components according to (6), then rank them in descending order and calculate the sum of their Fisher discriminability values J_{all}.

Step 3) Select the components with the first largest J_i values until a threshold T is satisfied, where T is the ratio of the sum of their values to J_{all}.

Fig. 2(a) shows a flowchart of this strategy. In accordance with our tradeoff strategy, we think that the value of T should not be too small or too large. We theoretically estimate that the value range of T might be around 0.5. The following experimental results on face and palmprint databases will show that the value range [0.4, 0.8] is appropriate for T, where the variance of the recognition rates is rather small. And in our experiments, T will be set as 0.6.

III. ILDA APPROACH

The ILDA approach, which synthesizes our three suggested improvements on LDA, can be described in the following four steps.

Step 1) Select the appropriate principal components according to the strategy defined in Section II-D and perform the Fisherface method using its equivalent form expressed by (2).

Step 2) From the discrimination vectors that are obtained, select those whose Fisher discrimination values are more than 0.5.

Step 3) Use the measure defined in Section II-C to make the selected vectors satisfy the statistical uncorrelation. Thus, the generated vectors construct the final linear discrimination transform W.

Step 4) For each sample x in X, extract the linear discrimination feature z

$$y = x * W. \quad (8)$$

This obtains a new sample set Y with the linear transformed features corresponding to X. Use the nearest neighbor classifier to classify Y. Here, the distance between two arbitrary samples, y_1 and y_2, is defined by

$$d(y_1, y_2) = \|y_1 - y_2\|_2 \quad (9)$$

where $\|\|_2$ denotes the Euclidean distance.

Fig. 2(b) shows a flowchart of ILDA.

IV. EXPERIMENTAL RESULTS

In this section, we first conduct the experiments on the three improvements to LDA. We then compare the experimental results of ILDA and other four linear discrimination methods: Eigenface, Fisherface, DLDA, and UODV, using different image data including a face database and a palmprint database.

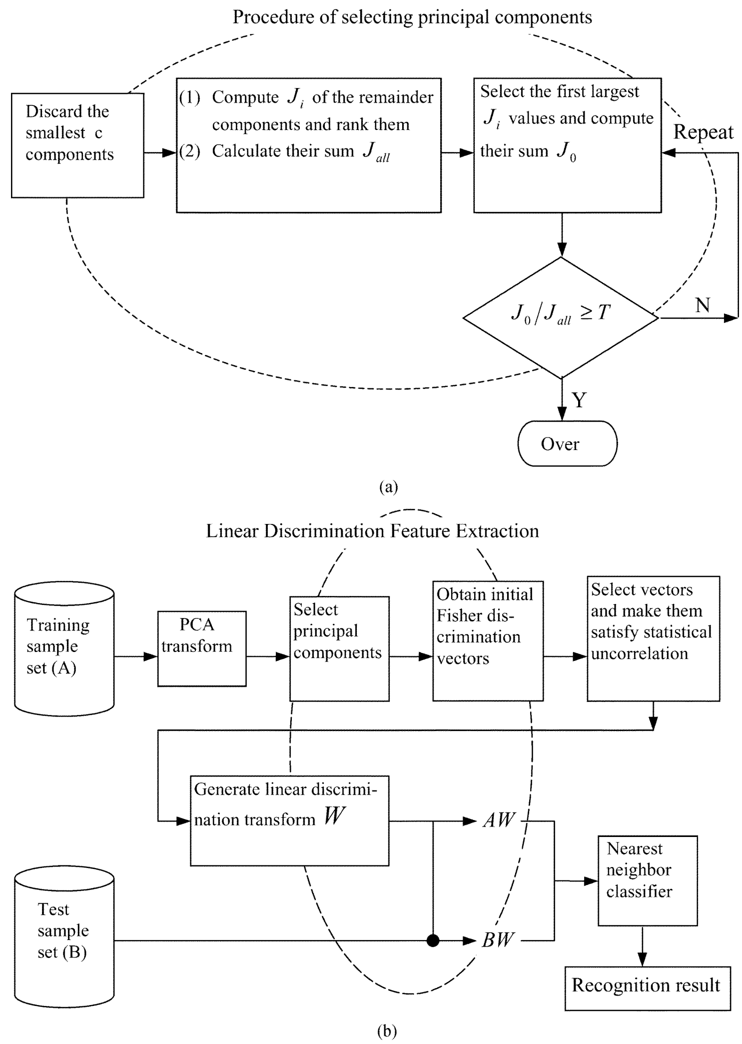


Fig. 2. Flowcharts of: (a) selecting the principal components and (b) ILDA.

The experiments are implemented on a Pentium 1.4-G computer and programmed using MATLAB language. Besides, we do not compare the test time for every method, because it is quite little (less than one second) when we test an image sample using any method in the experiments.

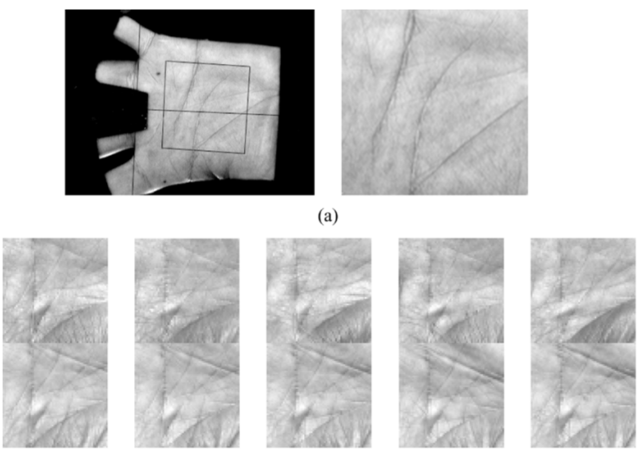
A. Introduction of Databases

The ORL facial image database [1] contains images with variations in facial expressions (open or closed eyes, smiling or not smiling), facial details (glasses or no glasses), facial poses (face tilted and rotated up to about 20 degrees), and in the scale of up to approximately 10%. The database contains 400 facial images made up of ten images of 40 individuals. Each image is 92 × 112 with 256 gray levels per pixel. Fig. 3 displays an example of ten images from one person.

For reasons such as its accommodation of low-resolution imaging, its ability to operate on low-cost capture devices, and the ease with which the palm can be segmented, palmprint recognition has become an important complement to personal identification. In [31], a Gabor-based method is applied to the online palmprint identification. In this paper, we use the LDA technique to perform offline palmprint recognition. Other two palmprint recognition methods that are Eigenpalm and Fisherpalm are presented in [32] and [33], respectively. These two methods are very similar to the Eigenface [6] and the Fisherface [8] methods, so we do not specially compare the Eigenpalm and the Fisherpalm methods in the following experiments of palmprint images. We collected palmprint images from 190 individuals using our self-designed capture device. The subjects mainly consisted of student and staff volunteers from the Hong Kong Polytechnic University. Of the subjects in this database, 130 persons are male, approximately 87% of the subjects are younger than 30 years old, about 10% are aged between 30 and 50, and about 3% are older than 50. The palmprint images were collected on two separate occasions, at an interval of around two months. After finishing the first collection, we slightly changed the light source and adjusted the focus of the CCD camera so that the images collected on the first and second occasions might be regarded as being captured by two different palmprint devices. On each occasion, the subjects were asked to each provide eight palmprint images for the right hand. Thus, each person provides 16 images and our database contains a total of 3 040 images from 190 different palms. The size of all the original palmprint images is 384 × 284 pixels with 75

[1]Available [Online] at: http://www.cam-orl.co.uk)

Fig. 3. Ten image samples from one person in the ORL database.



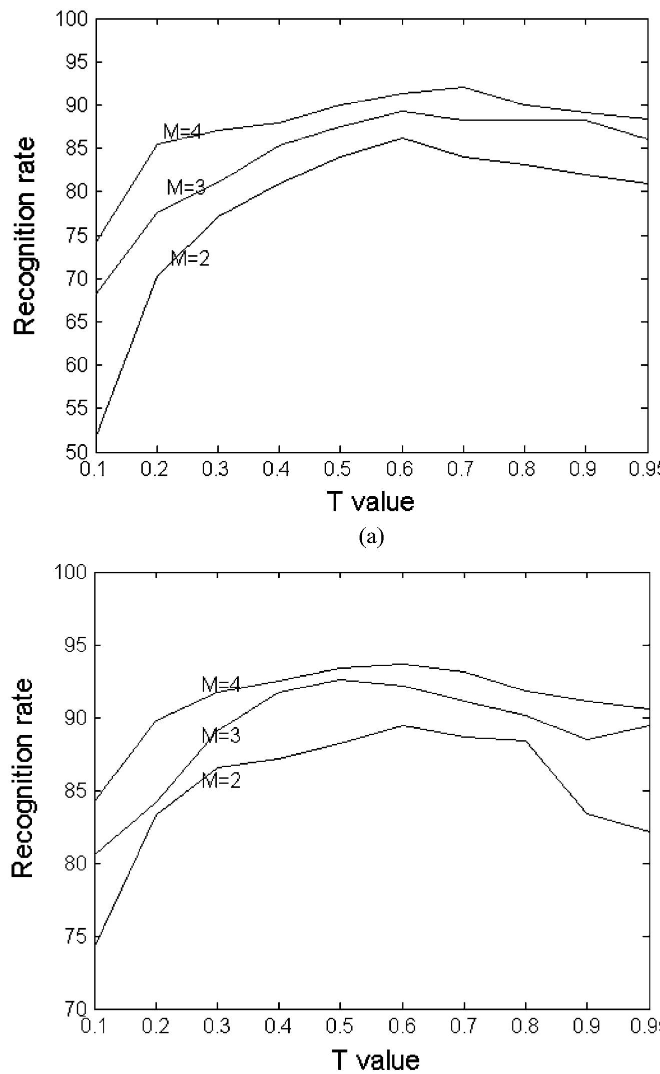
(a)

(b)

Fig. 4. Palmprint image data: (a) demo of a subimage acquired from a palm and (b) ten image samples from one person in the palmprint database.

dpi resolution. Using the preprocessing approach in [27], the subimages with a fixed size (128×128) are extracted from the original images. In order to reduce the computational cost, each subimage is compressed to 64×64. We use these subimages to represent the original palmprint images and to conduct our experiments. Fig. 4(a) displays the demo of a subimage acquired from a palm. Fig. 4(b) shows ten image samples of one person captured at different time. The first five were collected first collections and second five on the next occasion, the major changes being in illumination and position, including shift and rotation. Similar to the kinds of changes encountered in facial expressions, the image may also be slightly affected by the way the hand is posed, shrunk, or stretched.

In the following experiments, the first two samples of every person in each database are used as training samples and the remainder as test samples. Thus, the ORL database provides 80 training samples for and 320 test samples. The palmprint database provides 380 training samples and 2 660 test samples. Generally, it is more difficult to classify patterns when there are fewer training samples. This is also illustrated in Table I, where the recognition rates of the Fisherface methods are worst when the training sample number per class is 2. The experiments



(a)

(b)

Fig. 5. The recognition rates of the third improvement with different image data: (a) ORL face database and (b) palmprint database, while the value of T is varied.

take up that challenge and seek to verify the effectiveness of the proposed approach using fewer training samples.

TABLE II
AN ILLUSTRATION OF FISHER DISCRIMINATIVE VALUES OBTAINED USING THE FISHERFACE METHOD

Different data	Fisher discriminative values									
ORL face database	Number of discrimination vectors: 39									
	1.0000	1.0000	0.9997	0.9981	0.9973	0.9962	0.9950	0.9932	0.9917	0.9885
	0.9855	0.9845	0.9806	0.9736	0.9663	0.9616	0.9555	0.9411	0.9356	0.9151
	0.9033	0.8884	0.8517	0.8249	0.8003	0.7353	0.7081	0.6930	0.6493	0.5515
	0.4088	0.3226	0.2821	0.2046	0.0493	0.0268	0.0238	0.0081	0.0027	
Palmprint database	Number of discrimination vectors: 189									
	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9998	0.9998
	0.9998	0.9997	0.9997	0.9996	0.9996	0.9995	0.9995	0.9994	0.9993	0.9993
	0.9992	0.9991	0.9990	0.9989	0.9987	0.9986	0.9985	0.9983	0.9983	0.9982
	0.9982	0.9979	0.9976	0.9976	0.9974	0.9971	0.9970	0.9968	0.9967	0.9965
	0.9962	0.9960	0.9959	0.9952	0.9948	0.9947	0.9945	0.9943	0.9941	0.9937
	0.9932	0.9930	0.9928	0.9922	0.9917	0.9912	0.9910	0.9908	0.9903	0.9900
	0.9897	0.9892	0.9888	0.9883	0.9878	0.9870	0.9869	0.9862	0.9858	0.9846
	0.9843	0.9836	0.9833	0.9825	0.9822	0.9816	0.9800	0.9795	0.9792	0.9787
	0.9783	0.9767	0.9759	0.9752	0.9743	0.9731	0.9723	0.9718	0.9703	0.9701
	0.9686	0.9679	0.9656	0.9646	0.9635	0.9621	0.9613	0.9605	0.9591	0.9557
	0.9551	0.9535	0.9521	0.9507	0.9486	0.9481	0.9439	0.9436	0.9390	0.9384
	0.9371	0.9331	0.9318	0.9313	0.9273	0.9225	0.9194	0.9186	0.9147	0.9118
	0.9112	0.9088	0.9069	0.9050	0.9036	0.8889	0.8845	0.8821	0.8771	0.8747
	0.8709	0.8659	0.8607	0.8507	0.8488	0.8424	0.8340	0.8280	0.8220	0.8157
	0.8070	0.8007	0.7959	0.7825	0.7751	0.7639	0.7626	0.7434	0.7378	0.7284
	0.7060	0.6944	0.6613	0.6462	0.6372	0.6193	0.6121	0.5663	0.5436	0.5061
	0.4753	0.4668	0.4343	0.3730	0.3652	0.3024	0.2900	0.2273	0.2014	0.1955
	0.1758	0.1541	0.1270	0.1159	0.0858	0.0741	0.0683	0.0591	0.0485	0.0329
	0.0243	0.0205	0.0184	0.0107	0.0090	0.0049	0.0026	0.0004	0.0001	

B. Experiments on the Improvement of Discrimination Vectors Selection

We test the proposed improvement of discrimination vectors selection on the Fisherface method. Table II shows the Fisher discriminative values that are obtained, ranging from 0 to 1. Table III shows a comparison of the classification performance of the proposed improvement and the Fisherface method. The ORL database recognition rate improves 1.25% while that of the palmprint database improves 4.97%. This improvement can further reduce the dimension of discriminative features. There is little difference in the training time of the Fisherface method and the proposed improvement.

C. Experiments on the Improvement of Statistical Uncorrelation of Discrimination Vectors

We also test the proposed improvement to the statistical uncorrelation of discrimination vectors on the Fisherface method. Table III also provides a comparison of the classification performance of this improvement, the Fisherface method and UODV. The recognition rates of UODV and the improvement are the same, but on the ORL database this improvement is 53.45% faster than UODV, and on the palmprint database it is 43.47% faster. The reason for this, as can be seen in Table II, is that only a small number of Fisher discriminative values are equal respectively. In other words, most discrimination vectors obtained from the Fisherface method are statistically uncorrelated, so there is no need to calculate each discrimination vector using UODV. On the other hand, it is necessary to require the vectors to satisfy this favorable property, since, compared with the Fisherface method, our proposed approach can improve recognition rates by 0.31% on the ORL database, and by 7.03% on the palmprint database.

D. Experiments on the Improvement of Principal Components Selection

We test the proposed improvement to principal components selection on the Fisherface method. Table III also provides a comparison of the classification performance of this improvement and the Fisherface method. On the ORL database the improvement increases training time by 7%, and on the palmprint database by 11.32%, but improves recognition rates by 5.31% and 9.55%, respectively. The proposed improvement can also greatly reduce the dimension of discriminative features.

Fig. 5(a) and (b) illustrates the recognition rates of this improvement with different image data: the ORL face database [Fig. 5(a)] and the palmprint database [Fig. 5(b)] while the value of T is varied, where $2 \leq M \leq 4$ (assuming that M is the number of training samples per class). We find that the effective value ranges of the rates for ORL and palmprint databases are [0.4, 0.9] and [0.3, 0.8], respectively. Hence, an appropriate range for both data is [0.4, 0.8]. Table IV shows an analysis of the mean values and the variances of the recognition rates where the value range of T is [0.4, 0.8]. The variances are much smaller than the mean values. In other words, in this range, the recognition effect of our approach is rather robust. Fig. 5 and Table IV also demonstrate the former theoretical estimation in Section II-D, that is, the value range of T might be around 0.5. In the experiments, T is set as 0.6.

E. Experiments on All of the Improvements

ILDA synthesizes all the above improvements on LDA. Fig. 6 displays the demo images of discrimination vectors obtained from different methods on the ORL database. Table V shows a comparison of the classification performance of ILDA

TABLE III
A COMPARISON OF CLASSIFICATION PERFORMANCE OF THREE IMPROVEMENTS ON LDA, THE FISHERFACE METHOD, AND UODV

Classification performance	Different databases	Discrimination methods				
		Improvement 1	*Improvement 2*	*Improvement 3*	Fisherface [8]	UODV[22]
Recognition rates (%)	ORL	82.19	81.25	86.25	80.94	81.25
	Palmprint	86.32	88.38	90.9	81.35	88.38
Extracted feature dimension	ORL	30	39	21	39	39
	Palmprint	160	189	100	189	189
Training time (second)	ORL	14.55	14.7	15.28	14.28	31.58
	Palmprint	36.81	39.06	40.11	36.03	69.1

TABLE IV
AN ANALYSIS OF THE MEAN VALUES AND THE VARIANCES OF THE RECOGNITION RATES IN THE THIRD IMPROVEMENT WHEN THE VALUE RANGE OF T IS [0.4, 0.8]

Different data	ORL face database			Palmprint database		
	Number of training samples per class			Number of training samples per class		
	2	3	4	2	3	4
Mean recognition rates (%)	83.69	87.71	90.25	88.37	91.55	92.89
Variance	1.70	1.31	1.41	0.75	0.88	0.65
Total mean recognition rate (%)	87.22			90.94		
Average variance	1.47			0.76		

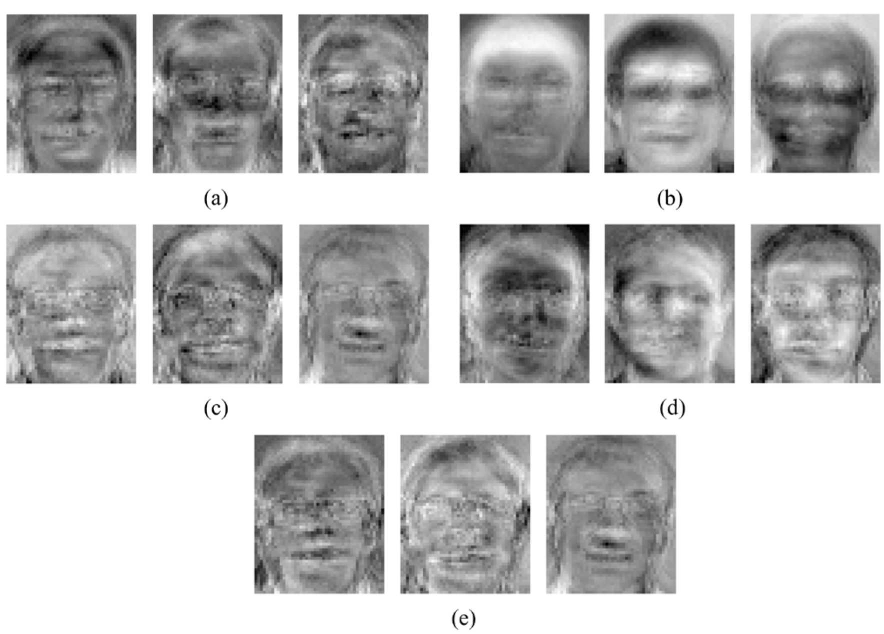


Fig. 6. Demo images of the discrimination vectors obtained from different methods on the ORL database: (a) ILDA; (b) Eigenface (c) Fisherface; (d) DLDA and (e) UODV.

and other methods. Using the ORL face database, the improvements in ILDA's recognition rates over Eigenface, Fisherface, DLDA, and UODV are 5.31%, 6.25%, 4.69%, and 5.94%, respectively. Using the palmprint database, the improvements in ILDA's recognition rates over Eigenface, Fisherface, DLDA and UODV are, again respectively, 18.3%, 12.18%, 19.43%, and 5.15%. In addition, compared with Fisherface, DLDA, and UODV (which uses the second least number of features), ILDA remarkably reduces the feature dimension by 51.28% and 50.26%, respectively for the ORL database and palmprint database. ILDA is much faster than UODV and its training time is rather close to those of Eigenface, Fisherface and DLDA. On the ORL database it is 50.29% faster than UODV, and on the palmprint database it is 39.28% faster. Compared to the Fisherface method, it only adds training time of 9.94% and 16.46%, respectively, for the ORL and palmprint databases.

TABLE V
A COMPARISON OF CLASSIFICATION PERFORMANCE OF ILDA AND OTHER LINEAR DISCRIMINATION METHODS

Classification Performance	Different databases	Discrimination methods				
		ILDA	Eigenface[6]	Fisherface[8]	DLDA[13]	UODV[22]
Recognition rates (%)	ORL	87.19	81.88	80.94	82.5	81.25
	Palmprint	93.53	75.23	81.35	74.1	88.38
Extracted feature Dimension	ORL	19	79	39	39	39
	Palmprint	92	379	189	189	189
Training time (second)	ORL	15.7	13.03	14.28	13.01	31.58
	Palmprint	41.96	32	36.03	37.54	69.1

V. CONCLUSION

This paper presents an ILDA linear discrimination approach for image recognition. ILDA effectively synthesizes three useful improvements on the current linear discrimination technique: it improves the selection of discrimination vectors, adds a measure so that the discrimination vectors satisfy the statistical uncorrelation using less computing time, and provides a strategy to select the principal components. We verify ILDA on different image databases. The experimental results demonstrate that it classifies better than major linear discrimination methods. Compared with the most representative LDA method, the Fisherface method, ILDA improves the recognition rates up to 12.18% and reduces the feature dimension by up to 51.28% while adding only up to 16.46% to training time of the Fisherface method. Consequently, we conclude that ILDA is an effective linear discrimination approach.

APPENDIX

Proof of Theorem 1:

Proof: $W_{\text{pca}}^T S_t W_{\text{pca}}$ is a diagonalized matrix, that is

$$W_{\text{pca}}^T S_t W_{\text{pca}} = \begin{bmatrix} \beta_1^T S_t \beta_1 & 0 & \cdots & 0 \\ 0 & \beta_2^T S_t \beta_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \beta_p^T S_t \beta_p \end{bmatrix} \tag{A1}$$

and

$$W_{\text{pca}}^T S_b W_{\text{pca}} = \begin{bmatrix} \beta_1^T S_b \beta_1 & \cdots & \beta_1^T S_b \beta_p \\ \vdots & \ddots & \vdots \\ \beta_p^T S_b \beta_1 & \cdots & \beta_p^T S_b \beta_p \end{bmatrix}. \tag{A2}$$

So, we have

$$\begin{aligned}&\left(W_{\text{pca}}^T S_t W_{\text{pca}}\right)^{-1} \left(W_{\text{pca}}^T S_b W_{\text{pca}}\right) \\ &\quad = \begin{bmatrix} \beta_1^T S_b \beta_1 / \beta_1^T S_t \beta_1 & \cdots & \beta_1^T S_b \beta_p / \beta_1^T S_t \beta_1 \\ \vdots & \ddots & \vdots \\ \beta_p^T S_b \beta_1 / \beta_1^T S_t \beta_1 & \cdots & \beta_p^T S_b \beta_p / \beta_p^T S_t \beta_p \end{bmatrix}.\end{aligned} \tag{A3}$$

Due to both (6) and (A3), we obtain (7). □

REFERENCES

[1] D. L. Swets and J. J. Weng, "Using discrimination eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 831–836, Aug. 1996.
[2] M. Loog, R. P. W. Duin, and R. Haeb-Umbach, "Multiclass linear dimension reduction by weighted pairwise fisher criteria," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 762–766, July 2001.
[3] A. Vailaya, H. Zhang, C. Yang, F. Liu, and A. K. Jain, "Automatic image orientation detection," *IEEE Trans. Image Processing*, vol. 11, pp. 746–755, July 2002.
[4] K. Nishino, Y. Sato, and K. Ikeuchi, "Eigen-texture method: Appearance compression and synthesis based on a 3D model," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 1257–1265, Nov. 2001.
[5] A. K. Jain, R. P. W. Duin, and J. Mao, "Statitical pattern recognition: A review," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 4–37, Jan. 2000.
[6] M. Turk and A. Pentland, "Eigenfaces for recognition," *Int. J. Cog. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.
[7] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 228–233, Feb. 2001.
[8] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherface: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 711–720, July 1997.
[9] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, pp. 178–188, 1936.
[10] C. Liu and H. Wechsler, "Robust coding scheme for indexing and retrieval from large face databases," *IEEE Trans. Image Processing*, vol. 9, pp. 132–137, Jan. 2000.
[11] ——, "A shape- and texture-based enhanced Fisher classifier for face recognition," *IEEE Trans. Image Processing*, vol. 10, pp. 598–608, Apr. 2001.
[12] L. Chen, H. M. Liao, M. Ko, J. Lin, and G. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognit.*, vol. 33, no. 10, pp. 1713–1726, 2000.
[13] 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.
[14] D. H. Foley and J. W. Sammon, "An optimal set of discrimination vectors," *IEEE Trans. Comput.*, vol. C-24, pp. 281–289, Mar. 1975.
[15] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic, 1990.
[16] Y. Q. Cheng, Y. M. Zhuang, and J. Y. Yang, "Optimal fisher discrimination analysis using the rank decomposition," *Pattern Recognit.*, vol. 25, no. 1, pp. 101–111, 1992.
[17] K. Liu, Y. Q. Cheng, J. Y. Yang, and X. Liu, "An efficient algorithm for Foley-Sammon optimal set of discrimination vectors by algebraic method," *Int. J. Pattern Recognit. Artific. Intell.*, vol. 6, no. 5, pp. 817–829, 1992.
[18] K. Liu, Y. Q. Cheng, and J. Y. Yang, "Algebraic feature extraction for image recognition based on an optimal discrimination criterion," *Pattern Recognit.*, vol. 26, no. 6, pp. 903–911, 1993.
[19] Z. Jin, J. Yang, Z. Hu, and Z. Lou, "Face recognition based on the uncorrelated discrimination transformation," *Pattern Recognit.*, vol. 34, no. 7, pp. 1405–1416, 2001.
[20] Z. Jin, J. Yang, Z. Tang, and Z. Hu, "A theorem on the uncorrelated optimal discrimination vectors," *Pattern Recognit.*, vol. 34, no. 10, pp. 2041–2047, 2001.

[21] X. Y. Jing, D. Zhang, and Z. Jin, "Improvements on the uncorrelated optimal discriminant vectors'," *Pattern Recognit.*, vol. 36, no. 8, pp. 1921–1923, 2003.

[22] ——, "UODV: Improved algorithm and generalized theory," *Pattern Recognit.*, vol. 36, no. 11, pp. 2593–2602, 2003.

[23] X. Y. Jing, D. Zhang, and J. Y. Yang, "Face recognition based on a group decision-making combination approach," *Pattern Recognit.*, vol. 36, no. 7, pp. 1675–1678, 2003.

[24] D. Zhang, H. Peng, J. Zhou, and S. K. Pal, "A novel face recognition system using hybrid neural and dual eighefaces methods?," *IEEE Trans. Syst., Man, Cybern. A*, vol. 32, pp. 787–793, Nov. 2002.

[25] W. Malina, "Two-parameter Fisher criterion," *IEEE Trans. Syst., Man, Cybern. B*, vol. 31, pp. 629–636, Aug. 2001.

[26] T. Cooke, "Two variations on Fisher's linear discrimination for pattern recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, pp. 268–273, Feb. 2002.

[27] M. H. Yang, "Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods," in *Proc. 5th IEEE Int. Conf. Automatic Face and Gesture Recognition (RGR'02)*, Washington, DC.

[28] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Network*, vol. 13, pp. 1450–1464, Sept. 2002.

[29] J. Yang and J. Y. Yang, "From image vector to matrix: A straightforward image projection technique—IMPCA vs. PCA," *Pattern Recognit.*, vol. 35, no. 9, pp. 1997–1999, 2002.

[30] X. Y. Jing, D. Zhang, and Y. F. Yao, "Improvements on the linear discrimination technique with application to face recognition," *Pattern Recognit. Lett.*, vol. 24, no. 15, pp. 2695–2701, 2003.

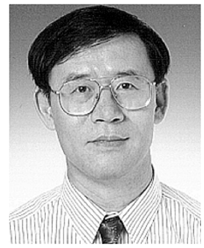
[31] D. Zhang, W. K. Kong, J. You, and M. Wong, "On-line palmprint identification?," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, pp. 1041–1050, Sept. 2003.

[32] G. Lu, D. Zhang, and K. Wang, "Palmprint recognition using eigenpalms features," *Pattern Recognit. Lett.*, vol. 24, no. 9–10, pp. 1463–1467, 2003.

[33] X. Wu, D. Zhang, and K. Wang, "Fisherpalms based on palmprint recognition," *Pattern Recognit. Lett.*, vol. 24, no. 15, pp. 2829–2938, 2003.

Xiao-Yuan Jing received the M.S. and Ph.D. degrees in pattern recognition from Nanjing University of Science and Technology, Nanjing, China.

He was an Associate Professor at the China Academy of Sciences and a Visiting Scholar at Hong Kong Polytechnic University and Hong Kong Baptist University. His current interests include pattern recognition, image processing, neural networks, machine learning, and artificial intelligence.



David Zhang graduated in computer science from Peking University, Peking, China, in 1974. He received the M.Sc. degree in computer science and engineering in 1983 and the Ph.D. degree in 1985, both from Harbin Institute of Technology (HIT), Harbin, China. In 1994, he received the Ph.D. degree in electrical and computer engineering from the University of Waterloo, ON, Canada.

From 1986 to 1988, he was a Postdoctoral Fellow at Tsinghua University, Tsinghua, China, and then an Associate Professor at the Academia Sinica, Beijing, China. He is currently with Hong Kong Polytechnic University, where he is the Founding Director of the Biometrics Technology Centre (UGC/CRC) (www4.comp.polyu.edu.hk/~biometrics/), a body supported by the Hong Kong SAR Government. He also serves as Adjunct Professor at Tsinghua University, Shanghai Jiao Tong University, Harbin Institute of Technology, and the University of Waterloo. His research interests include automated biometrics-based authentication, pattern recognition, and biometric technology and systems. He is the author of more than 130 journal papers, 20 book chapters, and ten books. He holds several patents in both the U.S. and China.

Dr. Zhang is the Founder and Editor-in-Chief of the *International Journal of Image and Graphics* (IJIG) (www.worldscinet.com/ijig/ijig.shtml); Book Editor for Kluwer International Series on Biometrics (KISB) (www.wkap.nl/prod/s/KISB); and Program Chair for the First International Conference on Biometrics Authentication (ICBA). He is an Associate Editor of more than ten international journals, including the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, the IEEE TRANSACTIONS ON PATTERN RECOGNITION. As a principal investigator, he brought to fruition many biometrics projects and won numerous prizes. In 1984, his Fingerprint Recognition System won the National Scientific Council of China's third prize, and in 1986 his Real-Time Remote Sensing System took the Council's first prize. In 2002, his Palmprint Identification System won a Silver Medal at the Seoul International Invention Fair, followed in 2003 by taking a Special Gold Award, a Gold Medal, and a Hong Kong Industry Award. He is a current Croucher Senior Research Fellow.

Yuan-Yan Tang (S'88–M'88–SM'96–F'04) received the B.S. degree in electrical and computer engineering from Chongqing University, Chongqing, China, the M.Eng. degree in electrical engineering from the Graduate School of Post and Telecommunications, Beijing, China, and the Ph.D. degree in computer science from Concordia University, Montreal, QC, Canada.

He is presently a Professor in Department of Computer Science, Hong Kong Baptist University, and an Adjunct Professor in Computer Science at Concordia University. He is an Honorary Lecturer at the University of Hong Kong and an Advisory Professor at many institutes in China. His current interests include wavelet theory and applications, pattern recognition, image processing, document processing, artificial intelligence, parallel processing, Chinese computing, and VLSI architecture. He has published more than 250 technical papers and is the author/co-author of 21 books/book-chapters on subjects ranging from electrical engineering to computer science.

Dr. Tang has served as General Chair, Program Chair, and Committee Member for many international conferences. He will be the General Chair of the 19th International Conference on Pattern Recognition (ICPR'06). He is the Founder and Editor-in-Chief of the *International Journal on Wavelets, Multiresolution, and Information Processing* (IJWMIP) and Associate Editor of several international journals related to pattern recognition and artificial intelligence. He is a Fellow of the IAPR.