Ear Recognition based on 2D Images

Li Yuan, Zhi-chun Mu

Abstract—Research of ear recognition and its application is a new subject in the field of biometrics authentication. Ear normalization and alignment is a fundamental module in the ear recognition system. Traditional manually normalization method is not suitable for an automatic recognition system. In this paper, an automatic ear normalization method is proposed based on improved Active Shape Model (ASM). This algorithm is applied on the USTB ear database for ear normalization. Then Full-space Linear Discriminant Analysis (FSLDA) is applied for ear recognition on the normalized ear images with different rotation variations. Experiments are performed on USTB ear image database. Recognition rates show that based on the right ear images, the acceptable head rotation range for ear recognition is between the right rotation of 20 degree to the left rotation of 10 degree.

I. INTRODUCTION

Earlier research has shown that human ear is one of the representative human biometrics with uniqueness and stability [1]. As an emerging biometrics technology, ear recognition is attracting more and more attention in biometrics recognition [2]. An ear recognition system usually involves ear detection, feature extraction and ear recognition/verification modules.

At present, there have already been some existing ear feature extraction methods for 2D images as follows: Alfred Iannarella presented a 12-measurement method, which is not suitable for machine vision because of the difficulty of localizing the anatomical points [1]. Moreno used feature points of outer ear contour and information obtained from ear shape and wrinkles for ear recognition [2]. Burge and Burger used the main curve segments to form Voronoi diagram, and build adjacency graph out from Voronoi diagram. Then a graph matching based algorithm was used for authentication [3]. Hurley treated the ear image as an array of mutually attracting particles that act as the source of Gaussian force field. The original image is described by a set of potential channels and positions of potential wells [4]. Wang used a high order moment-invariant method to extract ear features [5]. Mu presented a long axis based shape and structural feature extraction method, the shape feature is consisted of the curve fitting parameters of the outer ear contour, the structural feature is composed of ratios of between the length of key sections and the length of the long axis [6]. Chang used standard PCA algorithm for ear recognition, and gets the conclusion that ear and face does not have much difference on recognition rate (71.4% and 70.6% respectively on image database from USF) [7].

For ear recognition in 3D, Hui Chen [8] proposed an ICP based 3D ear recognition. Pin Yan [9] proposed an automatic 3D ear recognition system, which includes automatic ear extraction, and ICP based ear recognition.

In prior 2D ear recognition, the pre-processing of ear images has been a manual process, which is not applicable for an automatic system. The 2D ear recognition also has the problem caused by rotation variations. So this paper includes two major parts: automatic ear extraction from 2D images using improved active shape model, and the study of the acceptable head rotation range for ear recognition.

II. EAR EXTRACTION AND NORMALIZATION

In this section, an automatic ear normalization method is proposed based on improved Active Shape Model (ASM). In the offline training step, ASM is used to create the Point Distributed Model using the landmark points on the ear images of the training set. This model is then applied on the test ear images to search the outer ear contour. The final step is to normalize the ear image to standard size and direction according to the long axis of outer ear contour. The long axis is defined as the line crossing through the two points which have the longest distance on the ear contour. After normalization, the long axis of different ear images will be normalized to the same length and same direction.

A. ASM based Outer Ear Contour Extraction

Using the long axis as the standard for normalization requires accurate location of the outer ear contour. Active Shape Model is a powerful statistical tool for object location with higher accuracy and robustness. It mainly relies on Principle Component Analysis to statistically model the variability in the training set of example shapes, and then iteratively deforms to fit an example of the object in a new image. The outer ear contour extraction process has the following four steps.

1. Ear Landmarks for Statistical Shape Models

A shape model is described by n landmark points that represent the boundary internal features, or even external features. Each training shape $X_i$ is represented as a vector:

$$X_i = (x_{i1}, x_{i2}, ..., x_{in}, y_{i1}, y_{i2}, ..., y_{in})^T, (i = 1, 2, ..., N) \quad (1)$$

2. Point Distributed Model
ASM creates a Point Distributed Model using the statistical feature of the landmark points [10]. The training samples need to be aligned so that the corresponding points on different samples are comparable. The training shapes are all aligned by scaling, rotation and translation for minimizing the weighted sum of the squared distances between their equivalent landmark points. Here, the similarity transformation is applied using three factors: scaling factor $s$, rotation factor $\theta$ and translation factor $t$. As shown in Fig. 1, the shape distribution is more of central tendency after alignment.

![Fig. 1. Trainings samples before and after alignment: (a) before alignment, (b) before alignment.](image)

After alignment, principle component analyses is applied to the aligned shape. Therefore, an ear shape model can be approximated as follows:

$$X = T_{t,s,\theta}(X + pb)$$

where $\bar{X}$ is the mean shape, $p$ is a set of orthogonal models of shape variation and $b$ is a vector of shape parameters. $T_{t,s,\theta}$ is a translation including rotating by $\theta$, scaling by $s$ and translation of $(x_i, y_i)$.

The shape vector will be different when parameter $b$ changes. Generally, the variation range for $b$ is $-3\sqrt{\lambda_i} \leq b \leq 3\sqrt{\lambda_i}$, so that the shape is similar to the ear, where $\lambda_i$ is the eigen value of the shape covariance matrix.

Fig. 2 shows how the shape model changes when $b$ is different.

![Fig. 2. Different shape models when $b$ changes.](image)

3. Improved ASM for Outer Ear Contour Extraction

Here, a contour based gray level matching algorithm is used to make the shape model quickly find the matching feature point. For each landmark point on the training shape, gray level sampling is applied along the normal direction of the boundary. For the $j$th landmark point on the $i$th training image, we sample $k$ points along normal direction on each side, so a vector is composed of the $2k+1$ points, and its normalized difference is $y_j$. For the $j$th landmark point of $N$ images, the mean value is $\bar{y}_j$, the covariance is $S_{y_j}$, thus we get the gray level statistical model for this landmark point. Apply the same method on each landmark point, we get the statistical model of all the landmark points on the target contour.

In this paper, an iterative scheme is applied for searching the outer ear contour with the following steps:

1) Initialize the model shape. Firstly, set $b = 0$ , so $X = T_{t,s,\theta} \bar{X}$. Then adjust $s_i$, $\theta_i$ and $t_i$ to make the initial model $X_i$ as similar to an actual outer ear contour as possible;  
2) According to the Gray Level Average Model of each point, we can get its matching point, and then a new shape vector $X_i$ is formed;  
3) Get the new parameters $s_i, \theta_i, t_i$ and $b$ for $X_i$;  
4) Update $s_i, \theta_i, t_i$ and $b$ of $X_i$ with the new parameters, and a new shape will be formed.  
5) If the change of the shape is smaller than a threshold or a given iteration number is reached, the algorithm is converged. Otherwise, repeat step (2)-(4). Fig. 3 shows the searching process.

![Fig. 3. Searching process of the model.](image)

In ASM, local texture model matching is conducted under the assumption that the normalized first derivative profile $\{ y_j \}$ satisfies a Gaussian distribution. The matching degree of a probe sample $y_i$ to the reference model is given by $f(y_i) = (y_i - \bar{y}_j)^T S^{-1}_{y_j} (y_i - \bar{y}_j)$, where $\bar{y}_j$ is the mean of $\{ y_j \}$ for $1 \leq j \leq N$, and $N$ is the number of images for establishing Point Distribution Model (PDM), $S_{y_j}$ is the covariance [11]. The main idea of finding the matching point is to minimize $f(y_i)$ which is equivalent to minimizing the probability that $y_i$ comes from the distribution.

For most images interpretation task, the landmarks are usually selected in the points that have strong edge information. So the destination points should have also strong edge information. These points tend to lie on the edges.
extracted by an edge operator. Thus we can adjust the matching method by adding a weight to the Mahalanobis distance function:

\[ f(y) = (c e)(y_i - \bar{y}_i)^T \Sigma_y^{-1}(y_i - \bar{y}_i) \]  

(3)

where \( c \) is a constant, \( e \) is the Sobel edge intensity at the target point.

With this improvement, points with strong edge information are more probably chosen as the best candidate. For the ear images, the landmarks on the outer ear contour part have clearly strong edge intensity, while the landmarks on other parts are not necessary with this property. Therefore, we apply the edge constraints to the landmarks on the outer ear contour. Through experiment, we get the statistical value of \( c \), \( c = 260 \). As illustrated in Fig. 4, most mismatched points on the contour in the standard ASM have been correctly matched by applying the improved matching function.

![Fig. 4. Search result before and after match function improved (a) Sobel edge extraction, (b) conventional matching function, (c) improved matching function.](image)

**B. Ear Normalization**

After the outer ear contour is found using ASM, we choose 30 feature points on the contour. Search the longest distance between two points \((x_i, y_i)\) and \((x_j, y_j)\) among the 30 points on the outer ear contour, connect the two points to form the long axis, the length of the long axis is defined as \( LA \), as shown to be the AB in Fig. 5(a). After normalization, \( LA \) for each ear is the same, and the direction of \( LA \) is perpendicular to axis \( X \).

The normalization process has the following steps:

1) Rotate the image, so that the long axis is perpendicular to axis \( X \), as shown in Fig. 5(b).
2) Through experiment, we found that the mean value of the ration between ear height and ear width is 1.95. So in this paper, ear is normalized to be \( 82 \times 160 \) pixels. So the test image is zoomed to make sure that the long axis is 160, as shown in Fig. 5(c).
3) Extract the ear using four points: two points on the long axis \((x_i, y_i), (x_j, y_j)\), the left most point on the outer ear contour, and the right most point on the outer ear contour, as shown in Fig. 5(d).
4) Normalize the ear to be \( 82 \times 160 \) pixels, as shown in Fig. 5(e).
5) Lighting normalization using the histogram equalization method, as shown in Fig. 5(e).

![Fig. 5. Process of ear normalization: (a) original image; (b) rotation; (c) scaling, \( LA = 160 \); (d) ear extraction; (e) normalization to \( 160 \times 82 \).](image)

**C. Ear Normalization Experiment**

The experiments are performed on the dataset collected by us (USTB dataset) [12]. It contains subset1 and subset2. Subset1 collected ear images of 77 subjects, 4 images for each. The images of each subject are taken under two conditions: illumination variation and orientation variation and individuals were invited to be seated 2m from the camera and change his/her face orientation. The 4 images are 300x400 pixels in size as shown in Fig. 6. Fig. 6(a) and Fig. 6(d) are the profile view of the head under different lighting condition. Fig. 6(b) and Fig. 6(c) are -30 and +30 rotation respectively.

![Fig. 6. Example images of the USTB ear database.](image)

The ear normalization experiments are performed on subset1. In four experiments, the probe images are No.1 to No. 4 respectively, and the remaining three images for each subject are utilized to establish the PDM. Then, the PDM is utilized by the proposed ASM to search a shape for the one not participating in establishing the PDM. Fig. 7 shows some example images of successful ear normalization. The success rate for the four experiments are 84%, 75%, 86% and 75%.

![Fig. 7. Example images of successful ear normalization.](image)

In practical application, the ear images may be taken under lighting variation or with different accessory. The proposed method works well under these conditions. Fig. 8 (a) and (b) show ear normalization under different light conditions. Fig. 8 (c) and (d) show ear normalization with earrings and glasses.
III. EAR RECOGNITION

In this paper, a full-space linear discriminant analysis (FSLDA) algorithm is applied on the USTB ear database [12] which contains ear images with rotation variations. FSLDA makes full use of the discriminant information in the full-space of the within-class scatter, namely the discriminant information in the null space and non-null space of the within-class scatter.

IV. EXPERIMENTAL RESULTS

The ear recognition experiments use both subset 1 and subset 2. USTB ear subset 2 is a multimodal image database. It contains profile views of 79 subjects. The subjects in subset 1 and subset 2 are different. For each subject in subset 2, we take the right ear images of right rotation at 0, 5, 10, 15, 20, 25, 30, 40, 45 and 60 degrees, and also the right ear images of left rotation at 5, 10, 15, 20, 25, 30, 35, 40 and 45 degrees. Fig. 9 shows some example images.

The ear images from subset 1 are used for training to get the feature space with the FSLDA mentioned in section 3. The ear images of 0 degree from subset 2 are used for the gallery. The ear images of left rotation and right rotation at other degrees from subset 2 are used for the probe.

Fig.10 (a) shows the recognition rate at different degrees with left rotation. Fig.10 (b) shows the recognition rate at different degrees with right rotation. It can be seen that the recognition drops as the rotation angle increases.

When the dimension of the feature space is fixed at 80, Fig. 11 shows the ear recognition rate at different rotation angles. When the head turns left, the recognition rates at 5, 10, 15 and 20 degree are over 90%. The recognition rate drops dramatically after the rotation angles of 20 degree. When the head turns right, the recognition rates at 5 and 10 degree are over 90%. The recognition rates at 15 and 20 degree are over 80%. The recognition rate drops dramatically after the rotation angles of 20 degree. In the range of 20 degree to 60 degree, the recognition rate of right rotation is higher than that of left rotation at the same rotation degree.

![Ear images](image_url)
So in this experiment, the following conclusion can be made: when the right ear images are used for recognition, and if the recognition rate of over 90% is preferred in practical application, the acceptable head rotation range is between 20 degree of left rotation to 10 degree of right rotation.

V. CONCLUSIONS

This paper has presented the ear recognition using 2D images. The improved ASM is applied to solve the ear normalization problem. The main contributions are: (1) automatically search the outer ear contour on the probe ear image; (2) automatic ear extraction based on the long axis of the outer ear contour.

Another contribution is the study of sensitivity to rotation using the Full-space Linear Discriminant Analysis (FSLDA). Experiments are performed on USTB ear image database. Recognition rates show that based on the right ear images, the acceptable head rotation range is right rotation of 20 degree and left rotation of 10 degree if we want to get a desirable recognition rate. This result may be helpful for the design of a practical ear recognition system.

REFERENCES