

Human Ear Recognition from Face Profile Images

Mohamed Abdel-Mottaleb* and Jindan Zhou

Department of Electrical & Computer Engineering, University of Miami,
1251 Memorial Dr., Coral Gables, FL 33146

Abstract. In this paper, we present a novel system for ear identification from profile images of the face. The system has two steps. In the first step, the ear is automatically detected from the profile image of the face. In the second step, the ear image is transformed to a force field, then feature points are extracted and the best match is found from a database. We propose a method based on differential geometry to extract ear feature points. We use a transformation of the ear image to make it suitable for extracting the feature points using differential geometry. During recognition, the feature points obtained from a query image are aligned and compared with those in the database using Hausdorff distance. The experimental results show that our method is effective.

1 Introduction

Ear biometric has the advantage of being invariable with respect to changes in face expressions as compared to face recognition [4]. In 1989, Alfred Ianarelli [8], conducted a study on personal identification using ear features. Iannarelli gathered over 10000 ear images and used 12 features to represent each ear. His experimental results showed that indeed ear features vary from person to person in such a manner that they can be used to distinguish between people. His work provided a foundation for the feasibility of using the ear shape for biometrics.

Some approaches have been developed for ear recognition from 2D intensity images. Chang et al. [10] used principal component analysis for ear recognition. Hurley et al. [4][13] applied force field transform to ear images, then, used the positions of the extracted energy wells as features for recognition.. Burge and Burger [9] used Voronoi diagram and curve segments with a novel subgraph matching algorithm for ear authentication. Recently, Chen and Bhanu [11][12] proposed a local surface descriptor and used the ICP (Iterative Closest Point) procedure for matching 3D ear data from range images.

In this paper, we propose a biometric system based on the shape of the ear. The system automatically detects the ear region from color profile images of the face, and then extracts ear features and applies matching for ear recognition. For feature extraction, we apply a force field transform to the detected ear region and extract, from the transformed image, a set of feature points to represent the ear. The recognition is performed by calculating the Hausdorff distance between the set of feature points that represent the query and those that represent every candidate in a database.

* Corresponding author.

The rest of the paper is organized as follows: Section 2 describes the details of our ear identification system. This includes our method for ear segmentation, feature extraction and matching. Section 3 presents the experimental results of ear identification. The conclusions are given in Section 4.

2 Ear Identification System

2.1 Ear Detection

The goal of this step is to extract a sub-image that contains only the ear region from the profile image of a person's face. The detection of the ear is achieved in three steps: 1) skin-tone region detection, 2) edge detection within the region of the skin-tone followed by size filtering operation, 3) candidate ear region detection using Hausdorff matching with a simplified model of the ear. Then, only one candidate region, which has the minimum mean square error with an average ear image, is selected as the detected ear region. Figure 1 shows an example with the results of the three-step process of ear detection. The details are given in the following subsections.

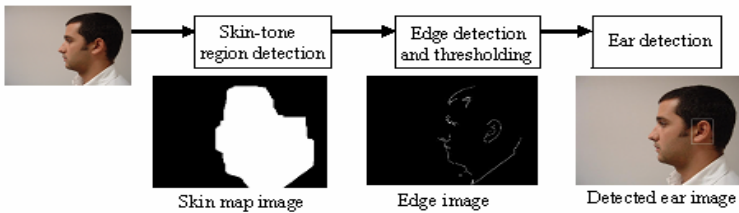


Fig. 1. System diagram for ear detection from profile images

2.1.1 Skin-Tone Region Segmentation

We use Flek's method [1] for skin-tone detection to locate the face region. This skin detection method uses a skin filter which relies on color and texture information.

Since the ear region is not as smooth as other skin regions, sometimes it will not be intact in the output of the skin filter. Therefore, we need to enlarge the output skin region to include the whole ear; this is achieved by applying a morphological dilation operation to the output of the skin filter.

2.1.2 Edge Detection on Skin Region

Canny edge detection is applied only to the region where skin is detected. The result is then filtered using an edge size filter to remove short and isolated edges that appear due to noise. Figure 2 shows an example of the results.

2.1.3 Ear Detection Using Template Matching

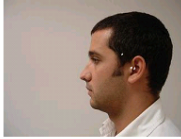
In this step, the ear region is detected using template matching. We use a simplified ear template that consists only of an edge that represents the ear's helix, as shown in Figure 3. Then, we use Hausdorff distance [2] to search for the template in the edge



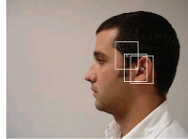
Fig. 2. Edge detecting in the skin-tone region



Fig. 3. Left ear's helix edge template



(a)



(b)

Fig. 4. Template matching and non-max suppression



Fig. 5. Anatomy of ear 1. Helix 2. Lobulo 3. Concha 4. Pliegue inferior 5. Pilegue anterior

map of the skin region, while allowing for affine transformation between the template and the data. The details of Hausdorff distance that we use for ear detection are similar to those we use for the recognition step and are given in Section 2.3.

Direct application of Hausdorff distance is inefficient and computationally expensive. Alternatively, we used the distance transformation, as discussed in [3], to solve this problem. In this case calculating Hausdorff distance is achieved by calculating the cross-correlation between the template and the distance transform of the edge map. Figure 4(a) shows a result of Hausdorff distance matching. The white points are the cross-correlation centers where the Hausdorff distance between the template and the area is smaller than a threshold. The non-max suppression method is applied to prune the matching results as shown in Figure 4(b), where only three points remain.

To obtain the correct ear region from the candidate regions, the candidate regions are scaled to the same size as an average ear image, and then we compute MSE (Mean Square Error) between each detected candidate region and the average ear image. The average ear image is defined as the average intensity of a set of training ear images. The detected region is the candidate region that gives the minimum MSE.

2.2 Ear Feature Extraction

Figure 5 shows the anatomy of the ear, the most distinctive features of the human ear are the shapes and locations of these anatomy structures, such as helix rim, inner ear, antihelix, lobule etc. The following sections present our method for feature extraction.

2.2.1 Force Field Transformation for Ear Feature Extraction

We use a differential geometry based method to extract feature points from ear images. In this approach, the 2D intensity image $f(x, y)$ is treated as a surface. The surface is defined as $S = (x, y, z)$, where the z value is represented by the image intensity at pixel (x, y) , e.g. $z = f(x, y)$. Under this representation, the most distinctive features of the ear can be considered as surface feature points that lie in ridge-ravine lines. The extraction of ridges and ravines of a surface is based on differential geometry of the surface and need computation of principle curvatures.

Since the calculation of principle curvatures involves second derivatives. To get a robust feature extraction results, the image needs to be smoothed while preserving the important features. We use an energy transformation method called force field transformation, which was used in [4], to preprocess the ear image before feature extraction. In this method, the image is transformed by considering the image to consist of an array of Gaussian attractors, which act as the sources of force field. It smoothes the original grayscale ear image and preserves the important structure of the ear as well.

The force field energy function method transforms the whole ear image into a force field by assuming that each pixel exerts an isotropic force on all the other pixels in the image [4]. The equations below summarize the force field transformation.

$$F_i(r_j) = P(r_i) \frac{r_i - r_j}{|r_i - r_j|^3} \tag{1}$$

$$F(r_j) = \sum_{i \in 0, N-1 | i \neq j} F_i(r_j) = \sum_{i \in 0, N-1 | i \neq j} P(r_i) \frac{r_i - r_j}{|r_i - r_j|^3} \tag{2}$$

Where N is the number of pixels in the image, and $F_i(r_j)$ is the exerted force on a pixel at position r_j by any other pixel at position r_i with pixel intensity $P(r_i)$. The force is proportional to the pixel intensity and inversely proportional to the distance from the pixel. Figure 6 is an example of the force field energy transformation. From the surface representation, it is clear that the transformed image is much smoother than the original image and preserves the image features as well.

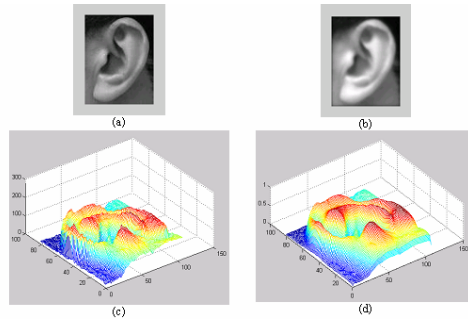


Fig. 6. Force field transformation (a) Original grayscale image and it’s surface representation (c). (b) The force field transformation and it’s surface representation (d).

2.2.2 Ear Image as Surface for Feature Extraction

Assuming a surface is represented by the function $z = f(x, y)$, the most important curvatures of the surface include the Gaussian curvature K , mean curvature H , principal curvatures k_1 and k_2 . The extraction procedure involves calculations of the second derivative as follows, where \vec{n} is the surface normal [5][6]:

$$\vec{n} = \frac{(-f_x, -f_y, 1)}{\sqrt{1 + f_x^2 + f_y^2}} \quad K = \frac{(f_{xx}f_{yy} - f_{xy}^2)}{(1 + f_x^2 + f_y^2)^2} \tag{3}$$

$$H = \frac{f_{xx} + f_{yy} + f_{xx}f_y^2 + f_{yy}f_x^2 - 2f_xf_yf_{xy}}{2(1 + f_x^2 + f_y^2)^{1.5}}$$

$$k_1 = H + \sqrt{H^2 - k} \quad , \quad k_2 = H - \sqrt{H^2 - k}$$

Our goal is to extract and use the points lying in ridge-ravine line as the feature points. These points correspond to the extreme ridge points on the considered surface. An extreme ridge point is a point where k_1 , has a local positive maximum. There are different ways to locate the ridges, here we threshold the k_1 value to find these points.

Figure 7 shows three examples for extracted feature points using the proposed method based on the force field energy image, where the feature points are superimposed on the energy images.



Fig. 7. Extracted feature points superimposed on the energy images

2.3 Recognition

Like most biometric systems, our ear identification system has two stages: off-line stage to prepare and archive the database and online stage in which recognition of an unknown query is performed. During archiving, the feature points extracted from an ear image are stored in a database. In the recognition stage, the test ear image is processed, and the extracted feature points are compared with those in the ear database for a match. We apply bi-directional Hausdorff distance [2] for the matching criterion. The partial bi-directional Hausdorff distance between the two sets of feature points that are transformed with respect to each other, is defined as:

$$H_{LK}(T(P), P') = \max(h_L(P', T(P)), h_K(T(P), P')) \tag{4}$$

Where P is the ear's feature points from the test image, $T(P)$ is the transformation of P , which can be represented as:

$$T(P) = A * P + t \tag{5}$$

$$= \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} * \begin{pmatrix} s & 0 \\ 0 & s \end{pmatrix} * \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

P' is the set of feature points of an ear in the database, $h_K(T(P), P')$ is the partial directed distance from $T(P)$ to P' , where $1 \leq K \leq q$ (q is the number of points of P), and $h_L(P', T(P))$ is the partial directed distance from P' to $T(P)$, $1 \leq L \leq q'$ (q' is the number of points of P'). The definition of partial directed distance, e.g., $h_K(T(P), P')$ denotes the K^{th} ranked value in the set of distances from $T(P)$ to P' . This process

automatically selects the K best matching points in $T(P)$, which means that matching only takes a portion of the feature points into consideration. In practice, to compute the partial directed distance, we specify some fraction f_1 , and f_2 , where $0 < f_1, f_2 < 1$, and let $K=f_1 \times q$, $L=f_2 \times q$. The fractions f_1 and f_2 determine how many missing points between the two sets of feature points can be tolerated.

The matching is performed by finding a rigid transformation, (i.e., rotation, scale and translation), that results in a minimum Hausdorff distance. The database ear image with the minimum Hausdorff distance to the test ear image is considered the best match. The minimum Hausdorff distance is found using Quadratic Programming optimization method [7].

3 Results

For the experiments, we collected a database of profile face images for 103 subjects. For each person, one image was used for training, where the ear region was detected and the extracted ear feature points were stored in the database. During recognition, we used different images from the ones used for the training. The recognition starts by locating the ear region in the test image and extracting the feature points. Then, the feature points are compared with all the sets stored in the database using Hausdorff distance.

The proposed ear recognition method was applied to 58 query images for 29 subjects, two images from each subject. Out of the 58 queries, 51 obtained the correct match as the first match. For the remaining 7 queries, four queries had their correct matches among the top three best matches, while for the other three queries, two obtained their correct matches in the top five best matches and one was ranked the 10th top match.

Our ear identification system was also tested for computational efficiency. The process of ear detection from profile images takes on average 4.6 seconds, i.e., 13 images per minute. The processes of ear recognition takes 3.1 second to calculate the Hausdorff distance between the feature points of a query and the feature points of a database image (this is the average time on the database of the 103 ear images). The result of the computational time is based on a Pentium 4/3GH processor desktop and Matlab platform. The computational time is expected to be much lower if a more efficient language is used.

4 Conclusions

We proposed an ear recognition system for human identification using profile face images. The proposed system includes a new automated method for ear detection. It also includes a new method for extracting ear features by using differential geometry approach applied to a transform of the ear image. The experiments show that the system is effective for ear recognition.

References

1. M. Fleck, D. Forsyth, and C. Bregler, "Finding Naked People," European Conference on Computer Vision, vol. 2, pp. 592-602, 1996.
2. D. P. Huttenlocher, G. A. Klanderman, W. J. Rucklidge, "Comparing images using the Hausdorff distance," IEEE Trans. on PAMI., vol. 15, no. 9, pp. 850-863, 1993.
3. G. Wei, D. Li, I.K. Sethi, "Detection of side-view faces in color images," Proc. Fifth IEEE Workshop on Applications of Computer Vision, pp. 79-84. Dec, 2000.
4. D. J. Hurley, M. S. Nixon and J. N. Carter, "Force Field Energy Functionals for Image Feature Extraction," In Proceedings of Proc. 10th British Machine Vision Conference BMVC99 2, pp. 604-613.
5. R. Huang, T.L. Kumii, "Parallel algorithms for extracting ridges and ravines," proceedings of the First Aizu International Symposium on Parallel Algorithms/Architecture Synthesis. Mar. 1995.
6. A. B. Hamza and H. Krim, "A topological variational model for image singularities," Proc. 2002 IEEE International Conference on Image Processing, 2002.
7. S. P. Han, "A globally convergent method for nonlinear programming," Journal of Optimization Theory and Applications, vol. 22, pp. 297, 1977.
8. A. Ianarelli. "Ear Identification," Forensic Identification Series. Paramount Publishing, California 1989.
9. M. Burge and W. Burger, "Ear Biometrics in Computer Vision," In the 15th International Conference of Pattern Recognition, ICPR 2000, pp. 826-830.
10. K. Chang, K. Bowyer and V. Barnabas, "Comparison and combination of ear and face images in appearance-based biometrics," IEEE Trans. Pattern Analysis and Machine Intelligence. vol. 25, pp.1160-1165, 2003.
11. B. Bhanu and H. Chen, "Human ear recognition in 3D," in Workshop on Multimodal User Authentication, pp. 91-98, 2003.
12. H. Chen and B. Bhanu, "Human Ear Detection from Side Face Range Images," International Conference of Pattern Recognition, ICPR 2004, vol3, p574-577.
13. D. J. Hurley, M. S. Nixon, and J. N. Carter, "Force field feature extraction for ear biometrics," Computer Vision and Image Understanding, 2005.