Ear-based biometric authentication through the detection of prominent contours

Aviwe Kohlakala and Johannes Coetzer

***Abstract*—In this paper novel semi-automated and fully automated ear-based biometric authentication systems are proposed. The region of interest (ROI) is manually specified and automatically detected within the context of the semi-automated and fully automated systems, respectively. The automatic detection of the ROI is facilitated by a convolutional neural network (CNN) and morphological postprocessing. The CNN classifies sub-images of the ear in question as either foreground (part of the ear shell) or background (homogeneous skin, hair or jewellery). Prominent contours associated with the folds of the ear shell are detected within the ROI. The discrete Radon transform (DRT) is subsequently applied to the resulting binary contour image for the purpose of feature extraction. Feature matching is achieved by implementing an Euclidean distance measure. A ranking verifier is constructed for the purpose of authentication. In this study experiments are conducted on two independent ear databases, that is (1) the Mathematical Analysis of Images (AMI) ear database and (2) the Indian Institute of Technology (IIT) Delhi ear database. The results are encouraging. Within the context of the proposed semi-automated system, accuracies of 99.20% and 96.06% are reported for the AMI and IIT Delhi ear databases respectively.**

***Index Terms*—ear shell, biometric authentication, convolutional neural network**

I. INTRODUCTION

A Biometric system performs personal authentication based on a specific physiological or behavioural characteristic of the individual. Biometric systems are increasingly utilised for security purposes, as they are more reliable and secure than most traditional modes of personal authentication, such as access cards, personal identification numbers and passwords. The human ear is one of the most distinctive human biometric traits that can be employed to establish or verify an individual's identity. Furthermore, the human ear constitutes a relatively stable structure that evolves very little with aging and may be acquired in a non-intrusive manner.

The concept of ear-based biometric authentication emerged relatively recently as an active field of research. Traditional systems rely on the extraction of hand-crafted features, while modern systems are able to learn so-called deep features by employing neural networks. A traditional ear-based biometric authentication system typically involves (1) segmentation of the ear or the detection of the region of interest (ROI), followed by (2) feature extraction and (3) feature matching (recognition or verification).

Different types of ear segmentation techniques, that is the localisation of the ear shell within an ear image have been investigated. A semi-automated ear detection technique based on an improved Adaboost algorithm and an active shape model (ASM) was proposed by Yuan and Mu [1], while an automated ear detection technique based on the combined use of the circular Hough transform and anthropometric ear proportions was presented by Vélez, Sánchez, Moreno and Sural [2].

A variety of algorithms have been proposed for extracting discriminative features from ear images. The extracted hand-crafted features are generally categorised into the following three types: geometrical features, local appearance-based features, and global features.

A number of geometrical feature extraction techniques that characterise the shape of the ear have been presented. Othman, Alizadeh and Sutherland [3] proposed a novel ear description technique based on a shape context descriptor. Annapurani, Sadiq and Malathy [4] proposed a technique that fuses the shape of the ear shell and tragus in such a way that a feature template is obtained. Omara, Zhang and Zuo [5] proposed a technique that uses the lines of minimum and maximum height associated with the ear's contour image to describe the outer helix.

Extensive techniques have been suggested to extract local appearance-based features from ear images. The scale invariant feature transform (SIFT) algorithm was employed by Anwar, Ghany and ElMahdy [6] and a robust algorithm for local similarity invariant extraction was proposed by Galdámez, Arrieta and Ramón [7]. A feature extraction technique based on the fusion of texture-based features (through local binary patterns) and geometric features (through the Laplacian filter) was proposed by Jiddah and Yurtkan [8].

Within the context of global feature extraction, techniques such as principal component analysis as proposed by Querencias-Uceta, Ríos-Sánchez and Sánchez-Ávila [9], and a technique based on a combination of the wavelet and discrete cosine transforms as proposed by Ying, Debin and Baihuan [10] have been investigated.

A number of feature matching protocols for quantifying the difference between two ears have been proposed, which include the utilisation of the Euclidean distance ([4], [7]), and the Hamming distance [4], as well as a minimum distance classifier [6], a k-nearest neighbour (KNN) classifier [8] and a nearest neighbour classifier that is based on a weighted distance [10].

As previously mentioned, the extraction of deep features from ear images has been investigated more recently. A number of deep learning techniques have been proposed. Dodge,

Based on "On automated ear-based authentication", by A. Kohlakala and J. Coetzer which was published in the Proceedings of the SAUPEC/RobMech/PRASA 2020 Conference held in Cape Town, South Africa from 29 to 31 January 2020. ©2020 SAIEE. "We would like to express our sincere gratitude to the Ball family for funding this research."

A. Kohlakala is with the Department of Mathematical Sciences, Stellenbosch University, Stellenbosch, South Africa (email: avi.kohlakala@gmail.com).
J. Coetzer is with the Department of Mathematical Sciences, Stellenbosch University, Stellenbosch, South Africa (email: jcoetzer@sun.ac.za).

Mounsef and Karam [11] used deep neural networks for the explicit purpose of feature extraction. Kacar and Kirci [12] introduced a novel architecture called ScoreNet for unconstrained ear recognition; the architecture fuses a modality pool with a learning approach based on deep cascaded score-level fusion. Hansley, Segundo and Sarkar [13] proposed an unconstrained ear recognition framework based on a convolutional neural network (CNN) model for ear normalisation and description, which is subsequently fused with hand-crafted features.

In this paper novel semi-automated and fully automated ear-based biometric authentication systems are developed. Within the context of the fully automated system, a CNN is designed for the purpose of facilitating the automatic detection of a suitable ROI which contains the entire ear shell. Within the context of the semi-automated system, the ROI is manually selected. Robust prominent contours that correspond to the folds of the ear shell are subsequently isolated within the ROI. These contours serve as input for a hand-crafted feature extraction protocol that is based on the calculation of the discrete Radon transform (DRT). A template matching protocol is employed that quantifies the difference between corresponding feature vectors through the calculation of the Euclidean distance. A rank-based verifier is finally constructed for the purpose of establishing the authenticity of a questioned ear image. The aforementioned steps are discussed in more detail in Section II.

The systems proposed in this paper are evaluated on the Mathematical Analysis of Images (AMI) and Indian Institute of Technology (IIT) Delhi databases. A detailed description of these databases is provided in Section III. In Table VII, the proficiency of existing systems also evaluated on the above-mentioned databases are compared to that of the semi-automated system proposed in this paper.

The paper is structured as follows. Section II details the design of the proposed systems. Section III introduces the data, outlines the experimental protocol, and analyses the results. Possible avenues for future research are laid out in Section IV.

II. SYSTEM DESIGN

An overview of the enrollment and authentication stages of the proposed semi-automated and fully automated ear-based biometric authentication systems are conceptualised in Figure 1.

A. Image segmentation

In the case of the semi-automated system a suitable ROI that contains the ear shell is manually selected, while for the fully automated system a suitable ROI is automatically detected. A CNN-based model (see Figure 2) is proposed to facilitate automatic ROI detection. The proposed CNN consists of four convolutional layers, where each of these layers is followed by a batch normalisation (BN), rectified linear unit (ReLU) and/or max pooling layer. The final pooling layer is followed by two fully connected (FC) layers. A detailed description pertaining to the mathematical underpinning of the CNN can be found in ([14], [15]).

For each database, ear images from different individuals are used for *training*, *validation* and *evaluation* purposes. The training set (seen data) is used to learn the parameters (weights) for the CNN in question, the validation set is used for avoiding overfitting by enforcing a stopping condition, while the evaluation set is used to measure the performance of the CNN on unseen data.

Each ear image (see Figure 3 (a)) is subdivided into overlapping regions by sliding a 82×82 square window across the image in question (see Figure 3 (b)). Each sub-image in the training and validation set is manually annotated as either positive (foreground) or negative (background). The positive training sub-images (see Figure 4) are representative of the foreground and typically forms part of the ear shell, while the negative sub-images (see Figure 5) represent the background and typically contains homogeneous skin, hair or jewellery. The objective of the CNN is to classify each patch within a test image as either foreground or background (see Figure 6).

The CNN is trained by employing stochastic gradient descent with momentum (SGDM) [16]. The weights of the first convolutional layer are initialised using normally distributed random numbers. The proposed CNN is trained from scratch. No fine-tuning of an existing pre-trained network (transfer learning) is conducted. After each epoch, the accuracy of the network is gauged by employing a validation set in order to avoid overfitting [17]. Morphological closing is subsequently applied to the resulting binary image (see Figure 6 (b) and (d)) in order to reduce noise and render the detected foreground boundaries more regular (see Figure 7).

B. Contour detection

After a Gaussian filter is applied to an input ear image (see Figure 8 (a)), a smoothed version (see Figure 8 (b)) is obtained. Canny edge detection is subsequently performed on the preprocessed image in Figure 8 (b), which results in a binary edge image (see Figure 9 (a)). Morphological dilation is applied in order to connect disconnected contours and remove noise (see Figure 9 (b)). Finally, the manually selected or automatically detected ROI is employed as a mask for the purpose of removing all of the edges not associated with ear contours, followed by the removal of the remaining small connected components (see Figure 10).

In the case of the IIT Delhi ear database, the same protocol (as the one for the AMI ear database) has been followed, except for the fact that the image borders are also cleared during the ROI-masking. In Figure 11 the detected prominent contours are presented for the IIT Delhi ear database.

C. Feature extraction and normalisation

Feature vectors are extracted from the prominent contours by applying the DRT to the binary edge image. The DRT is obtained when projections of an image are calculated

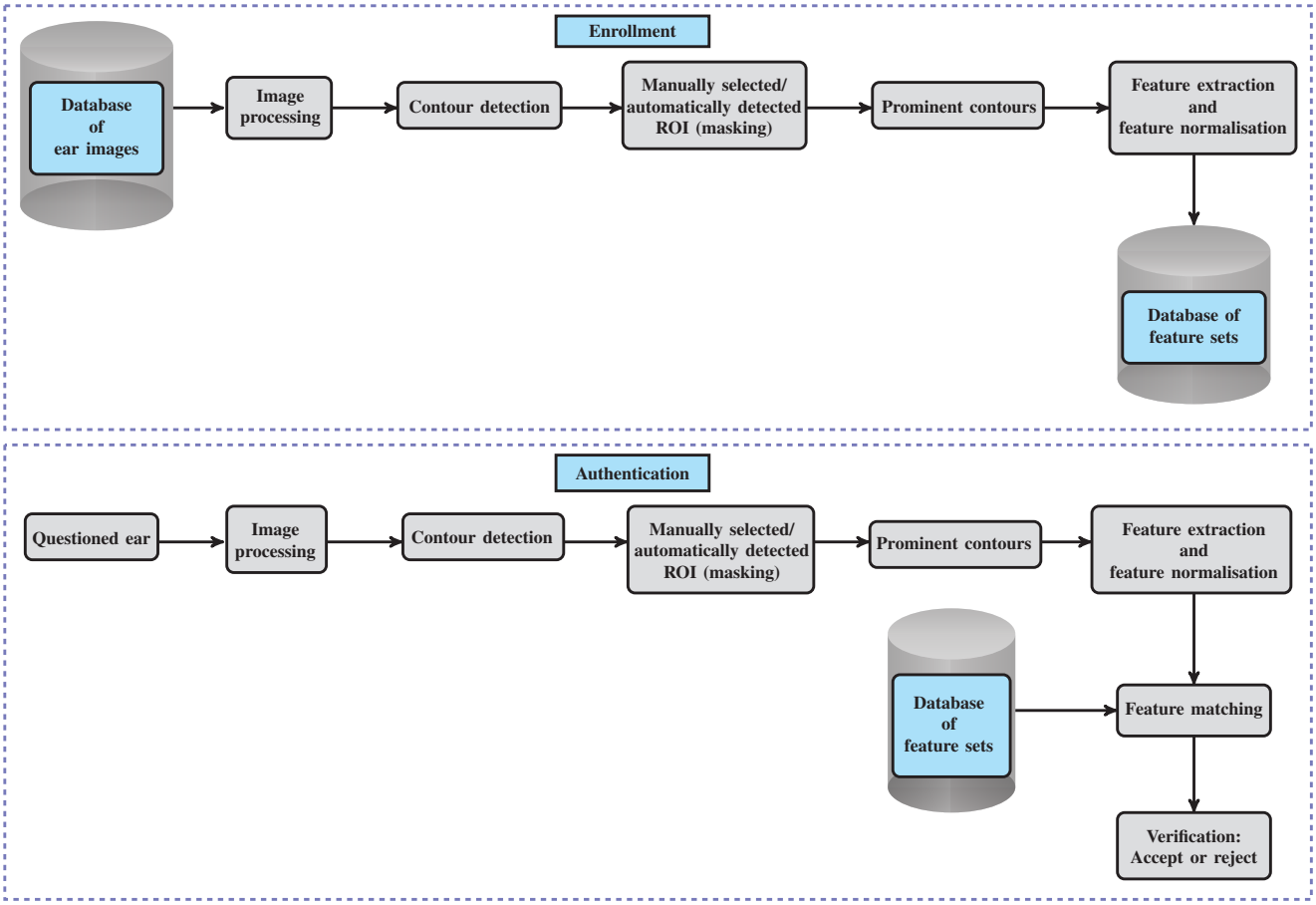


Fig. 1: Schematic representation of the system design.

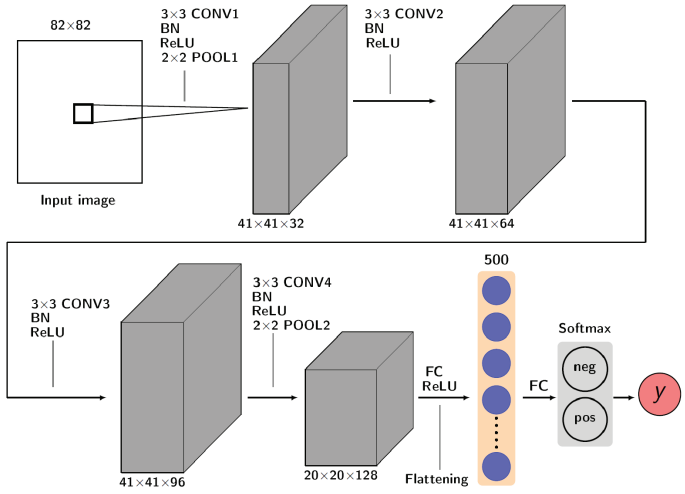
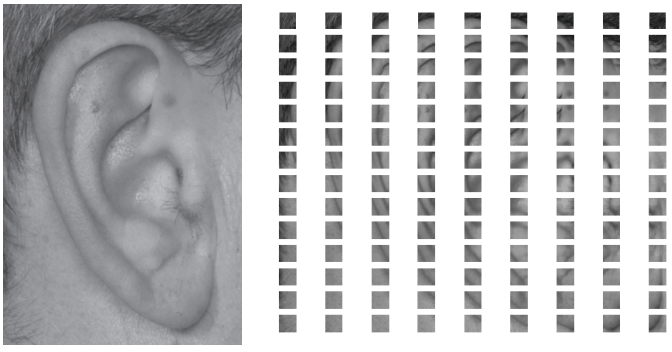


Fig. 2: The proposed CNN architecture for the purpose of automatically detecting a suitable ROI within an ear image.



(a) (b)

Fig. 3: (a) An example of a grey-scale image of size 702×492 pixels from the AMI ear database. (b) Example of the image depicted in (a) after being partitioned into 126 overlapping 82×82 sub-images.

from equally distributed angles within the interval $\theta \in [0°, 180°)$ [18]. The DRT of the binary image $I(m, n)$ of size $M \times N$ pixels containing the prominent contours associated with the ear shell can be expressed as follows

$$R_j = \sum_{i=1}^{MN} \delta_{ij} I_i, \quad \text{for } j = 1, 2, \ldots, \beta\Theta, \tag{1}$$

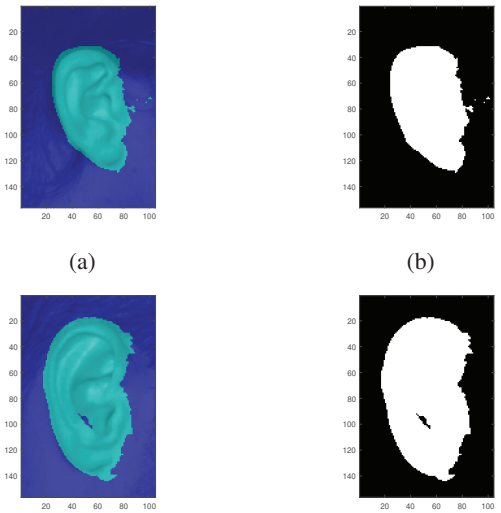
where R_j denotes the jth beam-sum which constitutes the cumulative intensity of the pixels that overlap with the jth beam, β denotes the number of non-overlapping beams per

(a) (b) (c)

Fig. 4: Examples of positively labelled sub-images of size 82×82 pixels.

(a) (b) (c)

Fig. 5: Examples of negatively labelled sub-images of size 82×82 pixels.



(a) (b)

(c) (d)

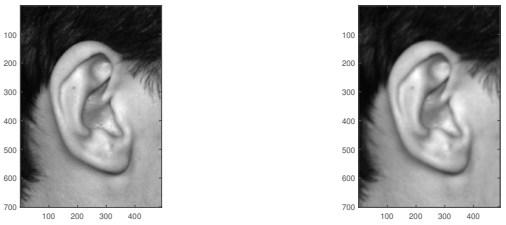
Fig. 6: (Left) The probability that a sub-image belongs to the foreground (contains contours associated with the shell of an ear) is represented by a shade of blue. (Right) Binary versions of the corresponding images on the left, after a suitable threshold has been applied.

angle, Θ represents the total number of angles and δ_{ij} denotes the weight indicative of the contribution of the *i*th pixel towards the *j*th beam-sum. A detailed description of the theory and implementation of the DRT can be found in [18].

Examples of contour images and their corresponding DRTs within the context of the AMI and IIT Delhi ear databases are depicted in Figures 12 and 13 respectively. In order to ensure translation and scale invariance, the zero-valued components are removed from each projection profile, after which the

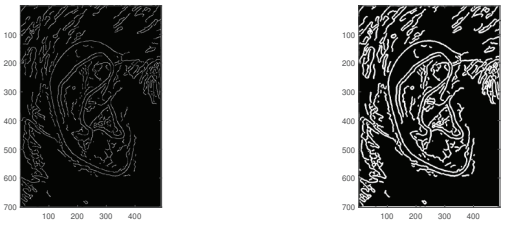
(a) (b)

Fig. 7: The automatically detected ROIs after a morphological closing operation has been applied to the binary images on the right of Figure 6.



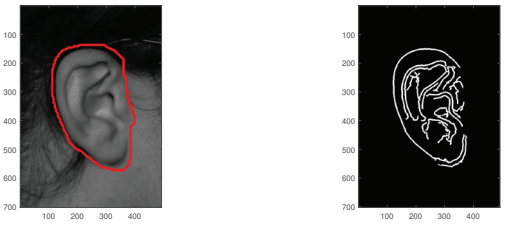
(a) (b)

Fig. 8: (a) Input image from the AMI ear database. (b) Smoothed version of the image on the left after the application of a Gaussian filter.



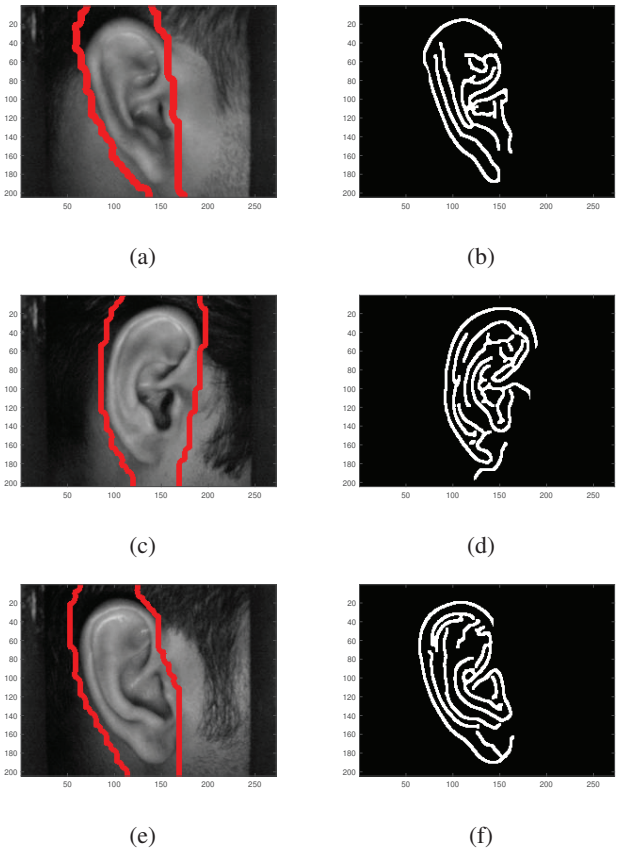
(a) (b)

Fig. 9: (a) Original edge map within the context of the AMI ear database. (b) Dilated edge image corresponding to the map on the left.



(a) (b)

Fig. 10: AMI ear database. (a) The ROI, which is automatically detected through deep learning, is enclosed by the red boundary. (b) Prominent contours after ROI-masking and the removal of small connected components.

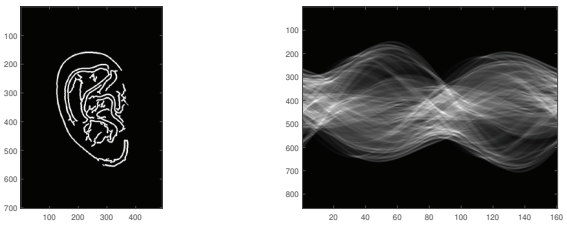


(a) (b)

(c) (d)

(e) (f)

Fig. 11: IIT Delhi ear database. (Left) These images are associated with three different individuals. The boundaries of the respective automatically detected ROIs are indicated in red. (Right) Detected prominent contours associated with the images on the left after the border has been cleared and small connected components have been removed.



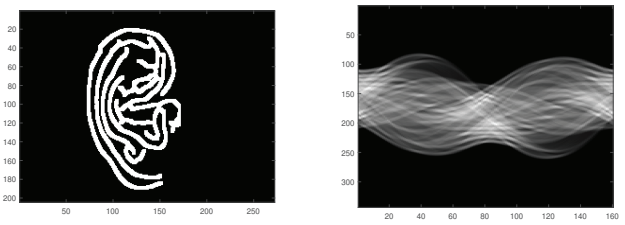
(a) (b)

Fig. 12: AMI ear database. (a) Prominent contours. (b) DRT image (where each column represents a projection profile) corresponding to the contour image on the left.

dimension of each projection profile is adjusted to a predefined value of 160 through linear interpolation. The DRT image intensities are also normalised in such a way that the standard deviation across all features equals one (see Figure 14).

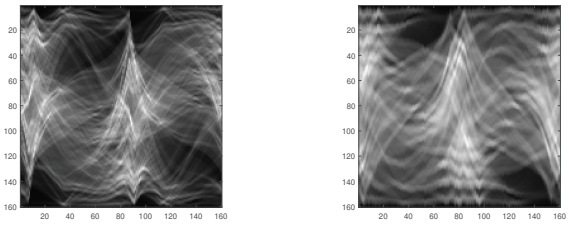
D. Feature matching and verification

The dissimilarity between two feature sets is quantified by the average Euclidean distance between the corresponding



(a) (b)

Fig. 13: IIT Delhi ear database. (a) Prominent contours. (b) DRT image (where each column represents a projection profile) corresponding to the contour image on the left.



(a) (b)

Fig. 14: (a) Normalised feature set that corresponds to the DRT image in Figure 12 (b). (b) Normalised feature set that corresponds to the DRT image in Figure 13 (b).

normalised feature vectors. The Euclidean distance between a normalised questioned and training feature vector, denoted by x and y respectively, is calculated as follows

$$\begin{aligned} D_{\text{Eucl}}(x,y) &= \sqrt{(x-y)'(x,y)} \\ &= \sqrt{(x_1-y_1)^2+\cdots+(x_d-y_d)^2}. \end{aligned} \tag{2}$$

In order to ensure rotational invariance, the normalised feature vectors associated with a questioned sample are iteratively shifted (with wrap-around) with respect to those belonging to a template. The alignment is deemed optimal when the average Euclidean distance is a minimum.

A questioned sample is compared to a reference sample (known to belong to the claimed individual), as well as to samples belonging to **other** (ranking) individuals. The dissimilarity between a questioned ear and the reference ear, as well as the respective dissimilarities between the questioned ear and those belonging to the ranking individuals are placed in a list, with the smallest dissimilarity at the top of the list and the largest dissimilarity at the bottom of the list. Verification is subsequently based on the relative position (ranking) of the dissimilarity associated with the reference ear in the list.

III. EXPERIMENTS

A. Data

Experiments are conducted on (1) the AMI ear dataset and (2) the IIT Delhi ear dataset. The AMI ear database consists of 700 images from 100 individuals. For each individual,

six images of the right and one image of the left ear are available and each ear image was captured at a resolution of 702×492 pixels. The IIT Delhi ear database consists of 375 images that belong to 125 individuals, i.e. three images per individual. Each of these images has a resolution of 272×204 pixels. Figures 15 and 16 depict samples of ear images from both databases.

(a) (b) (c)

Fig. 15: Examples of images from the AMI ear database [19].

(a) (b) (c)

Fig. 16: Examples of images from the IIT Delhi ear database [20].

B. Protocol

In this study three main experiments are conducted to investigate the proficiency of the proposed semi-automated and fully automated ear-based authentication systems. The experimental protocols are dichotomized as follows:

- **Experiment 1.** This experiment investigates the proficiency of the proposed *semi-automated* ear-based authentication system where the ROI is *manually* specified. This experiment is further dichotomized into two sub-experiments, i.e. Experiment 1A and Experiment 1B, which respectively considers so-called "Rank-1" and "Optimal ranking" scenarios as explained later in this section.
- **Experiment 2.** This experiment investigates the proficiency of the proposed *automated* ROI detection algorithm.
- **Experiment 3.** This experiment investigates the proficiency of the proposed *fully automated* ear-based authentication system, in which case the ROI is *automatically* detected through deep learning. Similar to Experiment 1, this experiment is further dichotomized into two sub-experiments, i.e. Experiment 3A and Experiment 3B.

It is assumed that only *one* positive sample is available for each individual enrolled into the system which serves as a reference sample for the corresponding individual during template matching. A k-fold cross-validation protocol is employed for each experiment as outlined below:

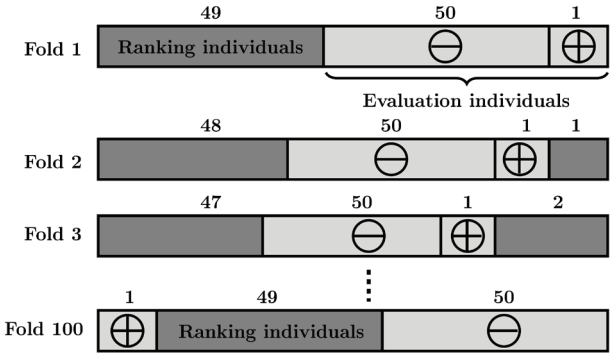


Fig. 17: Conceptualisation of the data partitioning protocol within the context of Experiment 1A, for the AMI ear database. Within each fold, 49 templates associated with 49 ranking individuals constitute the ranking set (dark gray), while three images associated with each of the respective 51 evaluation individuals constitute the evaluation set (light gray). One of the aforementioned evaluation individuals (⊕) constitutes the claimed individual. Technically, one image associated with the claimed individual is also employed for ranking purposes.

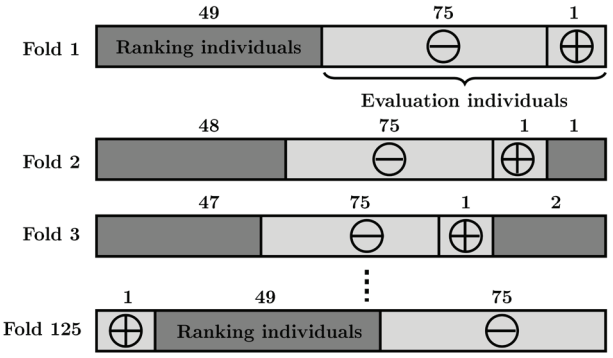


Fig. 18: Conceptualisation of the data partitioning protocol within the context of Experiment 1A, for the AMI ear database. Within each fold, 49 templates associated with 49 ranking individuals constitute the ranking set (dark gray), while three images associated with each of the respective 51 evaluation individuals constitute the evaluation set (light gray). One of the aforementioned evaluation individuals (⊕) constitutes the claimed individual. Technically, one image associated with the claimed individual is also employed for ranking purposes.

Experiment 1A (Rank-1 scenario): In this scenario a questioned ear is only accepted as authentic when the distance associated with the reference sample belonging to the claimed individual is the smallest, in which case the questioned ear has a ranking of one. This is referred to as the *rank-1 scenario*. For this experiment both of the ear databases are partitioned into two sets, that is the evaluation and ranking sets. For the AMI ear database, a 100-fold cross-validation procedure is conducted as depicted in Figure 17. A similar 125-fold cross-validation protocol is employed within the context of the IIT Delhi ear database (see Figure 18). The proposed data partitioning protocol for the evaluation individuals, within the context of the Rank-1 scenario and the AMI ear database is depicted in Figure 19. A similar data partitioning protocol is followed within the context of the evaluation individuals for the IIT Delhi ear database.

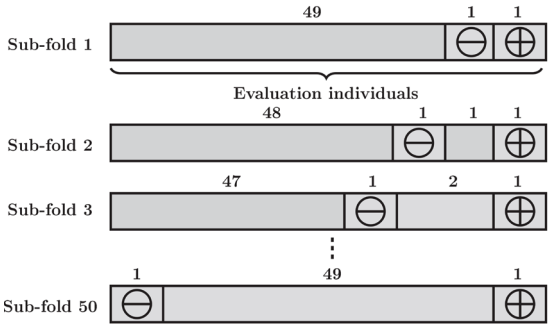


Fig. 19: Conceptualisation of the data partitioning protocol for the evaluation individuals within the context of Experiment 1A and the AMI ear database.

Experiment 1B (Optimal ranking scenario): In this scenario, the system is rendered more lenient such that a questioned ear is accepted when it has a ranking that is better than or equal to a very specific optimal ranking, which may be greater than one. The optimal ranking is estimated by considering a suitable data partitioning protocol. This is referred to as the *optimal ranking scenario*. Within the context of this experiment both of the ear databases are partitioned into a ranking set, an optimisation set and an evaluation set. The data partitioning and cross-validation protocol for both databases is presented in Figure 20.

As is the case for Experiment 1A, a 100-fold and 125-fold cross validation procedure are conducted for the AMI and IIT Delhi ear databases respectively. For a specific fold, cross-validation is conducted across the respective optimisation individuals according to the protocol depicted in Figure 19. The estimated optimal ranking based on *both* the average error rate (AER) and the equal error rate (ERR) is then employed to authenticate the ears associated with the evaluation individuals.

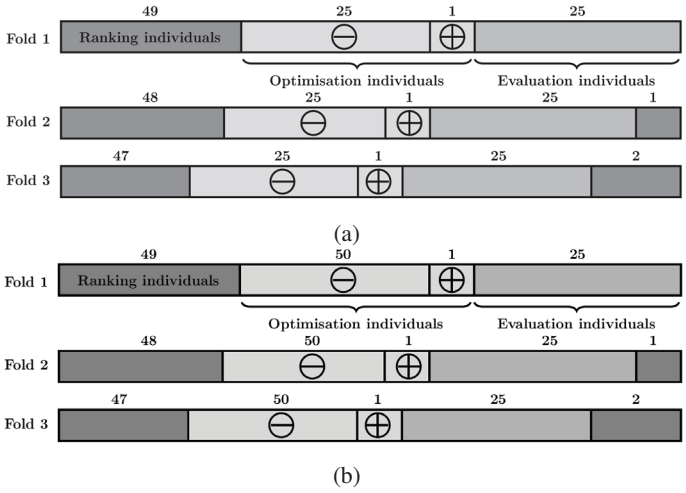


Fig. 20: The first three (out of a total of 100 and 125) folds of the proposed data partitioning and cross validation protocol within context of Experiment 1B for (a) the AMI ear database and (b) the IIT Delhi ear database.

Experiment 2 (Automated ROI detection): In this experiment the manually selected (specified) ROI serves as a ground truth for evaluating the proposed automated ROI-detection protocol. For both of the databases, the data is partitioned as depicted in Figure 21.

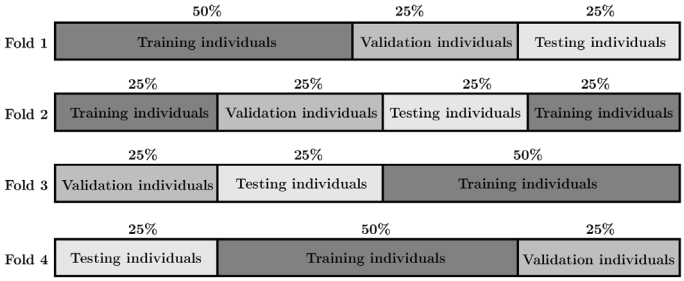


Fig. 21: Conceptualisation of the 4-fold cross validation protocol within the context of Experiment 2 for both the AMI and IIT Delhi ear databases.

Experiment 3 (Fully automated ear-based authentication): This experiment evaluates the proficiency of the proposed fully automated ear-based authentication system, where a suitable ROI is automatically detected through deep learning. Both of the ear databases are partitioned into four sets, that is a training set, a validation set, a ranking set and an evaluation set. Within the context of this experiment both the "Rank-1" and "Optimal ranking" scenarios are investigated in Experiment 3A and Experiment 3B, respectively.

C. Results

In this section, the performance of the proposed systems is reported and a comprehensive analysis of the results is presented. Table I lists the statistical measures employed for the purpose of quantifying the proficiency of the proposed systems. The statistical measures employed in this study constitute the most frequently utilised performance parameters for the purpose of evaluating ear-based biometric systems and include the accuracy (ACC), the false acceptance rate (FAR) and the false rejection rate (FRR) [21]. The results presented in Tables II, III, IV, V, and VI constitute *averages* across the relevant folds.

TABLE I:
THE STATISTICAL PERFORMANCE MEASURES EMPLOYED. THE NUMBER OF TRUE POSITIVES, FALSE POSITIVES, TRUE NEGATIVES, AND FALSE NEGATIVES ARE DENOTED BY TP, FP, TN, AND FN, RESPECTIVELY.

Performance measure	Definition
False acceptance rate (FAR)	FP/(FP+TN)
False rejection rate (FRR)	FN/(FN+TP)
Average error rate (AER)	(FAR+FRR)/2
Equal error rate (ERR)	FAR $\approx$ FRR
Precision (PRE)	TP/(TP+FP)
Recall (REC)	TP/(TP+FN)
Accuracy (ACC)	(TP+TN)/(TP+FN+FP+TN)
F_1 score	2 * PRE * REC/(PRE+REC)

Experiment 1A (Rank-1 scenario): The results are presented in Table II. It is clear that the proposed semi-automated system is more proficient in the case of the AMI ear database. This is presumably due to the fact that these images have a higher resolution.

TABLE II:
THE RESULTS FOR THE PROPOSED SEMI-AUTOMATED EAR-BASED AUTHENTICATION SYSTEM WITHIN THE CONTEXT OF THE RANK-1 SCENARIO.

Performance measure	AMI ear database	IIT Delhi ear database
FAR	0.04%	0.18%
FRR	4.76%	13.00%
AER	2.40%	6.59%
ACC	95.24%	89.12%

Experiment 1B (Optimal ranking scenario): The AER and EER were investigated as optimisation criteria for selecting the optimal ranking. The results are presented in Table III. For the AMI and IIT Delhi ear databases, only questioned ear images with a ranking of 5 (or better) and a ranking of 7 (or better) are accepted respectively.

TABLE III:
THE RESULTS FOR THE PROPOSED *OPTIMISED* SEMI-AUTOMATED EAR-BASED AUTHENTICATION SYSTEM WITHIN THE CONTEXT OF (A) THE AMI DATABASE (WITH AN OPTIMAL RANKING OF 5) AND (B) THE IIT DELHI DATABASE (WITH AN OPTIMAL RANKING OF 7).

Performance measure	Rank-5
FAR	1.75%
FRR	0.45%
AER	1.10%
ACC	99.20%

(a)

Performance measure	Rank-7
FAR	5.40%
FRR	3.63%
AER	4.52%
ACC	96.06%

(b)

Experiment 2 (Automated ROI detection): In Table IV the results for the AMI and IIT Delhi ear databases are presented. The precision, recall, accuracy, and F_1 score are employed as performance measures. In order to visually compare the manually selected and automatically detected ROIs, a few examples within the context of the AMI and IIT Delhi ear databases are presented in Figures 22 and 23 respectively.

Experiment 3A (Rank-1 scenario): The results for the proposed fully automated system within the context of the Rank-1 scenario are presented in Table V. The low FAR and high FRR are not unexpected.

Experiment 3B (Optimal ranking scenario): When optimal rankings (which do not necessarily correspond to a ranking of one) are investigated within the context of the proposed fully automated system, a significant improvement in proficiency is achieved (see Table VI). Only questioned ear images with a ranking of 7 (or better) and a ranking of 10 (or better) are accepted within the context of the AMI and IIT Delhi ear databases respectively. A similar improvement in proficiency is evident when the results presented in Table III

TABLE IV:
RESULTS FOR THE PROPOSED AUTOMATED ROI DETECTION PROTOCOL.

Performance measure	AMI ear database	IIT Delhi ear database
PRE	80.30%	70.26%
REC	90.88%	81.86%
ACC	91.01%	87.93%
F_1	87.66%	73.40%

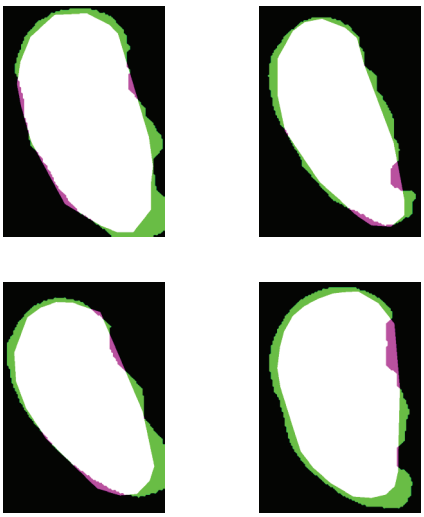


Fig. 22: Examples of ear images from the **AMI ear database** for the purpose of comparing the manually selected and automatically detected ROIs. The manually selected (specified) ROI serves as a ground truth for evaluating the proposed automated ROI detection protocol. The true positive, true negative, false positive and false negative pixels are depicted in white, black, green and pink respectively.

Fig. 23: Examples of ear images from the **IIT Delhi ear database** for the purpose of comparing the manually selected and automatically detected ROIs. The manually selected (specified) ROI serves as a ground truth for evaluating the proposed automated ROI detection protocol. The true positive, true negative, false positive and false negative pixels are depicted in white, black, green and pink respectively.

are compared to those presented in Table II within the context of the proposed semi-automated system.

TABLE V:
THE RESULTS FOR THE PROPOSED FULLY AUTOMATED EAR-BASED AUTHENTICATION SYSTEM WITHIN THE CONTEXT OF THE RANK-1 SCENARIO.

Performance measure	AMI ear database	IIT Delhi ear database
FAR	1.12%	2.45%
FRR	20.50%	39.25%
AER	10.81%	20.85%

TABLE VI:
THE RESULTS FOR THE PROPOSED *OPTIMISED* FULLY AUTOMATED EAR-BASED AUTHENTICATION SYSTEM WITHIN THE CONTEXT OF (A) THE AMI DATABASE (WITH AN OPTIMAL RANKING OF 7) AND (B) THE IIT DELHI DATABASE (WITH AN OPTIMAL RANKING OF 10).

Performance measure	Rank-7
FAR	3.11%
FRR	10.23%
AER	6.67%

(a)

Performance measure	Rank-10
FAR	5.38%
FRR	15.45%
AER	10.46%

(b)

Table VII places the proficiency of the proposed *optimised* semi-automated ear-based authentication system into perspective by comparing it to a number of recently developed systems that were also evaluated on either the AMI or IIT Delhi ear databases.

TABLE VII:
A SUMMARY OF EXISTING FEATURE EXTRACTION AND MATCHING TECHNIQUES FOR EAR-BASED BIOMETRIC AUTHENTICATION, AND THE REPORTED PERFORMANCE RATES WITHIN THE CONTEXT OF THE AMI AND IIT DELHI EAR DATABASES.

Publication	Feature extraction technique	Feature matching technique	AMI (%)	IIT Delhi (%)
Accuracies				
This paper	DRT	Euclidean distance	99.20	96.06
[8]	LBPs and Laplacian filter	KNN classifier	80	...
[4]	Fusion of features describing the shell and tragus	Euclidean distance	100	99.97
Recognition rates				
[6]	SIFT	A minimum distance classifier	100	95.20
[5]	Geometrical features describing the outer helix	Euclidean distance	...	99.6

D. Software and hardware employed

The systems proposed in this paper were developed in MATLAB™ (versions R2017b and R2018a). The following toolboxes were employed:

- Image Processing Toolbox™ (version R2017b);
- Neural Network Toolbox™ (version R2018a); and
- Statistics and Machine Learning Toolbox™ (version R2018a).

The algorithms were implemented on an 8*th* Generation Intel® Core™ i5 workstation with 8 GB RAM.

E. Conclusion

In the case of the proposed semi-automated system AERs of 2.4% and 6.59% are reported for the AMI and IIT Delhi ear databases respectively within the context of the Rank-1 scenario. These AERs are reduced to 1.10% and 4.52% respectively by employing optimal rankings. Accuracies of 91% and 88% are reported for the proposed CNN-based ROI detection protocol within the context of the AMI and IIT Delhi ear databases respectively. Within the context of the fully automated system AERs of 10.81% and 20.85% are reported for the AMI and IIT Delhi ear databases respectively within the context of the Rank-1 scenario. These AERs are significantly improved upon to 6.67% and 10.46% respectively by employing optimal rankings.

The proficiency of the proposed fully automated end-to-end system, in which the ROI is automatically detected, followed by feature extraction, feature matching, and verification is significantly lower than that of the semi-automated system in which case the ROI is manually specified for both the rank-1 and optimal ranking scenarios.

F. Contribution

The semi-automated and fully automated systems proposed in this paper employ an ensemble of pattern recognition techniques that has not been employed for the purpose of ear-based biometric authentication on any previous occasion and may therefore be considered *novel*. It is therefore reasonable to assume that either of the aforementioned systems will be *complementary* to any existing state-of-the-art system that invariably extracts *different* features or employ *different* feature matching techniques. It is therefore very likely that, when any of the proposed systems is *combined* with an existing system of comparable proficiency (see Table VII), a superior combined performance will be attained.

IV. Future work

Avenues for future research include an investigation into the feasibility of an end-to-end deep learning-based approach for ear-based biometric authentication, as well as an investigation into the feasibility of another machine learning-based approach, like a suitable support vector machine, for the second part of the fully automated system proposed in this paper. The proposed semi-automated and fully automated systems should also be evaluated on other databases that may become publicly available in the near future.

REFERENCES

[1] L. Yuan, and Z. Mu, "Ear recognition based on Gabor features and KFDA," *The Scientific World Journal*, 2014.

[2] J. F. Vélez, Á. Sánchez, B. Moreno, and S. Sural, "Robust ear detection for biometric verification," *IADIS International Journal on Computer Science and Information Systems*, vol. 8, no. 1, pp. 31-46, 2013.

[3] R. N. Othman, F. Alizadeh, and A. Sutherland, "A novel approach for occluded ear recognition based on shape context," in *Proceedings of the International Conference on Advanced Science and Engineering* (ICOASE), 2018, pp. 93-98, IEEE.

[4] K. Annapurani, M. A. K. Sadiq, and C. Malathy, "Fusion of shape of the ear and tragus–a unique feature extraction method for ear authentication system," *Expert Systems with Applications*, vol. 42, no. 1, pp. 649-656, 2015.

[5] I. Omara, F. Li, H. Zhang, and W. Zuo, "A novel geometric feature extraction method for ear recognition," *Expert Systems with Applications*, 2016, vol. 65, pp. 127-135.

[6] A. S. Anwar, K. K. A. Ghany and H. ElMahdy, "Human ear recognition using SIFT features," *Third World Conference on Complex Systems (WCCS)*, 2015, pp. 1-6, IEEE.

[7] P. L. Galdámez, A. G. Arrieta, and M. R. Ramón, "Ear recognition using a hybrid approach based on neural networks," in *Proceedings of the 17th International Conference on Information Fusion*, (FUSION), 2014, pp. 1-6, IEEE.

[8] S. M. Jiddah and K. Yurtkan, "Fusion of geometric and texture features for ear recognition," in *Proceedings of the 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies* (ISMSIT), 2018, pp. 1-5, IEEE.

[9] D. Querencias-Uceta, B. Ríos-Sánchez, and C. Sánchez-Ávila, "Principal component analysis for ear-based biometric verification," in *Proceedings of the International Carnahan Conference on Security Technology* (ICCST), 2017, pp. 1-6, IEEE.

[10] T. Ying, Z. Debin, and Z. Baihuan, "Ear recognition based on weighted wavelet transform and DCT," in *Proceedings of the 26th Chinese Control and Decision Conference*, (CCDC), 2014, pp. 4410-4414, IEEE.

[11] S. Dodge, J. Mounsef, and L. Karam, "Unconstrained ear recognition using deep neural networks," *IET Biometrics*, vol. 7, no. 3, pp. 207-214, 2018.

[12] U. Kacar and M. Kirci, "ScoreNet: deep cascade score level fusion for unconstrained ear recognition," *IET Biometrics*, vol. 8, no. 2, pp. 109-120, 2018.

[13] 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.

[14] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, "Deep learning", vol. 1, no. 2, Cambridge: MIT press, 2016.

[15] J. Wu, "Introduction to convolutional neural networks", *National Key Lab for Novel Software Technology*, Nanjing University, China, vol. 5, p.23, 2017.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.

[18] J. Coetzer, B. M. Herbst and J. A. du Preez, "Offline signature verification using the discrete radon transform and a hidden Markov model," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 4, pp. 559-571, 2004, Special issue: Biometric Signal Processing.

[19] E. Gonzalez, L. Alvarez, and L. Mazorra, "AMI ear database," 2012, http://ctim.ulpgc.es/research_works/ami_ear_database/

[20] A. Kumar, "IIT Delhi ear database version 1.0," 2007, http://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Ear.htm

[21] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, "A survey on ear biometrics," *ACM computing surveys (CSUR)*, vol. 45, no. 2, pp. 1-35, (2013).

Aviwe Kohlakala is a Ph.D. student in the Department of Mathematical Sciences at Stellenbosch University. She received a B.Sc. in Mathematics and Applied Mathematics from the University of South Africa in 2015 and an M.Sc. in Applied Mathematics from Stellenbosch University in 2019. Her research interests include machine learning, deep learning, biometric authentication and pattern recognition.

Johannes Coetzer was born in Bloemfontein, South Africa in 1971. He received an M.Sc. in Applied Mathematics from the University of the Free State in 1996 and a Ph.D. in Applied Mathematics from Stellenbosch University in 2005. From 1997 to 1998, he was a Junior Lecturer, and from 1999 to 2001, a Lecturer in Applied Mathematics at the University of the Free State. From 2002 to 2008, he was a Lecturer, and since 2009, a Senior Lecturer in Applied Mathematics at Stellenbosch University. His research interests include machine learning, biometric authentication and classifier combination.

CONTENTS PAGE