DISCRIMINANT IRIS FEATURE AND SUPPORT VECTOR MACHINES FOR IRIS RECOGNITION

Byungjun Son, Hyunsuk Won, Gyundo Kee,Yillbyung Lee

Division of Computer and Information Engineering, Yonsei University
134 Shinchon-dong, Seodaemoon-gu, Seoul 120-749, KOREA
{sonjun, basket98, kigd, yblee}csai.yonsei.ac.kr

ABSTRACT

In an iris recognition system, the size of the feature set is normally large. As dimensionality reduction is an important problem in pattern recognition, it is necessary to reduce the dimensionality of the feature space for efficient iris recognition. In this paper, we present one of the major discriminative learning methods, namely, Direct Linear Discriminant Analysis (DLDA). Also, we apply the multiresolution wavelet transform to extract the unique feature from the acquired iris image and to decrease the complexity of computation when using DLDA. The Support Vector Machines (SVM) approach for comparing the similarity between the similar and different irises can be assessed to have the feature's discriminative power. In the experiments, we have showed that that the proposed method for human iris gave a efficient way of representing iris patterns.

1. INTRODUCTION

Most of works on personal identification and verification by iris patterns have been done in 1990s, and recent noticeable studies among them include those of [1], [2] and [3]. Daugman and Wildes implemented a whole system for personal identification or verifications including the configuration of image acquisition device, whereas the Boles system only focused on the iris representation and matching algorithm without an image acquisition module.

We discuss feature extraction strategies for a class of iris features with very high dimension. In a iris recognition system, one may try to use large feature set to enhance the iris recognition performance. However, the increase in the number of the iris features has caused other problems. For example, the recognizer using higher dimension feature set requires more parameters to characterize the classifier and requires more storage. Thus, it will increase the complexity of computation and make its real-time implementation more difficult and costly. A larger amount of data is needed for training. To avoid these problems, a number of dimensionality reduction algorithms have already been proposed to obtain compact feature set.

Feature extraction methods reduce the dimensionality by projecting the original feature space into a smaller subspace through a transformation. Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) are two major methods used to extract new features [4]. In this paper, we use wavelet and direct discriminant analysis for high-dimensional data set [5]. Section 2 briefly describes the image preprocessing which mainly involves the quality check of an image and iris localization. In section 3, we overview a multilevel two-dimensional 2D discrete wavelet transform (DWT) to obtain the feature vector of an iris image with lower dimensionality and more robust features. Also, we describe the DLDA scheme to linearly transform the iris subimages obtained by wavelet transform to new feature space with higher separability and lower dimensionality. The same operations of DWT and DLDA are performed in training as well as testing phases. Section 4 describes iris matching based on SVM. Experimental results and analysis will be stated in section 5, and finally the conclusions are given in section 6.

2. IMAGE PREPROCESSING

The images acquired from an image acquisition device always contain not only the appropriate images but also some inappropriate ones. Therefore, we need to check the quality of eye image to determine whether the given images are appropriate for the subsequent processing or not and then to select the proper ones among them in real time. Some images ascertained as inappropriate ones are excluded from the next processing.

The images excluded from the subsequent processing include as follows; the images with the blink Fig.1.(a), the images whose the pupil part is not located in the middle

This work was supported in part by Brain Neuroinformatics Program sponsored by KMST.

and some parts of the iris area disappear Fig.1.(b), the images obscured by eyelids or the shadow of the eyelids Fig.1.(c), and the images with severe noises like Fig.1.(d).

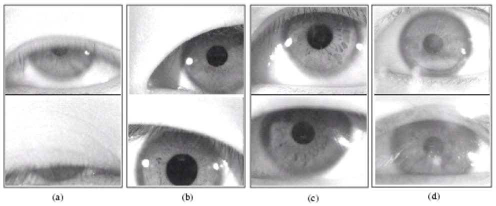

Fig. 1. Example of images with bad quality: (a) the images with the blink (b) the images whose the pupil part is not located in the middle (c) the images obscured by eyelids or the shadow of the eyelids (d) the images with severe noise.

An iris area can be localized from the eye image passed in the quality check step by separating the part of an image between the inner boundary and outer boundary. Fig.2. shows the results of finding the inner boundary, the outer boundary and the collarette boundary in the eye image and the image of iris area, where is used in feature extraction, localized by using these boundaries.

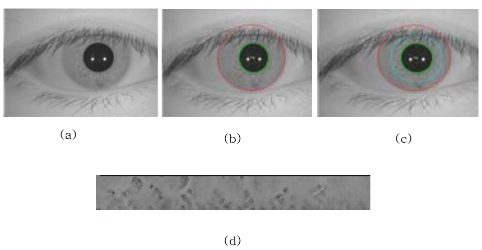

Fig. 2. (a) Original eye image (b) Image of the inner boundary and outer boundary (c) Image of the collarette boundary (d) localized iris image.

3. FEATURE EXTRACTION

Most applications emphasize finding a feature set that produces efficient and implementable results. If the dimension of features defining a problem is too high, we must select a robust set of features from an initial set to provide appropriate representation. We also must design an appropriate classifier to the selected features set. We have chosen the DWT and DLDA approach to obtain a robust and lower dimensional set of features with high discriminating power.

3.1. Wavelet Transform

The hierarchical wavelet functions and its associated scaling functions are to decompose the original signal or image into different subbands. The decomposition process is recursively applied to the subbands to generate the next level of the hierarchy. The traditional pyramid-structured wavelet transform decomposes a signal into a set of frequency channels that have narrower bandwidths in the lower frequency region. The DWT was applied for texture classification and image compression because of its powerful capability for multiresolution decomposition analysis. The wavelet decomposition technique can be used to extract the intrinsic features for iris recognition. We employ the multilevel 2D Daubechies wavelet transform to extract the iris features. Using the wavelet transform, we decompose the image data into four subimages via the high-pass and low-pass filtering with respect to the column vectors and the row vectors of array pixels. Fig.3. shows the process of pyramid-structured wavelet decomposition.

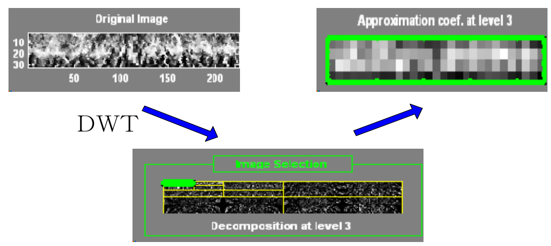

Fig. 3. Example of a 3-level wavelet transform of the iris image

In this paper, we always select low frequency subimage for further decomposition. The three-level lowest frequency subimage is extracted as the feature vector. Generally, low frequency components represent the basic figure of an image, which is less sensitive to varying images. These components are the most informative subimages gearing with the highest discriminating power. The level of low frequency subimage chosen to extract the feature vector depends on size of the localized iris image. If the size is smaller then our localized iris image, the one or two-level lowest frequency subimage might have higher discriminating power.

3.1. Direct Linear Discriminant Analysis

After extraction of the iris feature vector by wavelet transform, the original iris vector x of 7200 dimensions is transformed to the feature vector y of 113 dimensions. To further reduce the feature dimensionality and enhance the class discrimination, we apply DLDA to convert the feature vector y into a new discriminant vector z with lower dimensions then the feature vector y. Existing LDA methods first use PCA to project the data into lower dimensions, and then use LDA to project the data into an even lower dimension. The PCA step, however, can remove those components that are useful for discrimination. The key idea of DLDA method is to discard the null space of between-class scatter S_b - which contains no useful information - rather than discarding the null space of S_w , which contains the most discriminative information[5]. Each scatter is given as follows:

$$S_b = \sum_{i=1}^{J} n_i (\mu_i - \mu)(\mu_i - \mu)^T \qquad (n \times n)$$

$$S_w = \sum_{i=1}^{J} \sum_{x \in C_i} (x - \mu_i)(x - \mu_i)^T \qquad (n \times n)$$

where n_i is the number of class i feature vectors, μ_i is the mean of class i, μ is the global mean, and J is the number of classes.

The DLDA method is outlined below. We do not need to worry about the computational difficulty that both scatter matrices are too big to be held in memory because the dimensionality of input data is properly reduced by wavelet transform.

First, we diagonalize the S_b matrix by finding a matrix V such that

$$V^T S_b V = D$$

where the columns of V are the eigenvectors of S_b and D is a diagonal matrix that contains the eigenvalues of S_b in decreasing order. It is necessary to discard eigenvalues with 0 value and their eigenvectors, as projection directions with a total scatter of 0 do not carry any discriminative power at all [5].

Let Y be the first m columns of V (an $n \times m$ matrix, n being the feature space dimensionality),

$$Y^T S_b Y = D_b \qquad (m \times m)$$

where D_b contains the m non-zero eigenvalues of S_b in decreasing order and the columns of Y contain the corresponding eigenvectors.

The next step is to let $Z = YD_b^{-1/2}$ such that $Z^T S_b Z = I$. Then diagonalize the matrix $Z^T S_W Z$ such that

$$U^T (Z^T S_w Z) U = D_w \qquad (1)$$

where $U^T U = I$. D_w may contain zeros in its diagonal. We can sort the diagonal elements of D_w and discard some eigenvalues in the high end, together with the corresponding eigenvectors.

We compute the LDA matrix as

$$A = U^T Z^T \qquad (2)$$

Note that A diagonalizes the numerator and denominator in Fisher's criterion.

Finally, we compute the transformation matrix (2) that takes an $n \times 1$ feature vector and transforms it to an $m \times 1$ feature vector.

$$x_{recuced} = D_b^{-1/2} A x \qquad (3)$$

4. SVM-BASED PATTERN MATCHING

We only give here a brief presentation of the basic concepts needed. The reader is referred to [7] for a list of applications of SVMs. SVMs are based on structural risk minimization, which is the expectation of the test error for the trained machine. This risk is represented as $R(\alpha)$, α being the parameters of the trained machine. Let β be the number of training patterns and $0 \le \gamma \le 1$. Then, with probability $1 - \gamma$ the following bound on the expected risk holds:

$$R(\alpha) \le R_{emp}(\alpha) + \sqrt{\frac{h(\log(2\beta / h) + 1 - \log(\gamma / 4))}{\beta}} \qquad (4)$$

$R_{emp}(\alpha)$ being the empirical risk, which is the mean error on the training set, and γ is the VC dimension. SVMs try to minimize the second term of (4), for a fixed empirical risk.

For the linearly separable case, SVM provides the optimal hyperplane that separates the training patterns. The optimal hyperplane maximizes the sum of the distances to the closest positive and negative training patterns. This sum is called *margin*. In order to weight the cost of missclassification an additional parameter is introduced. For the non-linear case, the training patterns are mapped onto a high-dimensional space using a kernel function. In this space the decision boundary is linear. The most commonly used kernel functions are polynominals, gaussian, sigmoidal functions. Our experiment was performed using SVM with Gaussian kernel.

5. EXPERIMENTAL RESULTS

Eye images were acquired through CCD camera with LED (Light-Emitting Diode) lamp around lens under indoor light. The size of eye images is pixels with 256 grey intensity values, and the size of normalized iris images is 32×225 pixels. Our data set consists of 1200 eye data acquired from 120 individuals (left and right eye). In case of individuals with glasses, images are captured both removing their glasses and having their glasses; however, contact lenses remained place.

We randomly choose five images per person for training, the other five for testing. To reduce variation, each experiment is repeated at least 20 times. We applied DWT+PCA, LDA, DWT+LDA, DLDA, and DWT +DLDA to a training set. Also, we evaluated the recognition performances using nearest feature classifier and SVM.

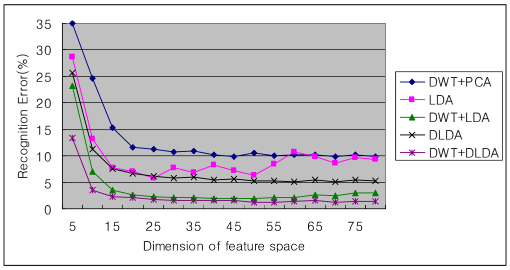

Fig. 4. Recognition Error Rate vs. Dimension of feature space.

Fig.4. shows the result of recognition error rate vs. dimension of feature space. The best recognition rate of the LDA approach is 93% with feature vector consisted of 20 components. The DLDA approach achieves 94.89% recognition rate with feature vector consisted of 70 components. The DWT+PCA, DWT+LDA and DWT+DLDA approach achieve 90.17%, 98.04% and 98.76% with 80, 40 and 55 features, respectively. From Fig.4, We can also see the DWT +DLDA method achieves the best recognition rate. The recognition rates of these DLDA methods are almost fixed after the number of features reaches around 25. In addition, recognition rate of the DWT+DLDA approach is 98.24% when the number of features is 25. It is higher than the best performance of the others and has lower dimension than others. This shows that the DWT+DLDA approach can achieve better performance although it uses smaller number of basis vectors than the others.

Dimension	DWT+DLDA+NN	DWT+DLDA+SVM
5	86.6	81.61
10	96.41	95.25
15	97.69	96.92
20	97.97	98.03
25	98.24	98.12
30	98.4	98.52
35	98.47	98.55
40	98.44	98.61
45	98.34	98.56
50	98.69	99.21
55	98.76	99.4
60	98.51	99.13
65	98.48	98.61
70	98.76	99.35
75	98.63	99.15
80	98.66	99.18

Fig. 5. DWT+DLDA+NN vs. DWT+DLDA+SVM.

As compared in Fig.5, we find that DWT+DLDA +SVM outperforms DWT+DLDA+NN. Such observations are almost consistent for different numbers of feature set. The best recognition rate of DWT+DLDA+SVM approach is 99.4% when the number of features is 55.

6. CONCLUSION

In this paper, we have presented effective methods for iris recognition. We specifically uses the multiresolution decomposition of 2-D discrete wavelet transform for extracting the robust feature set of low dimensionality. In addition, the DLDA method is used to obtain the feature set with higher discriminative power and lower dimensionality. These methods of feature extraction well suit with iris recognition system while allowing the algorithm to be translation and rotation invariant.

We showed that the DWT+DLDA method outperformed the PCA, LDA and DLDA, and DWT+DLDA in terms of classification rate. For the complex data consisting of many classes in the problem of iris recognition, the DWT+DLDA method can be used for an alternative of LDA. DWT +DLDA +SVM has the advantage of efficient testing, and good performance compared to other linear classifiers. For future works, it is necessary to conduct experiments on a large number of data so as to verify the efficiency and robustness of our approach.

7. REFERENCES

[1] John G. Daugman., “High confidence visual recognition of persons by a test of statistical indenpendence,” *IEEE Trans. On Pattern Analysis and Machine Intelligence,* 15(11):1148--1161, January 1993.

[2] R. P. Wildes., “Iris Recognition : An Emerging Biometric Technology, ” *Proc. of the IEEE*, 85(9):1348--1363 , 1997.

[3] W. W. Boles, B. Boashash, “A Human Identification Technique Using Images of the Iris and Wavelet Transform, ” *IEEE Trans. of Signal Processing,* 46(4):1185--1188, 1998.

[4] W. L. Poston and Marchette, D. J., “Recursive Dimensionality Reduction Using Fisher's Linear Discriminant,” *Pattern Recognition*, 31(7):881--888, 1988.

[5] Jie Yang, Hua Yu., “A Direct LDA Algorithm for High-Dimensional Data - with Application to Face Recognition,” *Pattern Recognition*, 34(10):2067--2070, 2001.

[6] Dimitrios Ioammou, Walter Huda, Andrew F. Laine., “Circle recognition through a 2D Hough Transform and radius histogramming, ” *Image and Vision Computing*, 7:15--26, 1999.

[7] A. Tefas , C. Kotropoulos , and I. Pitas., “Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication, ” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(7): 2001.