Gender and Kinship by Model-Based Ear Biometrics

Di Meng
Department of Electronics and Computer Science
University of Southampton
Southampton, United Kingdom
dm4n17@soton.ac.uk

Mark S. Nixon
Department of Electronics and Computer Science
University of Southampton
Southampton, United Kingdom
msn@ecs.soton.ac.uk

Sasan Mahmoodi
Department of Electronics and Computer Science
University of Southampton
Southampton, United Kingdom
sm3@ecs.soton.ac.uk

***Abstract*— Many studies in biometrics have shown how identity can be determined, including by images of ears. In the paper, we show how model an ear and how the gender appears to often be manifest in the ear structures, as is kinship or family relationship. We describe a new model-based approach for viewpoint correction and ear description to enable this analysis. We show that with the new technique having satisfactory basic recognition capability (recognizing individuals with performance similar to state of art), gender can achieve 67.2% and kinship 40.4% rank 1 recognition on ears from subjects with unconstrained pose.**

Keywords— model-based, geometric features, gender classification, kinship verification

I. Introduction

It has long been known that ears are unique to their owner [1]. As in other biometrics, there are wider aspects to identification, and more recently it has been shown that gender can be determined automatically from ears [8] [11] [15]. This paper is about these wider aspects, in particular about gender and kinship. We use model-based identification, as this is the only possible approach to achieve the focuses of this research.

Everyone has a pair of ears, and ears will not change much between 8 and 70 years old [16]. Recognition approaches [6] mainly use appearance-based features and there has been little work to date on model-based approaches (perhaps since a model is not intuitive in this rather unappealing biometric). The recent Unconstrained Ear Recognition Challenge [5] has highlighted the difficulties of performing person identification from ear images in the wild. For complete information of identity from ear images utilizing soft biometric traits [12], such as gender, can be supplementary. However, there is as yet limited work on gender classification from ears, though it is common in many other biometrics. This is perhaps because, like a model, it does not appeal to intuition that gender can be ascertained from the loose and apparently random structure of the inner ear.

The first paper to show that gender can be perceived from the ear was actually based on 2D [8] and using Gabor filters achieved 89.5% correct recognition. Later work improved recognition to 92.4% based on 3D imagery [11]. A more recent study used geometric and hand labelled features [15] to again demonstrate that gender is implicit in ears.

Determining kinship from face images has a quite recent history. This is motivated by the common observation that siblings resemble their parents. Technology and age have around 96.2% has been achieved [9]. However, there are the traditional difficulties in face recognition: expression, pose, occlusion and age. By way of contrast, ears do not suffer from change in expression and appear largely invariant with age (except to enlarge) though pose is innate, and occlusion by hair poses a natural impediment to recognition.

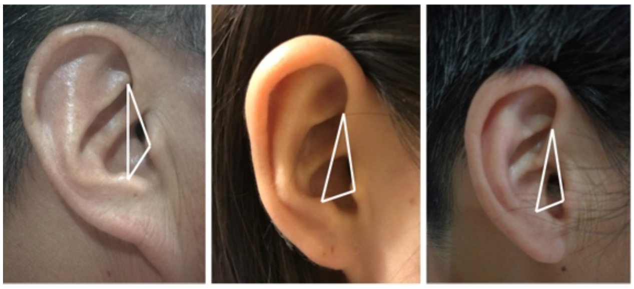


Fig. 1. On Kinship and Gender from Ear Images

This study is the first work on automated kinship verification from ear images and gender, both by a model basis. Fig.1 shows the ears of two parents together with one from their daughter, and in the other a son. By overall structure, the daughter's ear appears to resemble that of her mother, and the son's ear resembles his father's. As we shall later describe, the triangles are prescribed by the locations of interest points and reflect the similarity. Naturally, these are just one of the geometric features, and there are many more. For the mother and daughter, the triangles formed by the fossa, incisura and antitragus appears to be more similar than that of the father; in contrast the triangle of the son appears close to that of his father. The difference in gender is less intuitive, though on these images the helix (the outer rim) appears wider for the female subjects.

II. Model-based Analysis of Ears

It is possible to determine gender and kinship by appearance-based features. However, this would only suggest performance capability and would not identify salient structures for future investigation. Adding capability to handle change in pose with appearance- based approaches mandates a very large database of images. We therefore decided to use a model-based approach. Essentially, we model the ear as a flat plane attached to the side of the head. This means that the affine transform can be used to correct for changes in pose/ viewpoint (perspective correction is unnecessary given the scale of the ear relative to its context). We detect interest points in guided manner, thereby identifying salient structures within the ear itself.

A prime description is to use triangles given their strong invariant properties. Our model consists of 9 points, shown in Fig. 2(a), which form 84 triangles, from which geometric features can be extracted. Therefore, 36 different distances can be computed among these 84 triangles. The points for the geometric features are shown in Fig. 2(a), and these have reliably been detected on ear datasets. The red triangle shows one of the 84 triangles and the basis used for the affine viewpoint correction. Fig. 2(b) shows six triangles of the 84 model triangles.

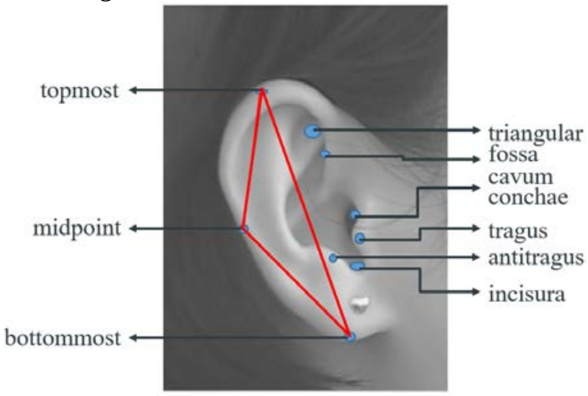


Fig. 2(a). Ear model points

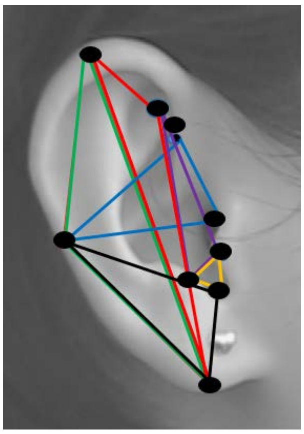


Fig. 2(b). Basis of ear model

A. Geometric Features

The images used in this study were from one database where subjects were looking in a plane normal to the camera view so the ear appears flat. We also have another unconstrained database where subjects' ears are not necessarily normal to the camera view. Since the ear can be viewed as a flat structure, the correction of the ears to appear as normal to the camera view can be made using an affine transform. For this paper, three points were used for the affine transformation, the top-most, bottom-most, and middle point.

In the first stage we use the Hough transform (HT) for ellipses to detect the ear, given good performance in noise and occlusion [2]. The result of applying the Hough transform for ellipses to an ear image is shown in Fig. 3.

The mapping of the ellipse defined as (1)

$$\frac{(x\cos\alpha + y\sin\alpha)^2}{a^2} + \frac{(x\sin\alpha - y\cos\alpha)^2}{b^2} = 1 \qquad (1)$$

where α is the rotation of points (x, y) with axes a, b.


Fig. 3. Hough transform for ear detection

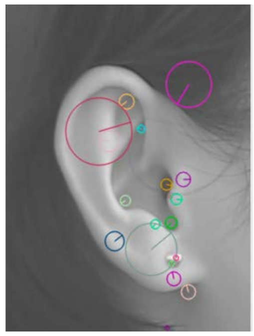


Fig. 4. Ear keypoints by SIFT

In this paper, the ellipse has been detected via edges detected by the Canny operator thus labelling the apices of the ellipse (the top-most and bottom-most points). These are the first two ear points to be detected and define a major axis from top to bottom. The third ear point is defined as the rear-most edge point along a line normal to the central major axis at its centre. As shown in Fig. 2(a)(b), this defines a point on the rear helix of the ear. These points were found to be robust except in ear images with severe occlusion by hair, when the ear can hardly be detected anyway.

The affine transform is defined as (2)

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_5 \\ a_6 \end{bmatrix} \qquad (2)$$

where x, y is original coordinate, x', y' is coordinate via affine transform, $a_1, a_2, a_3, a_4, a_5, a_6$ are the coefficients of shear, scale, rotation.

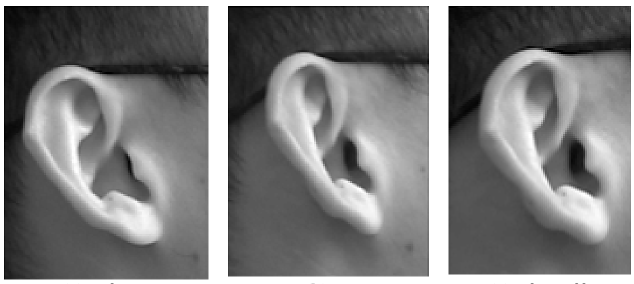

(a) reference (b) new (c) after affine

Fig. 5. Applying the affine transform

As observed from the above equation, the affine transform [14] has six degrees of freedom corresponding to the six parameters of the transformation (scale, rotation and translation parameters) can be calculated from non-collinear correspondences. It appears that a person's head may change only in yaw and pitch in the databases here, and there were no examples of roll. When capturing the ear from the experimental subject, that means the databases have some ear images which have yaw axis rotation and pitch axis rotation. Fig. 5 shows an example of the affine transformation. Fig.5 (a) is reference image, Fig. 5(b) is the registered image, which rotated around the yaw axis. Fig. 5(c) is the image after affine transformation.

SIFT (Scale-invariant feature transform) [10] is used to detect interest points and their description. SIFT is known to be a robust way for landmark extraction even in images with small pose variations and varying brightness conditions. Fig. 4 shows the result of SIFT, similar to a previous model-based approach [3].

B. Appearance features

The Histogram Of Gradients (HOG) [4] is employed here to describe the ear area surrounded by points detected in the previous section. There are five steps in the computing HOG descriptors. Fig. 6 shows the HOG of an ear image.

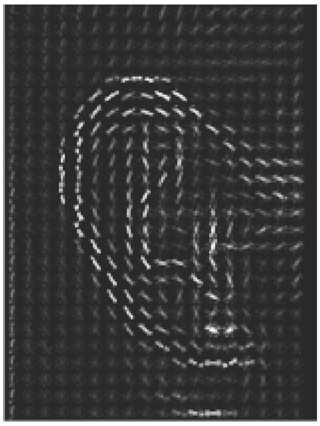


Fig. 6. HOG descriptions of an ear image

III. RESULT AND ANALYSIS

In this section, the classifier used for recognition is kNN (k nearest neighborhood algorithm), and uses the Euclidean, Manhattan, and Mahalanobis [12] distance measures. In addition, we use SFFS (Sequential Forward Floating Selection) [7] to do the feature selection.

We show initially how it is possible to achieve identity recognition by the new measures, on a standard database captured with the ear normal to the plane of view of the camera. We then show that the recognition on a new database captured for gender and kinship analysis, wherein the axis of view is not normal (the pose of the subjects was suggested rather than controlled), and the affine transformation is deployed to correct for viewpoint. We show gender on our new database and then kinship. Our database is sufficiently large for an initial study, which shows that indeed gender can be determined, and that kinship can be ascertained.

Fig.7. Ears from USTB databases 1 and 2

A. Ear recognition

The database1 and database2 of USTB database were used to investigate recognition capability. Database 1 has 180 images (60 subjects with 3 images for each subject), and database 2 has 308 images (77 subjects, 4 images for each subject). In addition, images in database 1 are not rotated whereas ears in database 2 are rotated. Fig. 7 shows the samples of database 1 and 2.

TABLE I. RANK 1 RECOGNITION ON DATABASE 1

Method	*Euclidean distance*	*Manhattan distance*	*Mahalanobis distance*
Geometric feature	81.7%	82.0%	76.3%
HOG	91.7%	90.0%	65.5%
Geometric features + HOG	92.8%	95.6%	76.3%

TABLE II. RANK 1 RECOGNITION ON DATABASE 2(WITHOUT AFFINE TRANSFORM)

Method	*Euclidean distance*	*Manhattan distance*	*Mahalanobis distance*
Geometric feature	53.5%	54.2%	37.2%
HOG	62.1%	66.3%	38.7%
Geometric features + HOG	62.5%	70.2%	37.4%

TABLE III. RANK 1 RECOGNITION ON DATABASE 2(VIA AFFINE TRANSFORM)

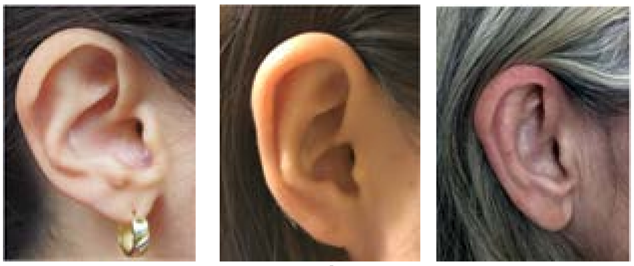
Method	*Euclidean distance*	*Manhattan distance*	*Mahalanobis distance*
Geometric feature	56.1%	59.8%	43.0%
HOG	59.9%	70.4%	30.0%
Geometric features + HOG	64.2%	80.0%	30.8%

TABLE I shows the result on database 1 which demonstrates that our approach can indeed match contemporary performance on constrained datasets. TABLE II shows the result of database 2 on raw images, and recognition is much reduced without use of the affine transform. TABLE III shows the result of database 2 with viewpoint correction via the affine transformation. In these three tables, the first column is the description method and first row is the distance measure used. All results use kNN (k=1) with leave one out cross validation. As TABLE I shows, the accuracy of HOG is a little higher than for the geometric features. When the two feature sets are fused the accuracy as high as 95.6%. For TABLE II and TABLE III, if the rotated image uses affine transformation, the accuracy is better than the image without affine transformation. The accuracies of geometric features are 59.8% and 54.2%, and the accuracies of HOG are 70.4% and 66.3%. The accuracies of geometric features and HOG are 80.0% and 70.2%.

B. Gender and kinship recognition

For gender classification, another database was constructed at the University of Southampton, which contains 134 images (67 subjects, 2 images for one subject) is used. There are 37 males and 30 females of the database. Fig. 8 shows the samples for gender classification. In our experiments in this section we use k=3.

The correct gender classification of geometric is 67.2%. It is important to notice that the classification accuracies in TABLE IV are calculated by using only the geometrical features. It is expected that these accuracies significantly improve with the fusion of HOG features with our existing geometrical features.



Female

Male

Fig. 8. Samples for Gender Classification

TABLE IV. RESULTS FOR GENDER CLASSIFICATION

Method	*Geometric feature*
Euclidean distance	67.2%
Manhattan distance	59.7%
Mahalanobis distance	65.7%

The experiments for Kinship verification are performed on our database with k=1, the database contains 134 images. There are 21 families (21 parents) and 25 children (16 sons, 9 daughters) of the database.

TABLE V. RESULT OF KINSHIP RECOGNITION (EUCLIDEAN DISTANCE)

Euclidean distance	*Father*	*Mother*
Son	40.4%	31.3%
Daughter	0%	33.3%
Son & Daughter	20.0%	28.2%

TABLE VI. RESULT OF KINSHIP RECOGNITION (MANHATTAN DISTANCE)

Manhattan distance	*Father*	*Mother*
Son	31.3%	25.0%
Daughter	0%	33.3%
Son & Daughter	28.0%	28.0%

As TABLE V and TABLE VI show the correct of kinship verification of father and son is 40.4%, and that of mother and daughter is 33.5%. However, the correct kinship verification between father and daughter is 0%. As observed from TABLE V and TABLE VI, the classification accuracy of kinship for father and son is 40.4%, and that of mother and daughter is 33.3%. However, the correct kinship verification between father and daughter is 0%. In TABLE V and TABLE VI, the classification accuracies are computed by using only geometrical features. We therefore expect to obtain significantly higher classification accuracies for Kinship once we fuse HOG features with our existing geometrical features.

Fig. 9 shows the Cumulative Match Characteristic (CMC) Curve of the results, where the horizontal axis signifies the correct identity was among the top n (number of rank) results. The basic capability of the technique is shown by the result on the USTB databases1 (d1) which shows 95.6% for the basic model-based approach. The inclusion of the affine transform allows wider application capability, but the recognition reduces to 80.0% (d2). Where the affine transform not to be included, recognition is 70.2% showing that the affine transform is operating with success. At present, all the ear recognition and gender classification results achieve 100% by rank 2 though the identity starts at 80.0% and gender 67.2% at rank 1. However, the kinship for FS (Father and Son) and MD (Mother and Daughter) starts at 40.4% and 33.3%, respectively, and achieves 100% by rank 20. Clearly, these results are well above random and encourage future study.

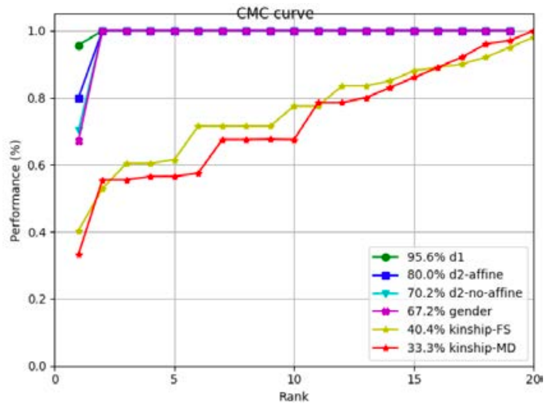


Fig. 9. Cumulative Match Characteristic Curve of Results

IV. CONCLUSIONS AND FUTURE WORK

This paper presents a study on a model-based ear recognition and on gender classification and kinship verification from ear images. In this study, the affine transform was used successfully for viewpoint correction. Our model based method has produced promising results, with an identification accuracy of 95.6% for non-rotated images. The affine transformation operates on the rotated images but reduces performance slightly. In addition, our model-based algorithm is also used for gender classification to produce 67.2% classification accuracy. Although the accuracy is not as high as appearance based results, it is a high performance among model-based techniques and can handle affine transformations. Therefore, we could benefit from more features for our system for gender classification. We can also use HOG features to improve gender classification and kinship verification accuracies which are currently. 40.4% accuracy for father and son, 33.3% accuracy for mother and daughter, 0% for father and daughter, 31.3% for mother and son. It is noted that in the kinship verification result, mothers' ears appear to be more influential. The accuracy of kinship verification indicates that ear biometrics can contribute to kinship verification. We intend to find more features to improve classification accuracy and deeper understanding of how these results are related to basic ear structures. The observation that the helix differs between female and male ear suggests that including the helix in our model-based system could conceivably improve the gender, and thus perhaps kinship, classification accuracy.

ACKNOWLEDGMENTS

I thank my supervisors Dr. Sasan Mahmoodi and Prof. Mark Nixon, for their help. The work was supported by my supervisors and my friends. Thanks for their families' ear images too.

REFERENCES

[1] Abaza, A., Ross, A., Hebert, C., Harrison, M. A. F., Nixon, M. S. "A survey on ear biometrics," ACM computing surveys (CSUR), 45(2), 22, 2013.

[2] Arbab-Zavar, B., Nixon, M.S., "On shape-mediated enrolment in ear biometrics," In: International Symposium on Visual Computing. Springer, Berlin, Heidelberg, 2007.

[3] Arbab-Zavar, B., Nixon, M.S., "On guided model-based analysis for ear biometrics," Computer Vision and Image Understanding, 487-502, 2011.

[4] Dalal, N., Triggs, B., "Histograms of oriented gradients for human detection," In: International Conference on computer vision & Pattern Recognition (CVPR'05). Vol. 1. IEEE Computer Society, 2005.

[5] Emeršič, Ž., et. al., "The unconstrained ear recognition challenge," In: 2017 IEEE International Joint Conference on Biometrics (IJCB). IEEE, 715-724, 2017.

[6] Emeršič, Ž., Štruc, V., Peer, P., "Ear recognition: More than a survey," Neurocomputing, 255, 26-39, 2017.

[7] Fallahpour, S., Eisa. N., Mohammad. H., "Using an ensemble classifier based on sequential floating forward selection for financial distress prediction problem," Journal of Retailing and Consumer Services 159-167, 2017.

[8] Khorsandi, R., Mohamed, A., "Gender classification using 2-D ear images and sparse representation," In: IEEE Workshop on applications of computer vision (WACV). IEEE, 2013.

[9] Kohli, N., Vatsa, M., Singh, R., Noore, A., Majumdar, A., "Hierarchical representation learning for kinship verification," In: IEEE Transactions on Image Processing 26.1, 289-302, 2016.

[10] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," International journal of computer vision, 91-110, 2004.

[11] Lei, J., Zhou, J.D., Mohamed, A., "Gender classification using automatically detected and aligned 3D ear range data," In: 2013 International Conference on Biometrics (ICB). IEEE, 2013.

[12] Nixon, M.S., Alberto S. Aguado, "Feature extraction and image processing for computer vision," Academic Press, 2012.

[13] Nixon, M. S., Correia, P. L., Nasrollahi, K., Moeslund, T. B., Hadid, A., Tistarelli, M., "On soft biometrics," Pattern Recognition Letters, 68: 218-230, 2015

[14] Song, Z., Zhou, S., Guan, J., "A novel image registration algorithm for remote sensing under affine transformation," In: IEEE Transactions on Geoscience and Remote Sensing 52.8, 4895-4912, 2013

[15] Yamen, D., Eyiokur, F.I., Sezgin, N., Ekenel, H.K., "Age and Gender Classification from Ear Images," In arXiv preprint, arXiv:1806.05742v1, 2018.

[16] Iannarelli, A., "Ear Identification," Paramont Publishing, 1989.