Audio Engineering Society

Convention Paper 10348

Presented at the 148th Convention
2020 June 2-5 , Online

Study on comparison of individuality of ear canal shape

Riki Kimura[1], Shohei Yano[1], Rui Fujitsuka[1], Naoki Wakui[2], Takayuki Arakawa[3], and Takafumi Koshinaka[3]

[1]*Yano lab. NIT, Nagaoka College*
[2]*Wakui lab. NIT, Nagaoka College*
[3]*NEC Corporation*

Correspondence should be addressed to Shohei Yano (syano@nagaoka-ct.ac.jp)

ABSTRACT

Ear acoustic authentication, a type of biometric authentication, uses the acoustic characteristics of the ear canal as a feature. Because ear acoustic authentication acquires features using earphones, the process of authentication is easy, and the method has attracted much attention recently. However, the mechanism of the acoustic characteristics of the ear canal has not been sufficiently studied. In this study, we verified two methods, the image matching method and Slicing method. In conclusion, Slicing method was found to outperform the image matching method, based on the results of this study.

1 Introduction

In today's digital society, countless devices such as computers, smartphones, cars, and home appliances around the world are connected by the Internet, and various types of information are exchanged. By using these, we can interact with social networking service(SNS) and shop online. However, there is an increasing number of cases of abusing digital technologies, such as crimes due to spoofing and leakage of personal information. Because these can cause significant damage to victims' rights and property, personal identification technology that prevents impersonation is important. Traditionally, password authentication, PINs, and card keys were used for personal authentication. Password authentication and PINs can be stolen when keyed in. Card keys can also be physically stolen. Recently, biometric authentication has been widely used. Biometric authentication has the advantage that it is less likely to be stolen because it uses unique information such as personal biometric information for authentication. Familiar biometric authentication techniques such as fingerprint authentication and face authentication are used in smartphones. Biometric authentication is used to not only unlock the device, but to also verify the identity of the user, such as in online payments, and must be re-authenticated each time. With conventional biometric authentication, authentication operations such as pressing a finger or facing a sensor are required. Repeating these increase the burden on the user. In addition, the safety of unlocking the device is not guaranteed. This is because it is difficult to detect if the person is indeed the one who operates after the unlocking changes. This makes impersonation easy if you lose your device while it is unlocked. In other words, there seems to be a need

for a system that does not require an authentication operation and can detect spoofing even after the service starts. We have developed ear acoustic authentication as a biometric authentication technique that overcomes these shortcomings. The ear acoustic authentication measures the acoustic characteristics of the external auditory canal using an earphone with a built-in microphone as shown in Fig. 1, and performs personal authentication based on individuality.

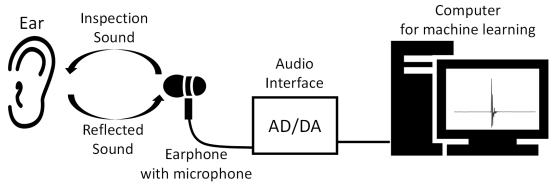

Fig. 1: The ear acoustic authentication system

Because continuous authentication is possible only by wearing the earphone, spoofing can be prevented in a hands-free manner. However, the mechanism by which individuality occurs in acoustic characteristics has not been elucidated. Shapes, skin types, lengths, and the likes can be cited as elements that constitute the external auditory canal. In this study, we focused on the shape. Therefore, the external ear shape was measured using a three-dimensional external ear shape scanner, and the relationship between the shape of the external auditory canal and acoustic characteristics was investigated from the viewpoint of similarity. However, no method for quantitatively evaluating the similarity of the three-dimensional ear canal shape has been studied. In simple polygon matching, a corresponding polygon cannot be searched for, and it cannot be said that similarity can be accurately evaluated. In this study, we proposed two new methods and evaluated their performance.

2 Measurement and editing of the ear canal shape

The outer ear shape was measured using a 3D outer ear shape scanner (United Science e-Fit scanner). Fig. 2 shows the measurement system reproduced with a dummy head.

Fig. 2: The measurement system

As shown in Fig. 3, The 3D ear canal data obtained by the measurement was cut using 3DCAD software, and only the ear canal part was extracted.

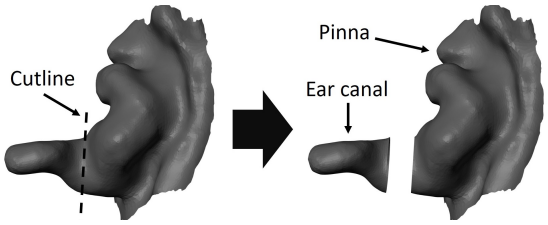

Fig. 3: The ear canal extraction processing

We proposed standardization to reduce human error in cut editing. This was to easily standardize by attaching a small marker to the ear as shown in Fig. 4 and measuring.

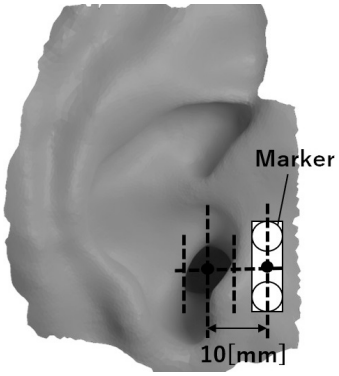

Fig. 4: The placement of marker on ear

Fig. 5 shows an ABS resin marker created by a 3D printer and the output image when scanning it.

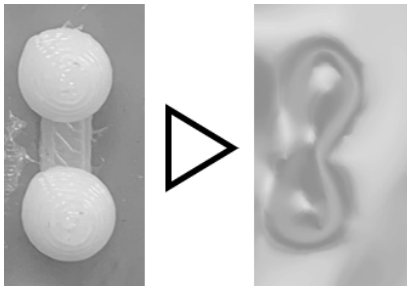

Fig. 5: The created marker and scan image

3 Analytical methods

3.1 Image matching method

In the past, a method for evaluating similarity by rendering 3D data to create a two-dimensional image and extracting local features using SIFT was studied[1][2]. In this method, two-dimensional images were created by rendering the ear canal shape data. The ear canal was scanned in 1° increments over a range of ±5° to reduce the angular error from the measurement, producing a total of 11 images. We extracted local features of the created image using A-KAZE, which is an improvement of KAZE[3][4]. The extracted features are matched by the brute force method, and one feature point is searched for all feature points of other data. In addition, a cross check was performed to improve the matching accuracy. This means that only the result of the best matching between the two feature point groups was left. In A-KAZE, the Hamming distance was used for evaluation because the descriptor was a binary representation. The Hamming distance between all the feature points matched between one set of images was calculated, and the average value was defined as the Hamming distance between the images. Hamming distance D when N sets of features A_1 to A_N and B_1 B_N were matched for one set of images A and B was represented by the following Eq. 1.

$$D = \frac{1}{N}\sum_{i=1}^{N}|A_i - B_i| \tag{1}$$

All calculations were performed on ±5° images and the image with the smallest D was defined as the distance between the final A and B images.

3.2 Slicing method

The ear canal shape data were sliced at equal intervals with respect to the Z-axis, and divided into many layers. By slicing at narrow intervals, the thickness of each layer in the Z-axis direction could be regarded as 0, and could be considered as data on the XY plane. When the center coordinates of each polygon were plotted with respect to the cut of the layer, an elliptical distribution, which was the shape of the ear canal, was shown. Least squares were applied to these point groups, fitting performed, ellipses fitted, and their constituent elements (center coordinates, short sides, long sides, and angles) calculated. Least square fitting was performed based on the method proposed by Pilu et al[5]. In this study, the Euclidean distance of the center coordinates was used to evaluate the similarity. We assumed that there were two ear canal shapes, data A and B, each of which was sliced into N equal layers, and the center coordinates of each layer were defined as A_1 to A_N and B_1 B_N from the bottom. First, coordinate reference matching was applied. The center coordinates $A_1 = (X_{A1}, Y_{A1})$ and $B_1 = (X_{B1}, Y_{B1})$ of the bottom layer were considered as deviations from the origin (X,Y)=(0,0). By correcting this shift, the center coordinates A_i and B_i of an arbitrary layer were represented by Eq. 2 and 3.

$$A_i = (X_{Ai} - X_{A1}, Y_{Ai} - Y_{A1}) \tag{2}$$

$$B_i = (X_{Bi} - X_{B1}, Y_{Bi} - Y_{B1}) \tag{3}$$

Therefore, the coordinates were set at the origin with reference to the center coordinates of the bottom layer. Next, an angle scan was applied to correct the angle shift at the time of measurement. The coordinates A_i' obtained by rotating the center coordinates $A_i = (X, Y)$ of an arbitrary layer by θ around the origin were represented by the following Eq. 4.

$$A_i' = (X\cos\theta - Y\sin\theta, X\sin\theta + Y\cos\theta) \tag{4}$$

The average of the Euclidean distances between all layers was defined as the Euclidean distance D of the ear canal shape. D when $A_i' = (X_{Ai}', Y_{Ai}')$ and $B_i = (X_{Bi}, Y_{Bi})$ was expressed by Eq. 5.

$$D = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(X_{Ai}' - X_{Bi})^2 + (Y_{Ai}' - Y_{Bi})^2} \tag{5}$$

It was determined that the smaller D was, the higher the similarity was. In Eq. 4, θ was scanned by 1° within a range of 360°, and the one with the smallest D was defined as the Euclidean distance between the ear canals.

4 Performance evaluation method

As a performance evaluation index of the two methods, the degree of separation S represented by the ratio between the intra-person distance(D_i) and others distance(D_o) was introduced. S was represented by Eq. 6.

$$S = \frac{D_o}{D_i} \tag{6}$$

D_o was excellent when large and D_i was excellent when small. In other words, the evaluation index was such that the larger S was, the better the method.

5 Experiment

We measured the external ear shape data of three people three times each. Only the ear canal part was extracted from the data and remeshed so that the number of polygons became 3200. The data created in this way was used for experiments.

5.1 Image matching method

Rendering the ear canal shape data using 3DCAD software (Blender). Rendering was performed at 1000 $\times$ 1000, 700 $\times$ 700, and 400 $\times$ 400 resolution using light source, no light source, and edge conditions respectively. The rendered images were matched, and the feature point distance D between the images was calculated from the average of the Hamming distance between the feature amounts. Finally, calculated the degree of separation S and use it for evaluation.

5.2 Slicing method

The ear canal shape data was sliced in the Z-axis direction at four intervals of 3 mm, 1 mm, 0.5 mm, and 0.1 mm. The Euclidean distance D was calculated at each slice interval. Finally, calculated the degree of separation S and use it for evaluation.

6 Results

6.1 Image matching method

Figs. 6 to 8 show examples of the rendered images at each setting.

Fig. 6: Rendered image with light source

Fig. 7: Rendered image with no light source

Fig. 8: Rendered image with edge processing

Fig. 9 shows the degree of separation S calculated for each setting.

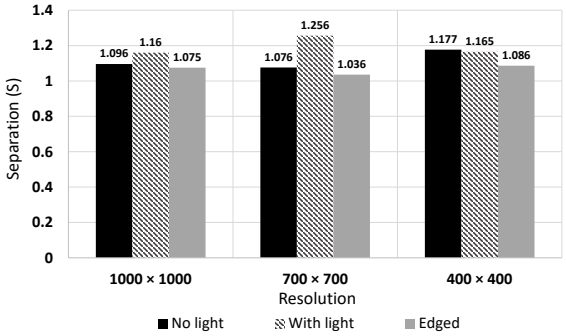

Fig. 9: Degree of separation S at each setting

From Fig. 9, it can be seen that the degree of separation S exceeds 1 in all the settings. Therefore, we conclude that similarity assessment using rendered images is possible. Also, the highest degree of separation was obtained when there was no light source with a resolution of 700 × 700. However, we think that the proper resolution and image type need further validation.

6.2 Slicing method

Fig. 10 shows the definition of the axes in this experiment and the cut line for a 1 mm slice interval.

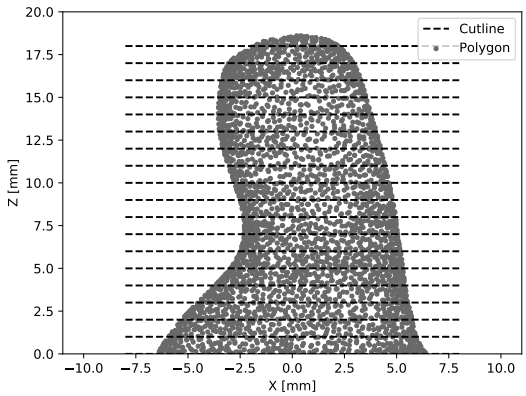

Fig. 10: Definition of each axis and cut line at 1mm intervals

Fig. 11 shows the bottom-level approximation at 1mm slice interval.

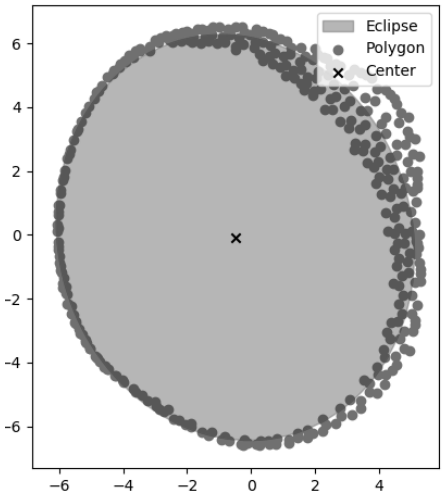

Fig. 11: The approximate ellipse at 1mm slice

Fig. 12 shows the degree of separation S calculated for each slice interval.

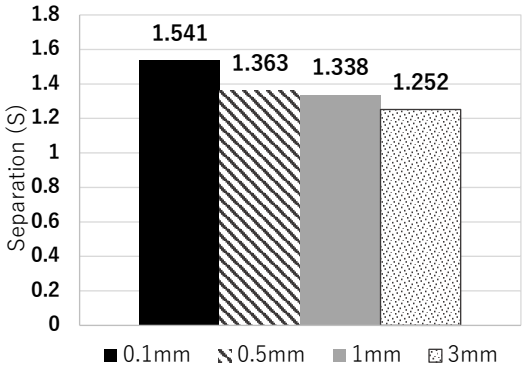

Fig. 12: Degree of separation S at each slice interval

From Fig. 12, it can be seen that the degree of separation S exceeds 1 at all slice intervals. Therefore, we conclude that similarity assessment by slice plane approximation is possible. Also, it can be seen that the smaller the slice interval, the greater the separation S.

7 Summary

First, we proposed a standardization method for measuring the ear canal shape. Next, to evaluate the similarity of the ear canal shape, we proposed the following two methods: an image matching method and a slice method. Thereafter, we evaluated their performance. For the evaluation, the degree of separation S, which is the ratio between the distance between individuals and the distance between other people, was used. In the image matching method, the degree of separation of 1.036 to 1.256 was obtained. In the slicing method, the degree of separation of 1.252 to 1.541 was obtained. From this result, it was concluded that the slice method showed a higher degree of separation and was an excellent method.

References

[1] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision (IJCV), 60(2), pp. 91-110, (2004).

[2] R. Ohbuchi, K. Osada, T. Furuya, T. Banno, “Salient local visual features for shape-based 3D model retrieval”, IEEE Shape Modeling International (SMI) 2008, pp. 93-102, (2008).

[3] Alcantarilla, Pablo Fernández, Adrien Bartoli, and Andrew J. Davison. "KAZE features." European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2012.

[4] Alcantarilla, Pablo F., and T. Solutions. "Fast explicit diffusion for accelerated features in nonlinear scale spaces. " IEEE Trans. Patt. Anal. Mach. Intell 34.7 (2011): 1281-1298.

[5] Pilu and Fischer in Fitzgibbon, A.W., Pilu, M., and Fischer R.B., Direct least squares fitting of ellipses, Proc. of the 13th International Conference on Pattern Recognition, pp 253–257, Vienna, 1996.