

# On Distinctiveness and Symmetry in Ear Biometrics

Di Meng<sup>ID</sup>, Mark S. Nixon<sup>ID</sup>, *Member, IEEE*, and Sasan Mahmoodi, *Member, IEEE*

**Abstract**—Previous works show that human ears can be used for identification, gender classification, and kinship verification and have investigated whether there is a symmetry between a person's ears; however, the symmetry performances have been less than satisfactory. Our paper extends the analysis of gender classification on ear images and analyses bilateral symmetry of human ears, in both cases aiming to determine the ear parts from which recognition is derived. We use model-based analysis and deep learning methods to capitalize on structure and performance, respectively. We consider the rotation of ear images under an affine transformation, by modelling the ear as a flat plane attached to the head. We address the question as to whether it is possible that given an image of one ear, a person can then be recognized from their other ear. Such a symmetry-based strategy could reduce constraints on applications of ear biometrics. We show that it is possible to recognise the gender with a 90.9% success rate and that the ear rim (the upper helix and lobe) dominates performance. To investigate symmetry, we compare one ear with a mirrored version of the other ear and achieve 93.1% CCR, which is the current state-of-the-art, with important regions different from those determined for gender. To extend the analysis we construct two groups of images, one of which contains both ears from the same subject and the other contains two ears from different subjects. The 100% CCR confirms the existence of symmetry between a subject's ears. By these approaches we show that there is actually a high chance that there exists symmetry between a person's ears and that it would be prudent for recognition systems to concentrate on the inner ear rather than the outer ear.

**Index Terms**—Gender classification, ear symmetry, deep learning, model-based, heatmaps.

## I. INTRODUCTION

EARS have long been considered unique to their owners and therefore constitute a reliable biometric. Ear recognition has been based on holistic and model-based approaches, with more recent work using deep learning. There are clear advantages to using an ear for biometric systems since it is immune to expression, though ears can be obscured by hair. Many approaches have been targeted at, and evaluated on, standardized datasets where the ear position is controlled (including early datasets like XM2VTS [29] and later and larger ones like SC face [23]). This is a necessary step in any biometric, since the prime consideration in biometrics is uniqueness to a subject. In order to translate ear biometrics

Manuscript received July 20, 2020; revised October 5, 2020 and January 7, 2021; accepted February 2, 2021. Date of publication February 12, 2021; date of current version March 24, 2021. This article was recommended for publication by Associate Editor Z. Sun upon evaluation of the reviewers' comments. (Corresponding author: Di Meng.)

The authors are with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: dm4n17@soton.ac.uk; msn@soton.ac.uk; sm3@soton.ac.uk).

Digital Object Identifier 10.1109/TBIOM.2021.3058562



Fig. 1. Examples where ear biometrics might be preferred for recognition.

to real world applications, we need to demonstrate and understand the capability on unconstrained ear images, where the ear is not necessarily in a plane orthogonal to the camera's view. It is not yet known in the literature, whether there is bilateral symmetry between ears, and whether any parts might best be focused on for recognition, if at all. This article describes work aimed to investigate these issues, as a part of the development of ear biometrics for applications beyond the laboratory.

The recent Covid-19 outbreak and the increased wearing of masks might hinder the development of using the face for recognition. Two examples for ear identification are demonstrated in Fig. 1. As shown in Fig. 1(a) a rioter might conceal his/her face, but not their ear and a mask obscures much of the face but not the ear. Fig. 1(b) shows different surveillance images recorded at a crime scene with low quality suggesting a need for the capability to handle noisy images. In all of these images, it might be preferable to use the ear for identification, since it is the only biometric that can be seen clearly. To achieve that, we need to reduce constraints on the deployment of surveillance/security systems to be able to handle images where the ear is not presented normal to a camera view and might also be rotated. Given that one ear often only can be seen, it is also prudent to investigate symmetry. Though it will always be best to store two images per subject it is conceivable that a match might be required when only one ear can be seen in different situations.

The external anatomy of the human ear has been studied for many years, and Fig. 2 shows the structure of outer ears. Each ear should have the same external anatomy, whether left ear or right. Naturally there might be covariate factors like injury or jewellery, but the underlying structure is presumed to be

1. (a-d) helix;
2. lobe;
3. antihelix;
4. cavum conchae;
5. tragus;
6. antitragus;
7. helix;
8. triangular fossa; and
9. incisura;



Fig. 2. Anatomy of the Human Ear [1].

the same, by human vision. Also, it is often considered that the left ear is similar to the right ear, but this has yet to be established empirically.

Compared with face biometrics, ears appear to change little with age (8-70 years old) [2]; compared with fingerprints ear images can be collected without a subject's consent and can be captured from a greater distance. Historically the first person recognizing biometric on human ears was the French criminologist Alphonse Bertillon in 1890 [3]. Iannarelli later examined 10,000 ear samples to prove that ears are unique to their owner and developed a manual system for ear identification [2]. Burge *et al.* were the first to introduce an automated ear biometric approach [4], and Hurley *et al.* [5] were the first to demonstrate recognition. The complete information of identity from ear images, utilizing soft biometric traits [6], such as gender, can be supplementary. Ear images have previously been used for identifying humans [5], [11], [24]–[26], classifying gender [7]–[10], and verifying kinship [13].

Ears are rarely recorded in identification documents and even with the heightened concerns on privacy, there is often little consideration given to obscuring a subject's ears.

There are many studies for ear recognition. Hurley *et al.* [36] presented one popular holistic approach to ear recognition, which was Force Field Transform and the recognition accuracy has achieved 99.2%. Banerjee and Chatterjee [37] also use Force Field Transform, with ERJU dataset to achieve a recognition rate of 97.71%. Considering occlusion of the ears by hair, local approaches can have a better performance than holistic approaches. Because the local approaches are based on the description of local parts. Arbab-Zavar and Nixon [11] utilizes SIFT for ear recognition with accuracies of 91.5% and 80.4% for non-occluded ear images and 20% for occluded ear images. Compared with a holistic approach (PCA), the accuracy of non-occluded ear images for ear recognition has been around 98.4%. However, the ear recognition accuracy with 20% occlusion in ear images has been only 12.7%. Deep learning has also been used for ear recognition in recent years, motivated by its advantageous performance. However, the deep learning methods can often be limited by a lack of training data. Eyiokur *et al.* [24] studies the unconstrained ear recognition problem. A deep convolution neural network (CNN) model is used for ear recognition to show the capability for domain adaption. Sinha *et al.* [38] discuss the feasibility of deep neural networks (DNNS) in

ear biometrics. They use Histogram of Gradient (HOG) and support vector machines (SVMs) for ear localization, followed by a CNN for ear recognition. This approach has achieved 97.9% recognition accuracy. Almisreb *et al.* [39] applies Transfer Learning to the AlexNet Convolution Neural Network (AlexNet CNN). Because deep learning needs much data, the Transfer Learning is used to solve classification problems where there is little data available. The validation accuracy has achieved 100% in this study. Emeršič *et al.* [40] builds a CNN model for ear recognition in the wild. They explore different strategies towards model training with limited amounts of training data and show that by selecting an appropriate model architecture, using aggressive data augmentation and selective learning on existing (pre-trained) models, they are able to learn an effective CNN-based model using little more than 1300 training images, and the recognition rate at rank 1 is 62%. Emeršič *et al.* [42] presents the results of the unconstrained ear recognition challenge. This article contains the results from five groups and evaluates six ear recognition techniques. The challenges of unconstrained ear images include head rotation, flipping, gallery size, large-scale recognition and the others. They have found that the robust performance is based on smaller part of dataset, but there still is a significant performance drop when the entire dataset is used for testing. Ear recognition in the wild are also discussed in [43], [44]. Ying *et al.* [41] describes a human ear recognition algorithm based on a convolutional neural network. A final fully connected layer is introduced in Dropout technology and an optimal activation function is selected to prevent network overfitting. For the non-rotated ear image dataset, the recognition accuracy has achieved 98.12%.

In terms of identity science, Khorsandi and Abdel-Mottaleb [7] presents the first study for the gender classification using 2D ear image analysis based on a sparse representation. Gabor filters are used for feature extraction to achieve a performance of 89.5%. Lei *et al.* [9] reports the first study for gender classification from 3D ear images and achieve the accuracy of 92.9% for gender classification. Yaman *et al.* [12] demonstrates the extraction of soft biometric traits, age and gender, from ear images. The accuracies of gender classification in [12] are 65% and 94% with the geometric and appearance features, respectively. Meng *et al.* [13] uses a model-based approach for gender classification and kinship verification. The geometric features are used for the gender classification, with an accuracy of 67.2%. Yaman *et al.* [10] reports excellent gender identification 98% and with fusion with face profiles can reach very high accuracy (>99%). This is the current state-of-the-art in ear biometrics for gender classification. However, these studies are only interested in performance and not specificity. An important question is also that which parts of ear guide the gender classification? Yoga *et al.* [32] analyses the gender differences based on anthropometric measurements of ears. They use manual measurements of the total ear index and the lobe index to demonstrate the differences between female and male. Arbab-Zavar and Nixon [11] is the first to consider potency of different ear parts for ear identification. The recognition accuracy in [11] is 85.7%, and the top ten most significant parts of ears for recognition are shown with the most important being the inferior crus of the antihelix. For



Fig. 3. Force Field Transform of a sample.

surveillance applications, we invariably capture a subject's ear from one side and often the other sides are not available, similar to Fig. 1(b). Thus, the question is when we examine ear images, do we need to consider left ears or right ears? If we can establish the possibility of ear symmetry, then we can remove the need to capture views of both sides of the head, as only one is necessary.

In symmetry, Yan and Bowyer [15] is the first study to mention the symmetry of ears and indicate that around 90% of people's right and left ears show bilateral symmetry. Yan and Bowyer [14] consider 3D ear symmetry and formulate results of oblique ( $15^\circ$ ,  $30^\circ$ ,  $45^\circ$ ) views with a 24-subject dataset. Although these are naturally welcome advances, they do not present a concerted study on a large database. Abaza and Ross [16] presents an analysis of symmetry of human ears, using a semi-automated mode and geometric features, with an Equal Error Rate (EER) of 16.8%. They also use Eigen-Ears (PCA) and the Shape From Shading (SFS), with EERs of 21.1% and 17.1%, respectively. Toygar *et al.* [17] also reports symmetric ear recognition with profile face fusion. They use LBP, LPQ and BSIF algorithms, and the best performance of using ear images with no profile faces is around 76.1%.

Previous studies do not consider rotation, and ears are recorded in laboratory conditions under which a subject looks straight ahead. Fig. 1(b) shows how a subject's head might be inclined or rotated, especially when committing a crime. Only three angles are discriminated in [14], where ear symmetry is considered based on the appearance of ears, rather than its structure. However, some subjects' heads are rotated along the yaw or pitch axes, which can change the appearance of the ears.

This article is originally based on a study of which ear parts are used most for identification in ear biometrics, and which contributes to gender recognition [IWBf] [30]. Beyond greater formalism of the previous paper, we now show:

1. that recognition can be achieved in unconstrained images;
2. that human ears appear largely to be bilaterally symmetric; and
3. which ear parts contribute most to recognition of bilateral symmetry.

We are the first to utilize deep learning to study ear symmetry. We also use a model-based technique for comparison with deep learning, and we consider unconstrained ear images, whilst we measure the performance by training pre-processing ear images, by using the affine transformation to register the two ears for comparison. The work presented here is the current state-of-the-art in ear symmetry. We show for the first time that there is actually a high chance of symmetry between a subject's two ears, confirming the previous observations from



Fig. 4. HOG descriptions of an ear image.

human vision, and we show which ear regions are significant for evaluating ear symmetry.

## II. IMAGE PRE-PROCESSING

### A. Force Field Transform

The force field transform is inspired by electrostatics, and electromagnetics and is used within one of the earliest approaches to ear biometrics [5]. The image is transformed by assuming that it consists of an array of Gaussian attractors, which act as the source of a force field. Each pixel in the image affects the other pixels, and it therefore contributes to the energy affecting other pixels in the image. Associated with the force field generated by each pixel, there is a spherically symmetrical scalar potential energy field denoted by  $E_i(r_j)$ . This energy is the potential energy imparted to a pixel of unit intensity at the pixel location with position vector  $r_j$  by the energy field of any other pixel with position vector  $r_i$  and pixel intensity  $P(r_i)$ , and is given by

$$E_i(r_j) = \frac{P(r_i)}{|r_i - r_j|}$$

where each pixel affects the other pixels with the different distances. The potential energy transformation is

$$F(r) = -\nabla(E(r))$$

Fig. 3 depicts the results of the force field feature extraction applied to the ear image. As shown in Fig. 3, Force Field Transform helps to reduce the effects of hair whilst preserving ear structure.

### B. Affine Transformation

The affine transformation is defined as

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

This has six degrees of freedom corresponding to the six parameters of the transformation (scale, rotation and translation parameters) [27] and can be calculated from non-collinear correspondences. In our study, we consider ear images in which a person's head may rotate only about yaw axis, and we do not consider roll and pitch rotations. Fig. 5 shows an example of the affine transformation. Fig. 5(a) depicts the reference image, the image to be registered is also shown in Fig. 5(b). This ear image is then rotated around the yaw axis to result in the ear image presented in Fig. 5(c) by using an affine transformation.



Fig. 5. Applying the affine transformation.

### III. MODEL-BASED METHODOLOGY

The Histogram of Oriented Gradients (HOG) [18] is also employed here to describe ear images. There are five steps in computing the HOG descriptors. The first step is optional and is global image normalization, equalization, and gamma (power law) compression, and also the calculation of square root or log of every color channel. The second step is to calculate the gradient of image. This can capture the edges and some texture information in the image, while illumination variations are mostly ignored. The third step is to compute gradient histograms. An ear image is divided into small spatial regions, named as ‘cells’ here. For all the pixels inside every cell, a 1-D local histogram of gradient or edge orientation is accumulated. This 1-D histogram forms the basic for the ‘orientation histogram’ representation. Gradient angle range is divided into a fixed number of predetermined bins. Using the gradient value of pixels in cells, votes are taken in the orientation histogram. The fourth step is to normalize across blocks using local groups of cells, with responses normalized before the next stage. The normalized block descriptors are described as HOG descriptors. The final step is to collect HOG descriptors from all the blocks of a dense overlapping grid of blocks covering the detection window (cell), and then to construct a feature vector as an input for classification algorithms. We apply HOG on the ear images after force field transform. Fig. 4 shows the HOG descriptors of an ear image.

### IV. DEEP LEARNING

#### A. ResNet-50

The deep residual learning method addresses the degradation problem by introducing a deep residual network [19]. This method allows some stacked layers to fit a residual mapping, instead of hoping that these layers directly fit a desired underlying mapping. The original mapping is recast into  $\mathcal{F}(x)+x$ . The formulation of  $\mathcal{F}(x)+x$  can be realized by feedforward neural networks with skip connections. We use a residual network with 18, 34 and 50 layers. Because of concerns on the training time, the building block is modified as a bottleneck design [19]. The 50-layer ResNet contains 3-layer bottleneck blocks instead of the 2-layer blocks in the 34-layer network.

#### B. Fine Tuning ResNet-50

Transfer learning is used to improve the performance of the network from one domain by transferring information from a related domain [20], specially for small datasets such as our ear dataset. The strategy of the transfer learning is to pre-train a model on a large labelled dataset. As we have a limited dataset for the training, we use ResNet-50, and transfer

the learning of the pre-trained model (ResNet-50) to the ear dataset. For ear recognition, we replace the last fully connected layer and the classification layer in the pre-trained ResNet-50 with a set of layers that can classify 137 classes to recognize 137 subjects. For the case of gender classification, we replace those layers with layers that can classify into two groups for male and female classes. Finally, for ear symmetry experiments, we replace those layers with layers that can classify 100 classes to recognize 100 subjects.

#### C. Evaluating Visualizations

Grad-CAM [22] (Gradient-weighted Class Activation Mapping) uses the gradients of any target concept (say logits for ‘female’ or even a caption), flowing into the final convolution layer to produce a coarse localization map highlighting the important regions in the image. The class discriminative localization map Grad-CAM  $L_{Grad-CAM}^c \in \mathbb{R}^{u \times v}$  of width  $u$  and height  $v$  for any class  $c$ ,  $y^c$  (before the softmax), with respect to feature map  $A^k$  of a convolutional layer is  $(\partial y^c)/(\partial A^k)$ . The neuron importance weights  $\alpha_k^c$  are calculated as:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

which are obtained by global average pooling of the gradients derived via backpropagation. The weight  $\alpha_k^c$  represents a partial linearization of the deep network downstream from feature map  $A^k$ , and captures the ‘importance’ of the  $k^{th}$  feature map for a target class  $c$ .  $Z$  is the number of pixels in the feature map.

A weighted combination of forward activation maps is developed in [22], and this is followed by a ReLU to obtain the linear combination written as,

$$L_{Grad-CAM}^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right).$$

### V. EXPERIMENTS FOR EAR RECOGNITION

#### A. Datasets for Ear Recognition

We use USTB dataset 1 [28] and USTB dataset 2 [28] for ear recognition. USTB dataset 1 has 180 ear images with 60 subjects. Each subject has 3 different images, they are normal frontal ear image, non-rotated frontal ear image and ear image under different lighting conditions. Images in USTB dataset 1 has already experienced rotation and shearing. USTB dataset 2 has 308 ear images with 77 subjects, each subject contains 4 images. The first and the fourth images are profile images with varying illuminations. Compared with the first image, the second and the third images have the same illumination. These two ears have been rotated by  $\pm 30^\circ$  respectively. Fig. 6 shows samples of ear recognition images.

#### B. Model-Based Methodology

We use kNN ( $k=1$ ) with a leave-one-out cross validation strategy for ear identification. Euclidean and Manhattan distances are used in kNN classifier to evaluate the model. The results of correct recognition are shown in Table I.

As shown in Table I, the model-based method achieves an accuracy of 95.1% in ear recognition, which shows the model can be used for identification, with good performance.



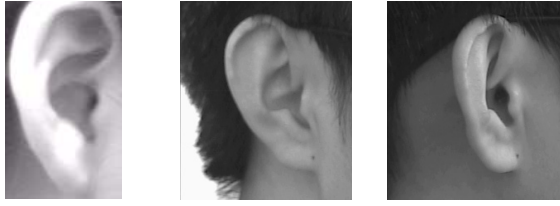


Fig. 6. Samples of ear recognition data.

TABLE I  
EAR RECOGNITION ACCURACIES BY MODEL-BASED METHOD

Method	Accuracy
Euclidean distance	94.5%
Manhattan distance	95.1%

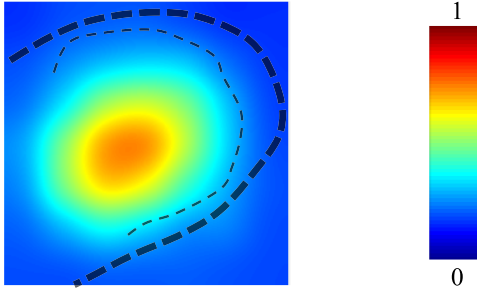


Fig. 7. Average heatmap for ear identification.

### C. Deep Learning

We deploy the transfer learning of ResNet-50 in this study. USTB Database 1 and Database 2 (308 images, 77 subjects with 4 images for each subject) are used for ear recognition. There are 351 images for training, and 137 images for testing. We select one test ear image from each subject randomly, so we select 137 test images from 137 subjects, and then we split the rest of images into training and validation datasets: 70% data for training and 30% data for verification. The accuracy of ear recognition is 92.9%, which means this model has good capacity for ear recognition. Compared with previous studies [24], [38]–[41], the highest accuracy was 100%. Almisreb *et al.* [39] uses 250 training images for 10 subjects. There is such a little amount of data in this dataset, and it contains just 10 subjects. However, when our model is trained on the rotated ear image datasets, we are training 348 images for 137 subjects. Also, the datasets used here include rotated ear images and the ear images are under weak illuminations. Therefore, our model appears to be robust to rotation and illumination. Then, we employ Grad-CAM to analyse which parts of the ear are significant for recognition. Fig. 7 shows the heatmap for ear recognition. The red parts in heatmaps are the most important parts, whereas the blue parts show those of less importance. As Fig. 7 shows, the central region of ear is more significant for ear recognition.

### D. Ear Recognition Verification

We also do verification experiments to evaluate the performance of our model. For the verification experiments, we select two ear images from each subject at random. We use the training model to represent each image by a vector,

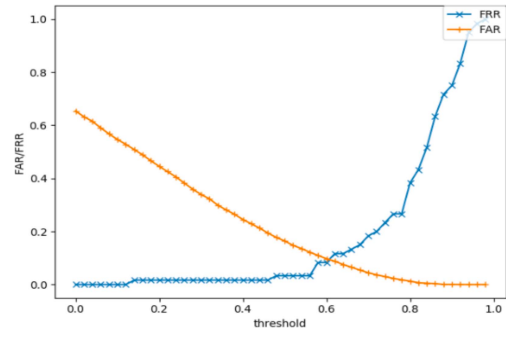


Fig. 8. Ear recognition verification.

and then calculate the cosine similarities as a distance metric between two images.

If the stored ear image of a subject  $A$  is represented by  $A_1$  and another ear image for verification is represented by  $A_2$ , thus, the null hypothesis ( $H_0$ ) and alternate hypothesis ( $H_1$ ) are defined as:

**Null hypothesis  $H_0$ :**  $A_1$  and  $A_2$  come from different subjects.

**Alternate hypothesis  $H_1$ :**  $A_1$  and  $A_2$  come from the same subject.

If the value of cosine similarity is higher, the system is more certain that  $H_1$  is correct. The system makes decision based on a threshold: if the cosine similarity value is higher than or equal to the threshold,  $H_1$  is correct. Otherwise, the cosine similarity value is lower than the threshold, and therefore  $H_0$  is confirmed.

False Reject Rate FRR is defined as rejecting the null hypothesis  $H_0$  when it is true, and False Accept Rate (FAR) is defined as failing to reject  $H_0$  when it is false. Equal error rate (EER) is the value defined as  $FRR = FAR$ , EER is an indicator of the performance of the model, the lower value for EER, the higher performance for the model.

Fig. 8 shows the recognition verification of ear images, and the Equal error rate (EER) for ear recognition verification is 7.24%.

## VI. ANALYSING THE DISTINCTIVENESS OF GENDER

### A. Datasets for Gender Classification

We explore the model-based technique for gender classification by using our own database (which we have made available on Github [33]) which contains 78 subjects: 38 females, and 40 males. Each subject contains two unconstrained profile images, which are from two sides. We use the USTB dataset 3 [28] and our own dataset for gender classification from ear images for deep learning. USTB dataset 3 contains 79 subjects. There are 10 ear images for each subject which include profile images with  $0^\circ$ ,  $5^\circ$ ,  $10^\circ$ ,  $15^\circ$ ,  $20^\circ$  rotation, and there are two images for every pose. Fig. 9 shows samples used for gender classification.

### B. Model-Based Methodology

We use the Euclidean, Manhattan and Mahalanobis distances for  $kNN$  with  $k = 1$  in a leave-one-out cross validation. Soft biometric data information includes gender, and relationship between subjects for all subjects, Table II shows accuracies of gender classification.

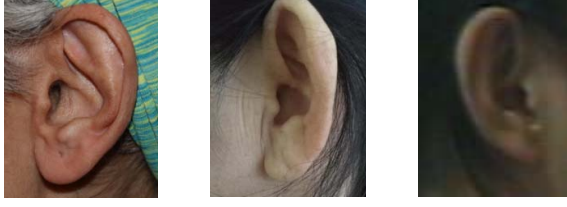


Fig. 9. Samples for gender classification.

TABLE II  
GENDER CLASSIFICATION ACCURACIES FOR OUR MODEL-BASED

Method	Accuracy
Euclidean distance	72.9%
Manhattan distance	80.0%
Mahalanobis distance	82.9%

Divided ( $2 \times 2$ ) regionsDivided ( $4 \times 4$ ) regions

Fig. 10. An example for a divided ear image.

r1	r2
r3	r4

r1	r2	r3	r4
r5	r6	r7	r8
r9	r10	r11	r12
r13	r14	r15	r16

Fig. 11. The order of regions (4 regions: left side, 16 regions: right side).

As observed in Table II, our model-based technique has achieved gender classification with accuracy of 82.9%, being 32.9% higher than random. The next question we address in this article is which parts of ears contribute to gender classification.

We divide the ear images into 4 and 16 parts, as shown in Fig. 10. The sub-regions are referred to as  $r_n$  (e.g.,  $r_1, r_2, \dots$ ) with order  $n$  shown in Fig. 11. We use each part for gender classification separately. Table III shows the accuracies for the 4 ear parts, and Table IV presents the accuracies for the 16 ear parts.

The accuracies of gender classification for each individual region are shown in Table III and Table IV. Table III shows the accuracies of the ( $2 \times 2$ ) sub-regions. Table IV shows the gender classification accuracies for 16 parts of ears. The results in Table III and Table IV are based on the Mahalanobis distance for  $kNN$  with  $k = 1, 3, 5, 7, 9$  in a leave-one-out cross validation strategy for gender classification.

As observed in Table III,  $r_1$  (in the four-region setting) enjoys the highest accuracy of 73.2%, which means  $r_1$  is the most important region for gender classification. Furthermore,  $r_2$  and  $r_4$  are the second most important regions for gender

TABLE III  
GENDER CLASSIFICATION ACCURACIES IN DIFFERENT PARTS OF EARS  
(4 REGIONS) BY USING OUR MODEL-BASED METHOD

Parts of ears	$k=1$	$k=3$	$k=5$	$k=7$	$k=9$
r1	62.0%	70.4%	<b>73.2%</b>	70.4%	67.6%
r2	53.5%	57.7%	60.5%	60.5%	57.7%
r3	47.8%	43.6%	33.8%	36.6%	35.2%
r4	52.1%	63.3%	60.5%	61.9%	60.5%

TABLE IV  
GENDER CLASSIFICATION ACCURACIES IN DIFFERENT PARTS OF EARS  
(16 REGIONS) BY USING OUR MODEL-BASED METHOD

Parts of ears	$k=1$	$k=3$	$k=5$	$k=7$	$k=9$
r1	63.4%	60.6%	52.1%	49.3%	50.7%
r2	53.5%	59.2%	59.2%	66.2%	69.0%
r3	35.2%	40.8%	39.4%	43.7%	42.3%
r4	<b>71.8%</b>	69.0%	64.8%	66.2%	66.2%
r5	61.9%	57.7%	53.5%	52.1%	53.5%
r6	42.3%	40.8%	49.3%	52.1%	47.9%
r7	50.7%	60.6%	56.3%	53.5%	49.3%
r8	50.7%	50.7%	50.7%	47.9%	47.9%
r9	42.3%	47.9%	56.3%	50.7%	49.3%
r10	57.7%	52.1%	49.3%	43.7%	42.3%
r11	67.6%	57.7%	59.2%	64.8%	62.0%
r12	52.1%	59.2%	60.6%	60.6%	63.4%
r13	43.7%	46.5%	45.1%	42.3%	38.0%
r14	50.7%	52.1%	50.7%	42.3%	45.1%
r15	49.3%	47.9%	49.3%	53.5%	59.2%
r16	52.1%	53.5%	52.1%	52.1%	56.3%

classification, whereas region  $r_3$  is considered as the worst region for gender classification. The accuracy in  $r_3$  is even lower than a random classifier.

For 16 region analysis, Table IV shows that the highest accuracy of 71.8% is associated with  $r_4$ . Region  $r_2$  has the second highest accuracy of 69.0%, both  $r_2$  and  $r_4$  being on the upper helix.

The results presented in Table III and Table IV, suggest that the upper helix is more important than lower parts of ears (such as the lower helix and the lobe) for gender classification. Regions  $r_4$  and  $r_2$  in the 16-region-division setting come from  $r_2$  and  $r_1$  in the 4-region-division setting, separately. This indicates that the upper regions of ears are more important than lower parts of ears.

### C. Deep Learning

We use the transfer learning with ResNet-50, VGG-16 and GoogleNet for gender classification. We use the USTB database 3 [28], our own database [33] and AWE dataset [34]. We select some images from each dataset, to have a merged dataset of ears to ensure a more balanced dataset between males and females. As a result, in our newly merged dataset there are 90 female subjects and 100 male subjects. There are 800 images for training and validation (split as: 70% data for training and 30% data for validation) and 110 images for testing. In all experiments in this article, we have removed the background of each ear image, to focus on just the ear and to exclude the effects of hair, head and neck regions.

As tabulated in Table V, the highest accuracy of the network based on transfer learning is associated with ResNet-50, which is 90.9%. This accuracy is sufficiently high, it is possible to use this model to classify subjects into females and males.

TABLE V  
GENDER CLASSIFICATION ACCURACIES FOR DEEP LEARNING

Method	Accuracy
ResNet-50	90.9%
VGG-16	80.9%
GoogleNet	81.8%

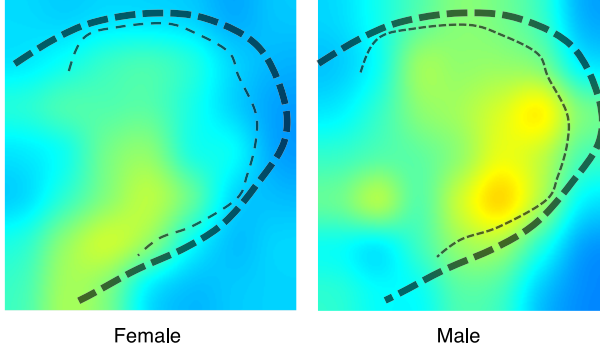


Fig. 12. Average heatmaps for different genders (ResNet-50).

We exploit Grad-CAM to compute the heatmaps based on ResNet-50 shown in Fig. 12 to determine which parts of ears contribute to gender classification from ear images.

Fig. 12 shows the average heatmaps for two genders: males and females, including a general outline of an ear to illustrate the important parts of the ear for gender classification. As depicted in Fig. 12, the difference between different genders is obvious, the upper helix, triangular fossa, tragus and antitragus of males is more important than the other parts of ears. For females, the lobe plays a more significant role in ear recognition. This may be due to the fact that women are more likely to wear earrings making the ear lobe more important for female recognition. Also, females are more likely than males to have long hair which increases the chance that the top of the ear is occluded and thus not used in their recognition. Conversely, males are more likely to have short hair, and their recognition is based on their parts that are more easily observed, including the antitragus and the antihelix. Some of the conclusions drawn for the heatmaps depicted in Fig. 12 differ from those obtained in our model-based technique (see Table III and Table IV), i.e., the upper helix is very important for gender classification, not only in our model-based technique but also as shown by the difference between the heatmaps in Figure 12. However, the lobe has not played an important role in the model-based technique for gender classification, but it appears very significant for the convolutional network to classify genders.

It is interesting to compare identification with gender recognition. Fig. 7 shows the average heatmap for ear identification. This heatmap demonstrates that the central parts of ears are very important for ear identification. Obviously, heatmaps for identification and gender classification are different, and some regions such as the helix, strike a balance between a strong positive response for males and a strong negative response for females. As the heatmaps of different genders demonstrate, the helix and lobe are important for gender classification. The central parts are also very important for ear recognition.



SCface dataset

AWE dataset

Fig. 13. Samples of AWE dataset.

## VII. EAR SYMMETRY

In this section, we consider bilateral symmetry for ear recognition. We use a pre-trained deep learning network as well as a model-based method for identification in an ear symmetry context. Our results are presented in Sections VII-B, VII-C, and VII-D.

### A. Datasets for Ear Symmetry

For our symmetry analysis, we exploit the SC face database [23] collected by The Technical University of Zagreb as well as Annotated Webs Ears (AWE) [34] database collected by University of Ljubljana.

The SC face database has 4160 images from 130 subjects. There is a subset of this database which contains each subject. The dataset contains 130 subjects, each subject has 9 images with different pose, one is mug shot at  $0^\circ$ , the other images rotated from  $-90^\circ$  to  $+90^\circ$  (4 left side images and 4 right side images). However, ears from some subjects have been entirely occluded by hair. We therefore choose 100 subjects with no occlusion for the experiments.

The AWE dataset contains 100 subjects with 1000 ear images. 520 left ear images and 480 right ear images. All the images are collected from the Web by first compiling a list of people, for whom it is reasonable to assume that a large number of images could be found online. Similar to other datasets collected in the wild, this list mostly features actors, musicians, politicians and the like [35]. Fig. 13 shows some examples of ear symmetry experiments.

### B. Model-Based Methodology

In this subsection, we use a model-based technique for matching a pair of ears which is based on the structure of ears. Hough Transform [31] is deployed to detect ears whose shapes are approximated as ellipses, and then we find the outer rectangle enclosing the ellipse. Finally, the histogram of gradients of ears is computed. We use one of SC face datasets for ear symmetry. In it, there are 300 left ear images for training, and 300 right ear images for testing. As both ears are used, we mirror the right ear images. Zero mean Gaussian noise is also added to ear images to result in noisy ear images with SNR 22 dB. The kNN classifier is then used with Euclidean and Manhattan distances for ear classification.

Tables 6, 7 and 8 show the results of our model-based technique, and they present the accuracies for original images, noisy images and noisy images after affine transformation

TABLE VI  
EAR RECOGNITION ACCURACIES FOR OUR MODEL-BASED METHOD  
(ORIGINAL IMAGES)

Distance	Accuracy
Manhattan distance	58.9%
Euclidean distance	60.7%

TABLE VII  
EAR RECOGNITION ACCURACIES FOR OUR MODEL-BASED METHOD  
(IMAGES WITH NOISE)

Distance	Accuracy
Manhattan distance	21.3%
Euclidean distance	23.1%

TABLE VIII  
EAR RECOGNITION ACCURACIES FOR OUR MODEL-BASED METHOD  
(IMAGES AFTER AFFINE TRANSFORMATION)

Distance	Accuracy
Manhattan distance	66.9%
Euclidean distance	65.0%

TABLE IX  
TWO-SIDE EAR RECOGNITION ACCURACIES FOR DEEP LEARNING

Images	Accuracy
Original image	91.2%
Images with noise	62.9%
Affine transformed Images	93.1%

respectively. Affine transformed ear images enjoy the highest accuracy of 66.9%. The recognition rate of original images is 60.7%, otherwise, the accuracy of ear images with Gaussian noise is 23.1%, which is lowered by 37.6%. We also use the force field transform and uniform local binary pattern (ULBP) for extracting features, but HOG has slightly better performance than force field transform and ULBP.

### C. Deep Learning

In the experiment described in this subsection, we transfer the pre-trained ResNet-50 to the SC face datasets. All the left ear images for training, and right ear images for testing.

As observed in Table IX, the accuracy of the network based on transfer learning is as high as 93.1%. The recognition rate of training original ear images is 91.2%, otherwise the accuracy of training the ear images with Gaussian noise is 62.9%, which decreases by 28.3%. Clearly, the presence of noise lowers the recognition rate, as expected, affine transformed ear images are responsible for the highest accuracy. In our experiments, we notice that the hair occlusion influences the correct prediction. Therefore, the hair occlusion is a challenge for verifying ear symmetry. For comparison purposes, we transfer pre-trained VGG-16 and GoogLeNet networks to the dataset, and the accuracies are 66.2% and 62.1% respectively. The performance of ResNet-50 network is better than the other two networks, therefore, we will analysis heatmaps of ResNet-50. In addition, we apply this model to an in-the-wild ear image dataset (AWE dataset), we use 520 left ear images for training, and 480 right ear images for testing, and the ear images have severe rotation in pitch, and yaw axes. Also, the ear images in the dataset have severe occlusions. The recognition accuracy for ear images without pre-processing is 20.2%.

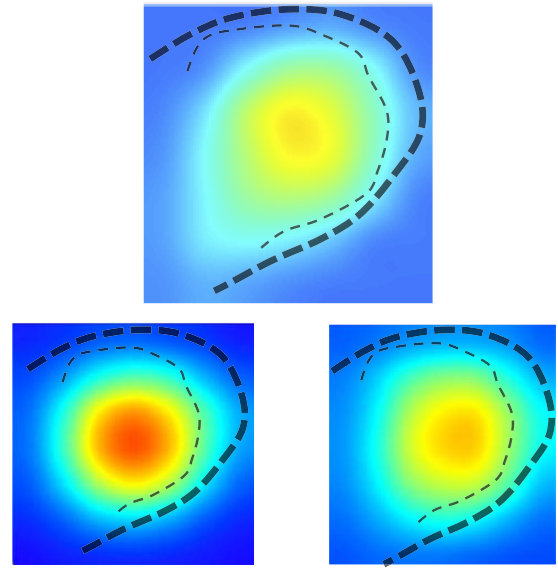


Fig. 14. Average heatmap for recognition of single ear from either side (Top: original images, Bottom left: affine transformed images, Bottom right: images with added Gaussian noise).

Although the accuracy is not high, it is still 19.2% higher than random chance. Compared with SC face dataset, the accuracy has a sharp decrease of 70.3%, therefore, our results indicate that applying the system to the in-the-wild ear images suggest a considerable challenge.

It is interesting to compare deep learning approaches with model-based techniques. As presented in all the tables, deep learning demonstrates better performance. The results of training original images indicate that deep learning is superior to model-based. Meanwhile, these two methods also use ear images after affine transformation in the training images. As a result, the accuracy of deep learning increases by 1.9%, while that of model-based method increases by 8%. As demonstrated in these experiments, affine transformation for model-based is actually more helpful than that for deep learning. Furthermore, we have contaminated the ear images with zero mean Gaussian noise for training. Additive Gaussian noise causes the accuracies to decline. The deep learning accuracy is lowered by 28.3%, while model-based decreases by 37.6%. Our results indicate that deep learning enjoys better performance than using a model, in a noisy environment.

Moreover, we determine which region of ears contributes most in symmetry-based recognition rates by using Grad-CAM to compute the heatmaps. We mirror right ears to left ones and match the right ears to the left ones. Fig. 14 demonstrates the contribution of different parts of ears.

For the heatmaps, red pixels are the most important, and blue pixels are less important. Therefore, the central part of ears is very important for recognition of single ears from either side, and it is also significant for noisy images and affine transformed images. Then, we compare the recognition of single ear from either side with one-sided ear recognition, and they have similar heatmaps. We also consider if the model can test the ear symmetry directly, by an image combining a subject's left ear and right ear, as shown in Fig. 15.

We also transfer the ResNet-50 learnings to one of SC face datasets. There are 300 combined two sides ear images from



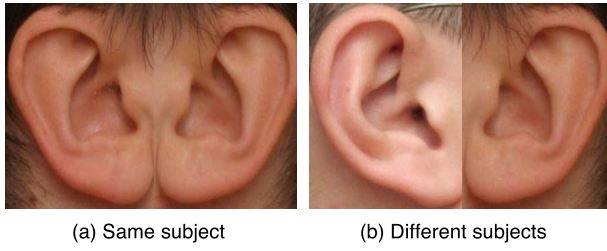


Fig. 15. Composite images of two sides.

TABLE X  
EAR SYMMETRY ACCURACIES FOR DEEP LEARNING

Images	Accuracy
Original images	100%
Noisy images	100%
Affine transformed Images	100%

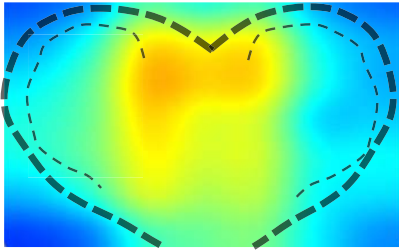


Fig. 16. Average heatmap for ear symmetry.

same subjects (100 subjects) and 300 combined ear images from different subjects (100 different subjects are selected randomly). 150 ear images from same subject (50 subjects) and 150 ear images from different subjects for training, and the others for testing. We divide the testing set to 5 sub-sets, every sub-set has 30 ear images from same subject and 30 ear images from different subjects. Table X shows the results of ear symmetry.

Table X shows the ear symmetry accuracies. The classification accuracy for original images, affine transformed images and noisy images are 100%. As observed from Table X, the model can classify all the ear images into the correct groups, and the model is not affected by noise. These accuracies along with the recognition rates reported in Table IX, indicate that human ears are bilaterally symmetric. Fig. 16 shows the average heatmap of an ear image in this experiment.

As shown in Fig. 16, the middle parts are very important for ear symmetry. If we compare the heatmap shown in Fig. 12 with the heatmaps of gender classification, we notice that the lobe is significant for gender classification, but not significant for ear symmetry. In fact, it appears reasonable that those sections of the ear that appear to dominate gender analysis contribute less to ear symmetry. The classification accuracy in the ear symmetry experiment is 6.9% higher than the recognition rate of single ear from either side, and by comparison with Fig. 14, we observe which parts of ears are not symmetric. As the heatmaps show the helixes and lobes are not significant for ear symmetry, and triangular fossa is most important for ear symmetry.

In the next experiment, we consider the effect of wearing earrings on heatmaps. Earrings are usually worn on the ear

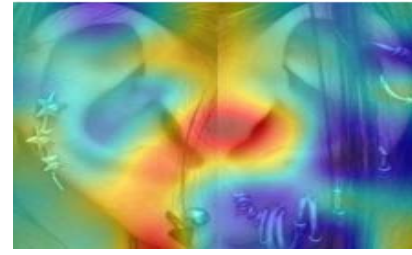


Fig. 17. Heatmap affected by earrings.

helix, and the positions of the earrings differs among helices. Fig. 17 shows a heatmap for ear images derived only from subjects wearing earrings. As shown in this figure, the red regions remain concentrated on the central section, which suggests that the ear symmetry is less affected by earrings.

#### D. Ear Symmetry Verification

Deep learning presents high accuracy for ear symmetry recognition; thus, we also do verification experiments to evaluate the performance of our model. For the verification experiments, we select a left ear image and a right ear image from each subject randomly. We use the training model to represent each image with a vector, and then calculate the cosine similarities of each image.

If the stored left ear biometric template of a subject  $A$  is represented by  $L_A$  and the right ear for verification is represented by  $R_A$ , thus, the null hypothesis ( $h_0$ ) and alternate hypothesis ( $h_1$ ) are defined as:

**Null hypothesis  $h_0$ :** Right ear  $R_A$  and left ear  $L_A$  come from different subjects.

**Alternate hypothesis  $h_1$ :** Right ear  $R_A$  and left ear  $L_A$  come from the same subject.

If the score of cosine similarity is higher, the system is more certain that  $h_1$  is more likely to be confirmed. The system decision is regulated by threshold  $T$ : if the score is higher than or equal to the threshold  $T$ ,  $h_1$  is confirmed. Otherwise, the score is lower than  $T$ , and the system infers that  $h_0$  is correct.

There are two incorrect conclusions. Type I error is defined as rejecting the null hypothesis  $H_0$  when it is true, also named False reject rate (FRR). Type II error or False accept rate (FAR) is defined as failing to reject  $H_0$  when it is false. Equal error rate (EER) is the value defined as  $FRR = FAR$ , EER is an indicator of the performance of the model, the lower value for EER, the higher performance for the model. Fig. 18 also presents FRR and FAR with respect to the threshold  $T$  for the affine transformed ear images. Fig. 18 indicates that the model enjoys high performance and high sensitivity for affine transformed ear images with lowest EER.

Table XI tabulates the EER for different experiments. EER for original ear images, affine transformed ear images and noisy ear images are 3.57%, 3.24% and 10.27% respectively.

Fig. 19 presents the inter- and intra-class variations for affine transformed ear images. In this figure, green bars represent the distance histogram between same subjects-client, and red bars show the distance between different subjects-imposter. The green bar is small, the red bar is large, the confusion region (the overlap region between the green and red histograms) is

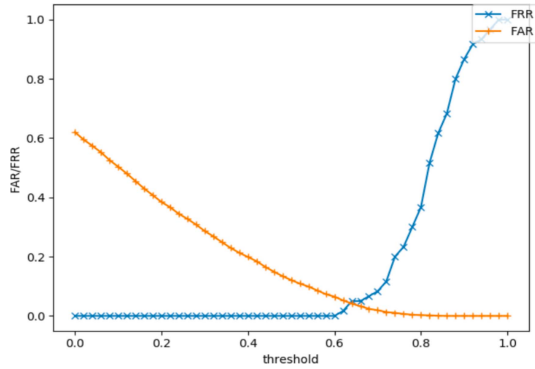


Fig. 18. Ear symmetry verification for affine transformed ear images.

TABLE XI  
EQUAL ERROR RATE OF DIFFERENT IMAGES

Images	EER
Original ear images	3.57%
Affine transformed ear images	3.24%
Ear images with Gaussian noise	10.27%

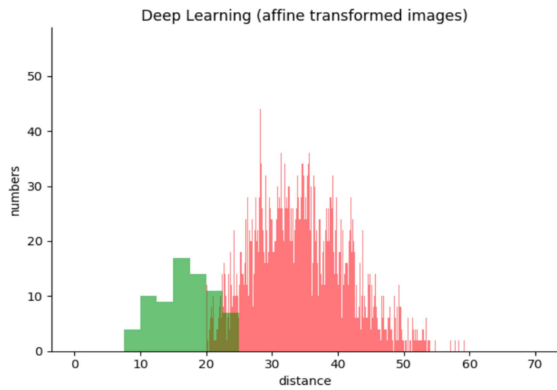


Fig. 19. Inter and intra class variations for affine transformed images.

smallest. Such a small confusion region indicates a high recognition rate when original ear images are used. The model has the best performance for the affine transformed ear images, affine transformation can compensate for rotation.

## VIII. CONCLUSION

This article presents a model-based technique as well as a deep learning method for ear recognition, gender classification and ear recognition with bilateral symmetry. For the model-based methodology, the accuracies are 95.1%, 82.9% and 66.9% for ear recognition, gender classification and ear recognition with bilateral symmetry, respectively. For deep learning, the performances for ear recognition, gender classification, and ear recognition with bilateral symmetry are 92.9%, 90.9% and 93.1% respectively. Also, we use deep learning for paired ear recognition. Our results confirm symmetry with a performance of 100%. We are the first to use deep learning for ear recognition with bilateral symmetry.

Meanwhile, we also consider which parts of ears contribute mostly to ear recognition, gender classification and ear bilateral symmetry. The heatmaps presented in this article, indicate that the central region is significant for ear recognition and

ear bilateral symmetry. Furthermore, we compare the heatmap associated with ear recognition with that of gender classification. The lobe and upper helix appear to be important for gender classification. However, the lobe does not play an important role in ear recognition and ear bilateral symmetry. Moreover, our analysis related to ear bilateral symmetry demonstrates that recognition rate is little affected by jewelry.

We demonstrate, in this article, through a set of experiments that there is an implied similarity between left and right ears. Therefore, ear recognition can be achieved regardless of which ear is used for analysis. This notion is important because human anatomy dictates that only one ear can be seen in a single image. Obviously, ears cannot be seen or analysed in images from a full-frontal view, like passport style images, which rarely occur in surveillance. As such our study therefore indicates that constraints on ear image acquisition are reduced. As such, this article establishes benchmarks for ear recognition by using ear bilateral symmetry and provides pointers for future applications and developments for ear symmetry particularly that it appears prudent for recognition to focus on the ear perimeter.

For future work, we aim to fuse the model-based approach with deep learning to improve the ear recognition performance. The effects of ear occlusion by hair is also another topic for future work. Another interesting topic for future work is to use deep learning attention networks for ear biometric to avoid non-ear regions of ear images in ear recognition. The last but not the least topic for future work is to avoid using transfer learning for the training of our deep learning network in order to produce heat maps with no effects of transfer learning, where non ear regions appear as least significant in heat maps. In such a scenario, the interpretation of heat maps will be clearer and more straightforward in the analysis of the deep learning methods for ear recognition, gender classification and ear symmetry. However, in such a case, a dataset with around 10,000 ear images for training would be needed.

## ACKNOWLEDGMENT

The authors remain grateful to the reviewers for their excellent comments and observations made on this article.

## REFERENCES

- [1] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, "A survey on ear biometrics," *ACM Comput. Surveys*, vol. 45, no. 2, pp. 1–35, 2013.
- [2] A. V. Iannarelli, *Forensic Identification Series: Ear Identification*. Fremont, CA, USA: Paramount, 1989.
- [3] A. Bertillon, *La Photographie Judiciaire: Avec Un Appendice Sur La Classification Et L'Identification Anthropométriques*. Paris, France: Gauthier-Villars, 1890.
- [4] M. Burge and B. Wilhelm, "Ear biometrics," in *Biometrics*. Boston, MA, USA: Springer, 1996, pp. 273–285.
- [5] D. J. Hurley, M. S. Nixon, and J. N. Carter, "A new force field transform for ear and face recognition," in *Proc. Int. Conf. Image Process.*, vol. 1, 2000, pp. 25–28.
- [6] M. S. Nixon, P. L. Correia, K. Nasrollahi, T. B. Moeslund, A. Hadid, and M. Tistarelli, "On soft biometrics," *Pattern Recognit. Lett.*, vol. 68, pp. 218–230, Dec. 2015.
- [7] R. Khorsandi and M. Abdel-Mottaleb, "Gender classification using 2-D ear images and sparse representation," in *Proc. IEEE Workshop Appl. Comput. Vision (WACV)*, 2013, pp. 461–466.
- [8] P. Gnanasivam and S. Muttan, "Gender classification using ear biometrics," in *Proc. 4th Int. Conf. Signal Image Process. (ICSIP)*, 2013, pp. 137–148.

- [9] J. Lei, J. Zhou, and M. Abdel-Mottaleb, "Gender classification using automatically detected and aligned 3D ear range data," in *Proc. Int. Conf. Biometrics (ICB)*, 2013, pp. 1–7.
- [10] D. Yaman, F. I. Eyiokur, and H. K. Ekenel, "Multimodal age and gender classification using ear and profile face images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 2414–2421.
- [11] B. Arbab-Zavar and M. S. Nixon, "On guided model-based analysis for ear biometrics," *Comput. Vis. Image Understanding*, vol. 115, no. 4, pp. 487–502, 2011.
- [12] D. Yaman, F. I. Eyiokur, N. Sezgin, and H. K. Ekenel, "Age and gender classification from ear images," in *Proc. Int. Workshop Biometrics Forensics (IWBF)*, 2018, pp. 1–7.
- [13] D. Meng, M. S. Nixon, and S. Mahmoodi, "Gender and kinship by model-based ear biometrics," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, 2019, pp. 1–5.
- [14] P. Yan and K. W. Bowyer, "Biometric recognition using 3D ear shape," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1297–1308, Aug. 2007.
- [15] P. Yan and K. W. Bowyer, "Empirical evaluation of advanced ear biometrics," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, 2005, p. 41.
- [16] A. Abaza and A. Ross, "Towards understanding the symmetry of human ears: A biometric perspective," in *Proc. 4th IEEE Int. Conf. Biometrics Theory Appl. Syst. (BTAS)*, 2010, pp. 1–7.
- [17] Ö. Toygar, E. Alqaralleh, and A. Afaneh, "Symmetric ear and profile face fusion for identical twins and non-twins recognition," *Signal Image Video Process.*, vol. 12, no. 6, pp. 1157–1164, 2018.
- [18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, 2005, pp. 886–893.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [20] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [21] M. S. Nixon and A. S. Aguado, *Feature Extraction and Image Processing for Computer Vision*. Amsterdam, The Netherlands: Academic, 2019.
- [22] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.
- [23] *SCface—Surveillance Cameras Face Database*. Accessed: Jan. 28, 2020. [Online]. Available: <http://www.scface.org/>
- [24] F. I. Eyiokur, D. Yaman, and H. K. Ekenel, "Domain adaptation for ear recognition using deep convolutional neural networks," *IET Biometrics*, vol. 7, no. 3, pp. 199–206, 2018.
- [25] I. I. Ganapathi, S. S. Ali, and S. Prakash, "Geometric statistics-based descriptor for 3D ear recognition," *Visual Comput.*, vol. 36, no. 1, pp. 161–173, 2020.
- [26] B. Arbab-Zavar, M. S. Nixon, and D. J. Hurley, "On model-based analysis of ear biometrics," in *Proc. 1st IEEE Int. Conf. Biometrics Theory Appl. Syst.*, 2007, pp. 1–5.
- [27] Z. Song, S. Zhou, and J. Guan, "A novel image registration algorithm for remote sensing under affine transformation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4895–4912, Aug. 2014.
- [28] *Introduction to USTB Ear Image Databases*. Accessed: Jan. 28, 2020. [Online]. Available: [http://www.ustb.edu.cn/resb/en/doc/Imagedb\\_123\\_intro\\_en.pdf](http://www.ustb.edu.cn/resb/en/doc/Imagedb_123_intro_en.pdf)
- [29] K. Messer, J. Matas, J. Kittler, J. Luetten, and G. Maitre, "XM2VTSDB: The extended M2VTS database," in *Proc. 2nd Int. Conf. Audio Video Based Biometrics Person Authentication*, vol. 964, 1999, pp. 965–966.
- [30] D. Meng, S. Mahmoodi, and M. S. Nixon, "Which ear regions contribute to identification and to gender classification?" in *Proc. 8th Int. Workshop Biometrics Forensics (IWBF)*, 2020, pp. 1–6.
- [31] B. Arbab-Zavar and M. S. Nixon, "On shape-mediated enrolment in ear biometrics," in *Proc. Int. Symp. Visual Comput.*, 2007, pp. 549–558.
- [32] S. Yoga, J. Balaihi, V. Rangdhol, S. Vandana, S. Paulose, and L. Kavya, "Assessment of age changes and gender differences based on anthropometric measurements of ear: A cross-sectional study," *J. Adv. Clin. Res. Insights*, vol. 4, no. 4, pp. 92–95, 2017.
- [33] *dm-vlc/dataset*. Accessed: Jun. 27, 2020. [Online]. Available: <https://github.com/dm-vlc/dataset/find/master>
- [34] *Ear Recognition Research*. Accessed: Sep. 18, 2020. [Online]. Available: <http://awe.fri.uni-lj.si/>
- [35] Ž. Emeršič, V. Štruc, and P. Peer, "Ear recognition: More than a survey," *Neurocomputing*, vol. 255, pp. 26–39, Sep. 2017.
- [36] D. J. Hurley, M. S. Nixon, and J. N. Carter, "Force field feature extraction for ear biometrics," *Comput. Vis. Image Understanding*, vol. 98, no. 3, pp. 491–512, 2005.
- [37] S. Banerjee and A. Chatterjee, "Ear recognition using force field transform and collaborative representation-based classification with single training sample per class," in *Intelligent Computing and Applications*. New Delhi, India: Springer, 2015, pp. 511–517.
- [38] H. Sinha, R. Manekar, Y. Sinha, and P. K. Ajmera, "Convolutional neural network-based human identification using outer ear images," in *Soft Computing for Problem Solving*. Singapore: Springer, 2019, pp. 707–719.
- [39] A. A. Almisreb, N. Jamil, and N. M. Din, "Utilizing AlexNet deep transfer learning for ear recognition," in *Proc. 4th Int. Conf. Inf. Retrieval Knowl. Manag. (CAMP)*, 2018, pp. 1–5.
- [40] Ž. Emeršič, D. Štepec, V. Štruc, and P. Peer, "Training convolutional neural networks with limited training data for ear recognition in the wild," 2017. [Online]. Available: [arXiv:1711.09952](https://arxiv.org/abs/1711.09952).
- [41] T. Ying, W. Shining, and L. Wanxiang, "Human ear recognition based on deep convolutional neural network," in *Proc. Chin. Control Decis. Conf. (CCDC)*, 2018, pp. 1830–1835.
- [42] Z. Emeršič *et al.*, "The unconstrained ear recognition challenge," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, 2017, pp. 715–724.
- [43] E. E. Hansley, M. P. Segundo, and S. Sarkar, "Employing fusion of learned and handcrafted features for unconstrained ear recognition," *IET Biometrics*, vol. 7, no. 3, pp. 215–223, 2018.
- [44] Y. Zhou and S. Zafeiriou, "Deformable models of ears in-the-wild for alignment and recognition," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, 2017, pp. 626–633.
- [45] *ImageNet*. Accessed: Sep. 28, 2020. [Online]. Available: <http://www.image-net.org/>



**Di Meng** received the M.Sc. degree from Hong Kong Baptist University. She is currently pursuing the Ph.D. degree in ear biometric with the University of Southampton, U.K. Her research interests include computer vision, biometrics, and image processing.



**Mark S. Nixon** (Member, IEEE) is a Professor with the Vision, Learning, and Control Research Group, School of Electronics and Computer Science, University of Southampton. His research interests are in image processing and computer vision and his team develops new techniques for static and moving shape extraction, which have found application in automatic face and automatic gait recognition and in medical image analysis. He has been a Chair/program Co-Chair BMVC'98, ICPR'04, IEEE BTAS'10, ICPR'16, IEEE ISBA'16, IAPR/IEEE IJCB'17, and ICB'19. He is the President of the IEEE Biometrics Council and a Vice Chair IEEE PSPB. He is a Fellow of the IET and IAPR (for services to biometrics and computer vision), and the Distinguished Fellow of the BMVA 2015.



**Sasan Mahmoodi** (Member, IEEE) received the Ph.D. degree from the University of Newcastle, Newcastle upon Tyne, U.K., in 1998. In 1999, he held a Research Fellowship post in mathematical analysis of human vision with the University of Newcastle. He joined the University of Southampton in 2007 as an Assistant Professor. He has also published more than 80 publications and chaired and co-chaired several conferences, such as MIUA. His current research interests are in biometrics, medical image analysis, computer vision, and machine learning. He has been in the editorial boards of some internationally prestigious journals, such as multidimensional systems and signal processing and has also been a guest editor in the special issues in various journals in pattern recognition.