

3D Ear Recognition Using SIFT Keypoint Matching

Xin Dong⁺ and Yin Guo,

University of Science and Technology Beijing, Beijing, China

Abstract. In this paper, we present a novel algorithm for 3D ear recognition. The basic idea is to rotate each 3D point cloud representing an individual's ear around the x, y or z axes, respectively generating multiple 2.5D images at each step of the rotation. Then we use SIFT descriptors to extract and describe the features of human ears. The test ear images are recognized by the application of a new weighted keypoint matching algorithm. Experimental results show that this method is both accurate and efficient.

Keywords: SIFT, keypoint detection, 2.5D image

1. Introduction

Biometrics is the technology using individuals' physiological or behavioral characteristics to recognize their identity [1]. It is safer and more reliable than traditional password and identification code. Compared to other biometrics, ear biometrics has certain advantages. Previous studies have shown that the ear is a promising new class of relatively stable biometrics. Ear, as a new class of biometrics, has drawn the attention of researchers around the world. Human ear has such characteristics as uniqueness and stability so that they can serve as the basis of identification of individuals. For example, the ear is rich in features; it is a stable structure that does not change much with age [2] and it does not change its shape with facial expressions [3]. In spite of all these advantages, how ear recognition is performed is still not thoroughly understood. In this paper, we propose a new approach which bases recognition on matching and voting keypoints that are extracted from multiple 2.5D views of each 3D ear. Our method is evaluated on the University of Notre Dame public data set (UND dataset) [4].

This paper basically has two important contributions.

Firstly, we rotate the 3D point cloud incrementally along one or more of the x, y and z axes to generate multiple images instead of taking only a single 2.5D image. A 2.5D image is a mapping of the 3D point cloud to a grey scale image in which the intensity is inversely proportional to the depth [5].

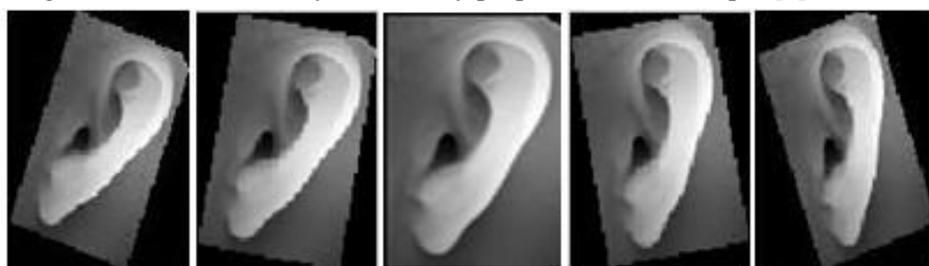


Fig. 1. Examples of different 2.5D ear images generated by rotating the same 3D point cloud about the z axis by -20° , -10° , 0° , $+10^\circ$, $+20^\circ$ respectively.

Figure 1 contains some 2.5D images generated by rotating the same 3D point cloud. Our experimental results demonstrate this provides important additional features and significant ear recognition accuracy.

The second main contribution is our new ear recognition algorithm. In order to recognize an individual's ear, we firstly project the ear's point cloud onto one or more 2.5D images.

⁺ Corresponding author. Tel.: +86-15210161216.
E-mail address: jamesdongxin@gmail.com.

Then we use SIFT descriptors to detect and extract keypoints from human ears. Then we match every test image to the set of training images and calculate the sum of the number of matches and the inverse distance between the best matching keypoints. The closest matching labeled image for every unlabeled test image is then determined.

2. Related Work

Moreno et al. [1] experiment with three neural net approaches to recognition from 2D intensity images of the ear. Their testing uses a gallery of 28 people plus another 20 people not in the gallery. They find a recognition rate of 93 percent for the best of the three approaches. They consider three methods (Borda, Bayesian, and weighted Bayesian combination) of combining results of the different approaches but do not find improved performance over the best individual method.

Choras [6], [7] introduces an ear recognition method based on geometric feature extraction from 2D images of the ear. The geometric features are computed from the edge detected intensity image. They claim that error-free recognition is obtained on “easy” images from their database. The “easy” images are images of high quality with no earring and hair covering and without illumination changes. No detailed experimental setup is reported.

Bhanu and Chen [8] presented a 3D ear recognition method using a local surface shape descriptor. Twenty range images from 10 individuals are used in the experiments and a 100 percent recognition rate is reported. In [9], Chen and Bhanu used a two-step ICP algorithm on a data set of 30 subjects with 3D ear images. They reported that this method yielded two incorrect matches out of 30 people. In these two works, the ears are manually extracted from profile images. They also presented an ear detection method in [10]. In the offline step, they built an ear model template from each of 20 subjects using the average histogram of the shape index [11]. In the online step, first, they used step edge detection and thresholding to find the sharp edge around the ear boundary and then applied dilation on the edge image and connected component labeling to search for ear region candidates. Each potential ear region is a rectangular box, and it grows in four directions to find the minimum distance to the model template. The region with minimum distance to the model template is the ear region. They get 91.5 percent correct detection with a 2.5 percent false alarm rate. No recognition results are reported based on this detection method.

Hurley et al. [12] developed a novel feature extraction technique using force field transformation. Each image is represented by a compact characteristic vector which is invariant to initialization, scale, rotation, and noise. The experiment displays the robustness of the technique to extract the 2D ear. Their extended research applies the force field technique to ear biometrics [13]. In the experiments, they used 252 images from 63 subjects with four images per person collected during four sessions over a five month period; any subject is excluded if the ear is covered by hair. A classification rate of 99.2 percent is claimed on this 63-person data set. The data set comes from the XM2VTS face image database [14].

Yan and Bowyer [15] presented a complete system for ear biometrics, including automated segmentation of the ear in a profile view image and 3D shape matching for recognition. We evaluated this system with the largest experimental study to date in ear biometrics, achieving a rank-one recognition rate of 97.8 percent for an identification scenario and an equal error rate of 1.2 percent for a verification scenario on a database of 415 subjects and 1,386 total probes.

Bay et al. [16] present a fast scale and rotation-invariant interest point detector and descriptor, namely SURF. In their research, they implement a kind of approximation of Hessian-matrix to detect interest points, and utilize a series of box filters which are approximate second order Gaussian derivatives to build scale spaces. Chandrasekhar et al. [17] propose a sort of new framework for computing low bit-rate feature descriptors. They test a various coding methods and distance measures, and compare their CHoG descriptor with other low-bit-rate descriptors. Their experiment results show that CHoG outperforms other descriptor schemes at the same bit rate.

The rest of the paper is organized as follows: In section 3, we describe the details of our ear recognition algorithm. Section 4 presents the experimental methods and the experimental results. Section 5 provides the conclusions.

3. Human Ear Recognition Algorithm

3.1. 3D Point Cloud to 2.5D image Conversion

First, the ear is automatically extracted from the side face range images. Then we have to convert 3D point clouds into 2.5D images because SIFT keypoints can only be extracted from 2D images [5]. Then keypoints could be extracted after the conversion.

We execute Algorithm 1 multiple times for each point cloud as we rotate it incrementally around its x, y, and z axes. A new 2.5D image is generated with each step of rotation. By the end of this process, we will have multiple 2.5D images.

Algorithm 1 3D point to 2.5D image Conversion

Input: A 3D point cloud

1: Compute the extrema of the point cloud along each of the three axes, obtaining $Xmin$, $Xmax$, $Ymin$, $Ymax$, $Zmin$, $Zmax$;

2: Create a 2D image of width $\frac{Xmax - Xmin}{2}$ and height $\frac{Ymax - Ymin}{2}$;

3: Scale the z-value of the points in the cloud to the range 1...255;

4: Project points onto the 2D image pixels, setting each pixel to the scaled z value. Pixels that do not have any 3D points projected on to them are set to zero.

Output: A 2.5D image

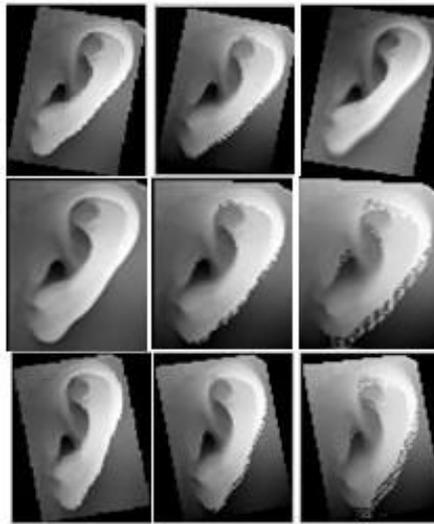


Fig. 2. Examples of different 2.5D images generated by rotating the same 3D point cloud about the x (the rows) and z (the columns) axes in increments of 10°

Then we can detect and extract keypoints from this set of 2.5D images instead of one single image. In this way, we can extract more quality keypoints and keypoints can capture the same feature at different viewing angles.

3.2. Keypoint Detection and Image Matching

Once 2.5D images are generated, we can detect and extract keypoints from both labeled training ears and unlabeled testing ears. In order to classify the test ears, we use a novel image matching algorithm depicted in Algorithm 2.

Algorithm 2 The image matching algorithm.

Inputs: (i) A set of labeled training images; (ii) A single unlabeled testing image.

1: Match the testing image to each training image. Find their SIFT features and the closest matching keypoints.

2: Compute the distance between the best match, $dist$ and the number of matches found, num . For efficiency in Matlab, it is cheaper to compute dot products between unit vectors rather than Euclidean distances. Note that the ratio of angles (acos of dot products of unit vectors) is a close approximation to the ratio of Euclidean distances for small angles.

3: Set the weight of each training image to $(\frac{1}{dist} + num)$.

4: The classification of the testing image is the class of the training image with the greatest weight.

Output: A classification of the unlabeled testing ear.

The function of Algorithm 2 is: a test image is matched to every training image respectively and the closest matching keypoints of every two images can be detected. Then the distance between the closest matching keypoints can be computed and we can give each training image a weight, which is defined as the sum of the inverse distance between the best matching keypoints and the number of matches found. Figure 3 is the matching result of ears of two different individuals.



Fig. 2. The matching result of ears of two different individuals

4. Evaluation

Our 3D ear recognition algorithm was evaluated in two different experiments. We performed our experiments on the University of Notre Dame public dataset (UND dataset) [4]. The main question is whether our method of rotating point clouds to project multiple 2.5D images really works, and if so, which axes of rotation produced the best performance. We describe first of all the dataset used for the experiments, and then experiments and results.

4.1. Dataset

The experiments are performed on the University of Notre Dame public dataset (UND dataset). The UND dataset is the biggest available public dataset for ear recognition. The data are acquired with a Minolta Vivid 910 camera. The camera outputs a 480×640 range image and its registered color image of the same size. The UND dataset consists of scans from 415 different individuals, with 2 different images of each individual, giving a total of 830 images [3].

4.2. Experiment 1

In this experiment, we used one 3D ear per person for training, and the other for testing. First of all, we used our algorithm without any rotation at all. We generate only one 2.5D image out of every training and testing ear. Then we use our algorithm to match them. This represents the baseline case as given in the second row of the table. Experimental results are given in the remaining rows of Table 1.

4.3. Experiment 2

In the second experiment, we tried many different types of point cloud rotation. In each column of Table 1, we rotate the training images in increments of 10° starting at -30° and ending at $+30^\circ$. The axes of rotation were either x, y or z (generating five 2.5D images per scan) or a pair of axes (e.g. x and z) or all the three axes.

The results show that rotating the training ears about the x, y, and z axes by $\pm 30^\circ$ yields the most accurate recognition rates, which reaches 96.49%.

Finally, we tested the idea of rotating both the test ears and the training ears to obtain more test keypoints. Then we perform the same experiments as in the case of Experiment 1. The results of this method are given in Table 1.

As is shown in the tables, rotating both the training and test ears about the x, y, and z axes yields the most accurate rate, which reaches 98.87% -- quite an increase over the baseline of 89.16%. Our method of rotating point clouds really works.

Overall, the experimental results show significant increases in recognition rate. The recognition rate improves as the angle of rotation increases and rotating both the training and test ears yields better results than rotating only the training images.

Table 1. Experiment 1&2 Results

Axes Angles	Training Images Rotated							Training and Test Images Rotated						
	x	y	z	xy	xz	yz	xyz	x	y	z	xy	xz	yz	xyz
Baseline	89.16	89.16	89.16	89.16	89.16	89.16	89.16	89.16	89.16	89.16	89.16	89.16	89.16	89.16
$\pm 10^\circ$	91.62	90.20	92.84	91.64	93.73	93.37	94.66	93.65	92.09	93.87	95.47	96.14	95.33	96.68
$\pm 20^\circ$	92.78	92.01	93.00	93.02	94.38	93.38	95.78	95.61	94.85	95.08	96.43	97.88	96.11	97.59
$\pm 30^\circ$	94.01	92.23	94.54	94.88	95.12	94.89	96.49	97.49	96.38	97.23	97.76	98.25	97.31	98.97

5. Conclusion

We have presented a human recognition system using 3D ear biometrics, which includes SIFT keypoint detection and 3D ear identification. We proposed a novel algorithm for 3D ear recognition. First, rotating the 3D point cloud to generate multiple images enabled us to obtain more quality keypoints. Second, we computed dot products between unit vectors rather than Euclidean distances for efficiency in Matlab to match images, which is an improvement over previous work.

Extensive experiments are performed on the UND dataset and our experimental results show that recognition rate increases significantly by rotating images. The experimental results on color image datasets also demonstrated the potential of the proposed algorithms for robust ear recognition in 3D. Ear biometrics has the potential to be used in the real-world applications to identify humans by their ears. It can be used in both the low and high security applications and in combination with other biometrics such as face. With the decreasing cost (few thousand dollars) and size of a 3D scanner and the increased performance, we believe that 3D ear biometrics will be highly useful in many real-world applications in the future [3].

There are several directions for future work [15]. Further study should result in guidelines that provide best practices for the use of 3D images for biometric identification in production systems. Also, speed and recognition accuracy are still very important issues. We have done some work to improve the speed of the algorithm, but the algorithm might benefit from adding feature classifiers.

6. References

- [1] B. Moreno, A. Sanchez, and J. Velez, "On the Use of Outer Ear Images for Personal Identification in Security Applications," Proc. IEEE Int'l Carnahan Conf. Security Technology, pp. 469-476, 1999.
- [2] A. Iannarelli, Ear Identification. Paramount Publishing, 1989.
- [3] Chen Hui, Bhanu B. Human ear recognition in 3D[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2007, 29(4): 718–737
- [4] The Computer Vision Research Laboratory at the University of Notre Dame. University of Notre Dame Biometrics Database Distribution [DB/OL].(2006-08-11)[2007-07-28]. <http://www.nd.edu/~cvrl/UNDBiometricsDatabase.html>
- [5] M Mayo, E Zhang, "3D face recognition using multiview keypoint matching", Advanced Video and Signal Based Surveillance, 2009. AVSS '09. Sixth IEEE International Conference on.
- [6] M. Choras, "Ear Biometrics Based on Geometrical Feature Extraction," Electronic Letters on Computer Vision and Image Analysis, vol. 5, pp. 84-95, 2005.
- [7] M. Choras, "Further Developments in Geometrical Algorithms for Ear Biometrics," Proc. Fourth Int'l Conf. Articulated Motion and Deformable Objects, pp. 58-67, 2006.
- [8] B. Bhanu and H. Chen, "Human Ear Recognition in 3D," Proc. Workshop Multimodal User Authentication, pp. 91-98, 2003.
- [9] H. Chen and B. Bhanu, "Contour Matching for 3D Ear Recognition," Proc. Seventh IEEE Workshop Application of Computer Vision, pp. 123-128, 2005.
- [10] H. Chen and B. Bhanu, "Human Ear Detection from Side Face Range Images," Proc. Int'l Conf. Image Processing, pp. 574-577, 2004.
- [11] J. Koenderink and A. van Doorn, "Surface Shape and Curvature Scales," Image and Vision Computing,

vol. 10, pp. 557-565, 1992.

- [12] D. Hurley, M. Nixon, and J. Carter, "Force Field Energy Functionals for Image Feature Extraction," *Image and Vision Computing J.*, vol. 20, pp. 429-432, 2002.
- [13] D. Hurley, M. Nixon, and J. Carter, "Force Field Energy Functionals for Ear Biometrics," *Computer Vision and Image Understanding*, vol. 98, pp. 491-512, 2005.
- [14] K. Messer, J. Matas, J. Kittler, J. Luetin, and G. Maitre, "XM2VTSDDB: The Extended M2VTS Database," *Audio and Video-Based Biometric Person Authentication*, pp. 72-77, 1999.
- [15] Yan Ping, Bowyer K W. Biometric recognition using three dimensional earshape[J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2007, 29(8): 1297—1308
- [16] Herbert. Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110 (3), June 2008.
- [17] V. Chandrasekhar, G. Takacs, D. Chen, S. Tsai, R. Grzeszczuk, B. Girod. CHoG: Compressed Histogram of Gradients: A Low Bit-Rate Feature Descriptor. *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.