Robust multimodal biometric authentication on IoT device through ear shape and arm gesture

Feriel Cherifi, Kamal Amroun, Mawloud Omar

Multimedia Tools and Applications manuscript No.
(will be inserted by the editor)

Robust multimodal biometric authentication on IoT device through ear shape and arm gesture

Feriel Cherifi · Kamal Amroun · Mawloud Omar

Received: date / Accepted: date

Abstract Nowadays, authentication is required for both physical access to buildings and internal access to computers and systems. Biometrics are one of the emerging technologies used to protect these highly sensitive structures. However, biometric systems based on a single trait enclose several problems such as noise sensitivity and vulnerability to spoof attacks. In this regard, we present in this paper a fully unobtrusive and robust multimodal authentication system that automatically authenticates a user by the way he/she answers his/her phone, after extracting ear and arm gesture biometric modalities from this single action. To deal the challenges facing ear and arm gesture authentication systems in real-world applications, we propose a new method based on image fragmentation that makes the ear recognition more robust in relation to occlusion. The ear feature extraction process has been made locally using Local Phase Quantization (LPQ) in order to get robustness with respect to pose and illumination variation. We also propose a set of four statistical metrics to extract features from arm gesture signals. The two modalities are combined on score-level using a weighted sum. In order to evaluate our contribution, we conducted a set of experiments to demonstrate the contribution of each of the two biometrics and the advantage of their fusion on the overall performance of the system. The multimodal biometric system achieves an equal error rate (EER) of 5.15%.

Feriel Cherifi
Laboratoire d'Informatique Médicale (LIMED), Faculté des Sciences Exactes
Université de Bejaia, 06000 Bejaia, Algérie. E-mail: cherififeriel@gmail.com

Kamal Amroun
Laboratoire d'Informatique Médicale (LIMED), Faculté des Sciences Exactes
Université de Bejaia, 06000 Bejaia, Algérie. E-mail: kamalamroun@gmail.com

Mawloud Omar
LIGM, ESIEE Paris, Université Gustave-Eiffel, Noisy-le-Grand, France.
E-mail: mawloud.omar@univ-eiffel.fr

Keywords Biometrics · multimodal authentication · ear · arm gesture · score-level fusion.

1 Introduction

Nowadays, smartphones are used for various purposes, such as store sensitive personal data, banking transactions, etc. The safety of these devices is therefore fundamental. Security systems based on passwords, PINs and schemas are vulnerable to multiple attacks and cumbersome to remember. To remedy these problems, biometrics are gradually becoming part of our daily life as an alternative to ensure the security of the connected systems. Indeed, modern connected objects and especially smartphones offer a plethora of sensors embedded in a single omnipresent device (cameras, magnetometer, gyroscope, accelerometer, touch screen, microphones, proximity detector, etc.) that promotes the development of biometric authentication systems in real-world applications.

Biometrics are based on natural and innate characteristics of each individual. These characteristics can be physical, such as fingerprints, iris and face, or behavioral, such as gait and voice. However, biometric systems using a single biometric modality, called unimodal biometric systems, suffer from intraclass variability, interclass similarity, data quality, non-universality, sensitivity to noise, and other factors [1]. These problems can be reduced by implementing multimodal biometric systems using and combining several biometric traits which makes it possible to evaluate the system more precisely and significantly improves its performances. Indeed, the major challenge in designing secure, large-scale real-world deployable and marketable biometric systems is the unavailability of high quality data. This data are highly variable depending on different lighting conditions, distance to the camera, equipment, etc. Multimodality solves these problems by providing the system with more data to rely on. Note that, in order to take the full advantage of the multimodal approach, it is essential to implement a good method for fusing different sources of biometric information [2]. Indeed, authors of [2] , [3] and [4] have demonstrated that a multimodal system can be fooled by spoofing only a subset of the biometric traits.

In recent years, arm gesture and ear biometrics gained considerable attention and are regarded as promising biometric traits, but both of them have their strengths and weaknesses when used for unimodal biometric systems. Arm gesture signals are generated from sophisticated human body functions very difficult to counterfeit. In addition, unlike physiological biometric features such as face, they can only be captured in a living individual. Ear recognition is one of the active fields of biometric authentication that received much more attention. The ear has a rich and stable structure that is permanent with increasing age and does not suffer from facial expression variations at the same time [5]. Most of the proposed arm gesture and ear recognition techniques in literature obtained satisfactory recognition results under controlled

conditions. Though, their performances deteriorate in uncontrolled conditions. In real-world applications, arm gesture signals are affected by noise and ear images are affected by scaling, rotation, illumination and occlusions by hair strains or accessories and some of the proposed techniques need user interaction. Therefore, arm gesture and ear authentication systems for mobile devices in uncontrolled conditions still a challenging problem. Indeed, mobile devices degrade the biometric data quality, which reduces the accuracy of identification and limits their security applications. Consequently, efficient and robust authentication systems in real-world applications for both arm gesture and ear are still a challenging problem and need to be improved.

In this paper, we present an efficient and robust multimodal biometric authentication system based on the combination of ear as a physical biometric and arm gesture as a behavioral one both captured through user smartphone to restrict smartphone access only to the legitimate user. The proposed system is non-intrusive, the arm gesture signals and ear images can be collected in the same action of responding to a call. First, the two biometric traits (ear and arm gesture) are separately evaluated. Next, we combine them using score-level fusion to arrive at the final decision. In ear biometric, we first apply a pre-processing step which corrects non uniform illumination, suppresses noise and enhances contrast. The feature extraction process has been made locally using Local Phase Quantization (LPQ) [6], so that the system gets robust with respect to pose and illumination variations. In this work we propose a new method based on image fragmentation that is quite robust with respect to ear partial occlusions. This step allows us to ensure the robustness of the proposed system as the ear image is taken under relaxed and unconstrained settings. Indeed, occlusion is one of the biggest problems of ear recognition [7]. For a real application, the users cannot be asked to remove their earrings, piercings and hair each time they authenticate themselves. The classification of ear images is done using Euclidean distance. For arm gesture, we propose the use of a set of four statistical features to identify a user. The classification is done using random forest classifier. Fusion was done on score-level using weighted-sum which has many advantages with respect to implementation, computation and recognition performances. We evaluated the performance of the proposed system on our proprietary database comprising arm gesture signals and ear images collected under uncontrolled conditions collected from 7 volunteers. To enlarge the set of ear images we also used the Annotated Web Ears (AWE) database [8]. Also to enlarge the set of arm gesture we used the Hand Movement, Orientation, and Grasp (HMOG) database [9]. AWE and HMOG collect data from 100 subjects under uncontrolled conditions. In order to carry out the experiments on a larger database than ours, AWE and HMOG were mixed to create a virtual biometric database which makes a total of 107 subjects (100 from AWE and HMOG and 7 from ours). In this paper, we also did a comparative study with [10] which proposes a multimodal biometric system that uses the same biometric traits as ours. Experimental results confirm the effectiveness and robustness of the proposed system.

The significant specific contributions of this paper are as follows:

1. Proposal of an efficient, robust and non-intrusive multimodal authentication system combining arm gesture and ear to authenticate smartphone users.
2. Proposal of a new method based on image fragmentation to solve the occlusion problem in ear recognition systems.
3. Proposal of the use of four statistical metrics for feature extraction from inertial sensor signals.

All in all, the proposed multimodal biometric system successfully overcomes real-word application challenge and has advantages such as non-intrusive, low recognition error rates and strong security, even if under uncontrolled environments.

The rest of this paper is organized as follows. Section 2 is a review of previous works related to multimodal biometric systems. Section 3 describes the proposed system in detail, while Section 4 provides the results of the experiments conducted on the multibiometric database. Finally, conclusion and future directions of research are drawn up in Section 5.

2 Related work

The various sensors existing on intelligent mobile devices have enabled the implementation of different biometric authentication systems that have made it possible to circumvent the problems of traditional systems (PINs, passwords, etc.). However, the acquisition of biometric data in real environments exposes the data to degradation and presents challenges to the security community. To minimize the weaknesses of different biometric methods while enhancing the advantages of them, several multimodal systems combining different biometric features in mobile devices, especially for smartphones, were proposed in the literature. We will present some of them in this section.

To authenticate a smartphone user, Akhtar et al. [11] proposed a multimodal biometric system that combines face, hand micro-movements and touchstroke patterns of the user while he/she is entering a text-independent PIN of fixed length which achieves 3% EER obtained with 10 training samples. In [12], the authors combined electrocardiogram (ECG) signals and fingerprint that can be applied for airport traveler authentication at border controls. They achieve an EER equal to 0.46% and a FAR equal to 2.96 %. Face and voice were fused together in order to enhance access security to user's mobile device by Memon et al. [13] with an EER = 11.87%. The iris has also been combined with other biometric features on smartphone to overcome the challenges presented by this platform. Zhang et al. [14] have combined iris with periocular region which can be obtained simultaneously with the irises when using mobile devices. The results show an EER = 0.6%. Sequeira et al. [15] have been motivated by the raising interest in mobile biometrics applications. Moreover, by the increasing interest in multimodal biometrics to develop MobBIO, a multimodal database comprising face, iris and voice samples acquired in unconstrained conditions from 105 volunteers. Gofman et al. [16] have studied

the effects of face and voice sample quality on recognition accuracy in mobile devices. They have developed a multimodal biometrics, which integrate information from the face and voice using a novel score-level fusion scheme driven by the quality of the captured biometric samples. The experimental results have shown that the approach increases accuracy by 4.14% and 7.86% compared to using face and voice recognition individually, respectively. Face and voice were fused by Abozaid et al.[17] to enhance the Wireless Body Area Network (WBAN) security. Mahmoud et al. [18] have presented a multimodal biometric authentication method to confirm the identity of a person based on his face and iris features. Wu et al. [19] proposed LVID, a multimodal biometrics authentication system for smartphones, which combines the advantages of lip movements and voice.

Dhvani et al. [20] have used the Raspberry Pi as a low-cost, wireless and remote enrollment node to develop a multimodal authentication system combining fingerprints and facial images. The authentication was done by biometric service hosted on the cloud as a Software-as-a Service. To secure the transmission, the authors have applied the modified AES-256 (Advanced Encryption Standard). To develop an authentication system for the IoT in eHealth, Habib et al. [21] have proposed to combine: patient's physiological biometric, patient's behavioral biometric and patient's smart phone radio fingerprint. They have bound together the biometric modalities and radio fingerprinting technique as a unique identifier to not only authenticate the patient but also the device transmitting health parameters. Face and iris biometric traits, collected by a single high quality camera, were used by Macek et al. [22]. The image of the face was used to identify the user and the image of the iris was further used to verify the identity. Olazabal et al. [23] have used discriminant correlation analysis (DCA) to fuse features from face and voice and used the K-nearest neighbors (KNN) algorithm to classify the features. The approach was implemented on the Raspberry Pi IoT device. The results show that fusion increased recognition accuracy by 52.45% compared to using face alone and 81.62% compared to using voice alone. To authenticate a mobile device user, Gofman et al. [24] have proposed feature-level fusion of face and voice. They applied Discriminant Correlation Analysis (DCA) to the fusion of incompatible features and compared the identification accuracy and execution times of several classifiers. The experimental results have shown that the multimodal approach outperformed the unimodal face and voice approaches for all classifiers. The greatest improvement was achieved using the Support Vector Machines (SVMs). Teeth image and voice collected from smartphone were used by Kim et al. [25] to develop a multimodal biometric system. The performance evaluation was conducted using a database acquired in smartphone for 40 subjects. The experimental result have shown 8.59% of EER in case of teeth verification, 11.73% in case of voice verification and the multimodal speaker authentication result presented the 4.05% of EER. The authors of [26] have presented a multimodal biometric system for mobile authentication combining voice, face and chirography-based. The fusion strategy is based on modality score level fusion that takes into account a voice quality measure. An EER of

0.1% has been obtained for use in a quiet office, and an EER of 0.3-0.6% for use in a noisy cafeteria. The authors of [27] have shown the potential of multimodal biometrics for protecting consumer mobile devices from unauthorized access. They have implemented Proteus based on face and voice biometrics. They have compared Proteus recognition accuracy to unimodal systems based on face and voice biometrics (In both cases: score-level fusion and feature-level fusion.). For their experiments, they created a CSUF-SG5 homegrown multimodal database of face and voice samples collected from University of California, collected in a variety of real-world settings.

Arm gesture is a new biometric trait that has attracted the attention of researchers in recent years. Ear contour has been investigated for several decades as a biometrics trait, however it is still an open field of study due to the number of unresolved challenges. It presents such as a tendency of ears to continue growing linearly as we age and partial or total occlusion by hair, earrings, etc. This is why researchers have thought of combining the ear with other biometric modalities to compensate for these disadvantages. In [28], a study was presented on a multimodal biometrics, fusing ear and speech biometrics approach to authenticate smartphone users. Abate et al. [10] have presented a multibiometric system based on the ear and arm gesture when a smartphone user is responding to a phone call. The ear feature extraction was done using Local Binary Patterns (LBPs) algorithm [29] and they used Euclidean distance and Dynamic Time Warping (DTW) for features matching. A score-level weighted data fusion was exploited to fuse the two biometrics. They achieved 10.04% equal error rate (EER).

Table 1 summarizes the work presented in the related work. We have taken the best accuracy rates obtained by the authors in their tests. Note that the works are not directly comparable as the biometric traits and the datasets used by the authors are different.

3 Proposed method

Each time a user answers a call or makes a call, the sensors of the mobile device could record both the movement associated with the arm holding the device and the image of the ear. We propose to combine these two features to create a real-time multimodal biometric authentication system consisting essentially of three elements: ear subsystem; arm gesture subsystem; and the fusion subsystem. As shown in Figure 1, initially the two biometric traits are evaluated separately. The two authentication techniques are then combined.

In Table 2, we present the important notations used in this paper. In the following subsections, we detail each subsystem.

Table 1 Summary of the surveyed work on multimodal biometric authentication for mobile devices.

Reference	Biometric trait	Dataset size	Accuracy
[17]	Face and voice	100 subjects	EER = 0.62 %
[18]	Face and iris	106 subjects	Accuracy = 99 %
[19]	Lip movements and voice	104 subjects	Accuracy = 95%
[11]	Touch-stroke, phone-movement and face	95 subjects	EER = 3 %
[12]	ECG and fingerprint	73 subjects	EER = 0.46 %
[13]	Face and voice	54 subjects	EER = 11.87 %
[14]	Iris and periocular	630 subjects	EER = 0.6 %
[16]	Face and voice	54 subjects	EER = 14.56 %
[20]	Face and voice	Not stated	Not stated
[22]	Face and iris	500 subjects	FRR = 1 %
[23]	Face and voice	80 subjects	EER = 8.04 %
[24]	Face and voice	Not stated	EER = 20.59 %
[25]	Teeth image and voice	50 subjects	EER = 2.13 %
[26]	Chirography, face and voice	100 subjects	EER = 0.5%
[27]	Face and voice	54 subjects	EER = 2.14%
[28]	Ear and voice	50 subjects	Accuracy = 93.5 %
[10]	Ear and arm gesture	100 subjects	EER = 0.1 %

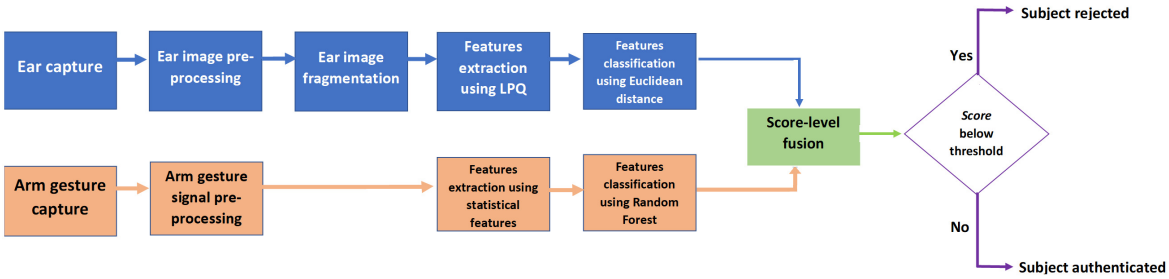

Fig. 1 Architecture of the proposed multimodal biometric authentication system based on ear images and arm gesture signals.

3.1 Ear subsystem

3.1.1 Ear acquisition and feature extraction

Ear image capture can be performed either contextually to a call, or deliberately for authentication purposes. To acquire the ear image, we record a video that is triggered when a call is received. The videos were recorded using the volunteers' smartphones, each volunteer used his/her own smartphone (e. g. Huawei GR3 2017, Sony Xperia P). We preferred to record a video instead of taking a single picture directly, because of the difficulty of choosing the ideal time to trigger the camera when answering a call. To detect the images that

Table 2 Notations.

Notation	Description
$\overrightarrow{A}, \overrightarrow{G}$	Acceleration and angular velocity vectors
$\|\overrightarrow{U}\|$	Magnitude of vector $\overrightarrow{U}$
$\overline{X}$	Mean of the sequence X
σ_X	Standard deviation of the sequence X
Med_X	Median of the sequence X
K_X	Kurtosis of the sequence X
$d(p,q)$	Euclidean distance between p and q
$score_{ear}$	Matching score of ear system
$score_{arm}$	Matching score of arm gesture system
$MULTI_{fusion}$	Final fused score
C	Accuracy of the system

contain the ear and to choose the image with the highest quality for feature extraction, we followed the same method proposed by [10].

In any biometric authentication system, the feature extraction phase is very important. Errors in this phase will propagate through the next phases which will cause the authentication system to fail. Many studies on ear feature extraction approaches have been suggested. For a detailed review of the 2D ear feature extraction approaches, we cordially refer the reader to [8]. However, their applications are restricted to controlled environments in laboratories, which slow down the deployment of ear-based authentication systems. Their performances deteriorate as soon as the images are affected by scaling, rotation, illumination and occlusions by hair strains or accessories. Ear recognition in unconstrained settings is still an unsolved problem and a challenging task. To meet this challenge, we first carry out a pretreatment on the original image to improve its quality before moving on to feature extraction. The pretreatment improves the visibility of image characteristics and details by enhancing image contrast and dynamics and also reduces noise, which mainly comes from the acquisition phase. First, the original image is normalized and converted into a gray level image, this will allow us to reduce the cost of memory and calculation. Then, the gray level image is filtered by a median filter to reduce noise. The histogram of the image is then equalized to improve contrast. Once the pretreatment is done, we apply feature extraction phase. In [8] the authors have presented an overview of the field of automatic ear recognition. They introduced to the research community a new, fully unconstrained dataset of ear images gathered "in the wild" called *Annotated Web Ears (AWE)*. They also conducted an objective ranking between 8 different techniques (LBP [29], LPQ [29], BSIF [30], POEM [31], HOG [32], Dense SIFT [33], RILPQ [34] and Gabor [35] features) for ear feature extraction made on 3 different databases (IITD II [36], USTB II [37] and AWE [8]). To take into account the results obtained on the various databases, we calculated the average of the EERs obtained for each technique on each database. LPQ achieves the lowest EER

= 14.8%. Based on these results we have chosen to use LPQ for ear feature extraction.

We will first present the two techniques LPQ and LBP:

- Local Phase Quantization (LPQ): In [6] the authors have proposed a new descriptor for texture classification that is robust to image blurring and invariant to uniform illumination changes. LPQ operates in the Fourier phase computed locally for a window in every image position. The LPQ value is computed for every pixel. Next, local histograms with 256 bins are computed within a sliding window. The window is moved with a certain overlap between two neighboring windows, in the horizontal and vertical directions over the image and concatenate the resulting local histograms [38].
- Local Binary Pattern (LBP): It is a texture descriptor for images which thresholds the neighboring pixels based on the value of the current pixel [39]. The original LBP operator assigns a label to every pixel of an image by thresholding the 3×3 neighborhoods of each pixel with the center pixel value and considering the result as a binary number. Then, the histogram of the labels can be used as a texture descriptor [40]. Recently, several LBP variants have been developed in order to improve the texture description [41].

To address the problem of occlusion, we propose a new, efficient and robust process based on image fragmentation that is applied before the execution of feature extraction algorithm (in our case LPQ) but the proposed fragmentation method can be applied with any other feature extraction method. The main steps of our method are:

- Step 1 : divide the images into 4 vertical sub-parts of equal size, as illustrated in the figure 2. The vertical sub-parts can be perceived as the internal and external parts of the ear.
- Step 2 : divide the images into 3 horizontal sub-parts of equal size. The horizontal sub-parts can be perceived as the up, middle and down parts of the ear. The number of sub-parts has been chosen to balance efficiency and minimum calculation cost. Other configurations, with a higher number of segments, generate a higher calculation cost. If we reduce the number of segments the system becomes ineffective against occlusion.
- Step 3 : calculate independently for each subpart $i/i \in \{1, \cdots, 7\}$ the LPQ feature vector v_i.
- Step 4 : fuse the results obtained from each subpart in a single vector p representing the biometric feature of the ear. The fusion is done according to the following formula:

$$p = v_1 \cdot v_2 \cdot v_3 \cdot v_4 \cdot v_5 \cdot v_6 \cdot v_7 + v_2 \cdot v_3 \cdot v_6 + v_2 \cdot v_3 \cdot v_4 \cdot v_5 \cdot v_6 + v_2 \cdot v_3 \cdot v_6 \cdot v_7. \quad (1)$$

The focus of this formula is to get the system robust with respect to hair and ear occlusion. Indeed, by analyzing the ear images we noticed several cases and explained below:

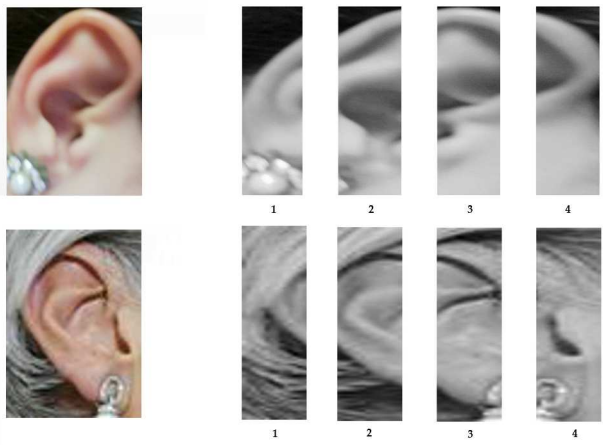

Fig. 2 In the first step we split the ear image in 4 vertical sub-parts of equal size indexed from 1 to 4. Image on top is a sample from our database, image on the bottom is a sample from the AWE database [8].

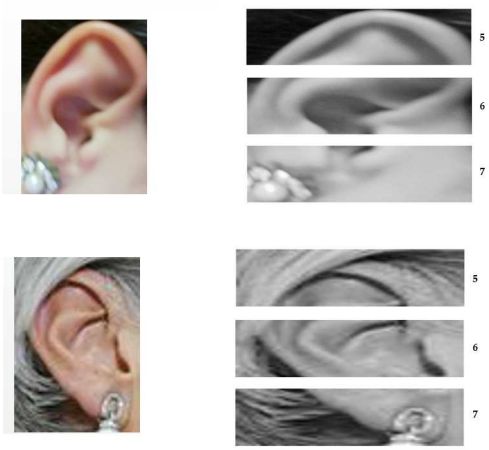

Fig. 3 In the second step, we split the ear image in 3 horizontal sub-parts of equal size indexed from 5 to 7.

1. In the case of hair occlusion, we have noticed that the subparts affected by this occlusion are 1, 4 and 5. These subparts as they are hidden by hair, their feature vectors v_1, v_4 *and* v_5 are useless, hence $v_2 \cdot v_3 \cdot v_6 \cdot v_7$, where we take into consideration only the v_i of the subparts that present useful information from the ear.
2. In the case of earrings occlusion, we have noticed that the subparts affected by this occlusion are 1 and 7. We therefore take into considera-

tion that the feature vectors of subparts where the ear is clearly visible, hence $v_2 \cdot v_3 \cdot v_4 \cdot v_5 \cdot v_6$.
3. In case of earrings and hair occlusion, subparts 1, 4, 5 and 7 are completely hidden and the ear is invisible, hence $v_2 \cdot v_3 \cdot v_6$.
4. Images have no occlusion. In this case we take into consideration the characteristic vectors of all sub-parts, hence $v_1 \cdot v_2 \cdot v_3 \cdot v_4 \cdot v_5 \cdot v_6 \cdot v_7$.

3.1.2 Ear classification

In classification phase, we perform a matching between the feature vector of the input image with those of the template collected during the enrollment phase. The classification is done by applying the Euclidean distance given by the equation 2:

$$d(p,q) = \sqrt{\sum_{i=0}^{n}(p_i - q_i)^2}, \tag{2}$$

where $p = (p_1, \cdots, p_n)$ and $q = (q_1, \cdots, q_n)$ are the two n-dimensional feature vectors. The matching score $d(p,q)$ represents the number of matched key-points between the two templates. The decision of acceptance or rejection of a person depends on this score. Positive authentication is declared when $d(p,q)$ is below a threshold. The threshold was chosen based on the experimental results presented in Section 4.1, the threshold chosen is the one that performs the lowest EER. The Euclidean distance is very simple and does not consume many resources so it is ideal for real-time applications.

3.2 Arm gesture subsystem

It has been reported in [10] and [42] that each user responses to a call in a unique way. This uniqueness of movement pattern increases the authentication accuracy on the one side and makes challenging for impostors to generate exactly the same movement patterns [10]. We take advantage of this uniqueness to authenticate the smartphone user. The proposed scheme collects data from multiple 3-dimensional smartphone sensors in the background and matches the query pattern with the pre-stored pattern to authenticate the smartphone owner. First, we collect the sensory data from the accelerometer and gyroscope sensors, then relevant features are extracted. The feature vectors are fed as input to the classifier for testing.

3.2.1 Arm gesture acquisition

In order to collect the arm gesture data for our analysis, we developed an Android application that collects the sensory data from the accelerometer and gyroscope sensors of the smartphone denoting the user motion when he/she

answers the phone. To ensure participants privacy, we did not collect any information that can be used to identify them. When the user answers the phone, the accelerometer and the gyroscope in the smartphone detect respectively, the acceleration values in meter per second square (m/s^2) and the change in the orientations in radians per second (rad/s) along x, y and z axes. For each sensory readings, we collect 3-dimensional values: $\overrightarrow{A} = (x_{acc}, y_{acc}, z_{acc})$ for accelerometer data, $\overrightarrow{G} = (x_g, y_g, z_g)$ for gyroscope, and their respective magnitudes $||\overrightarrow{A}|| = \sqrt{x_{acc}^2 + y_{acc}^2 + z_{acc}^2}$, $||\overrightarrow{G}|| = \sqrt{x_g^2 + y_g^2 + z_g^2}$. The magnitude has the advantage of making the inertial sensor data independent of device orientation.

The accelerometer and gyroscope sensors on smartphones are extremely sensitive and record not only minor hand shakes associated with hand movement, but also interference such as noise which corrupts useful information. Therefore, in order to be able to perform arm gesture recognition the removal of unwanted noise from the raw collected data is necessary. Along these lines, we clean the raw smartphone inertial sensor signals using an average smoothing filter of size 1×3.

3.2.2 Arm gesture feature extraction

Once the data has been filtered, the signals are segmented into fixed-size windows. For each window W_n and for each signal (accelerometer and gyroscope), we calculate separately for each axis (x, y, z and also for the magnitude) statistical features that can discriminate the legitimate user. We propose a set of four different features from the time domain, namely the Mean, Standard deviation, Median and Kurtosis. Table 3 provides the details of the selected features. We chose time domain features instead of frequency domain ones because their calculation is computationally cheaper and also for their performance. So for each window W_n, we get a feature vector of size 32 (4 features $\times$ 8 signals).

Table 3 List of the extracted features from all the dimensions of accelerometer and gyroscope.

Feature	Formula
Mean: $\overline{W_n}$	$\frac{1}{N}\sum_{i=1}^{N} W_n(i)$
Standard deviation: σ_{W_n}	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(W_n(i) - \overline{W_n})^2}$
Median: Med_{W_n}	midpoint of W_n
Kurtosis: K_{W_n}	$\frac{m_4}{(\sigma^2)^2}$, where m_4 is the fourth moment

To illustrate the process of feature extraction, we present in the figure 4 a sample data collected from the accelerometer in x, y *and* z direction, composed of 400 points. As a first step of the feature extraction process, we

segment the signal into fixed size windows (in the example we have 8 windows of one second). The next step consists to compute for each window W_n (in the exemple $n \in 1..8$) the statistical features: Mean, Standard deviation, Median and Kurtosis for the x, y *and* z directions. The same steps are applied on the gyroscope data.

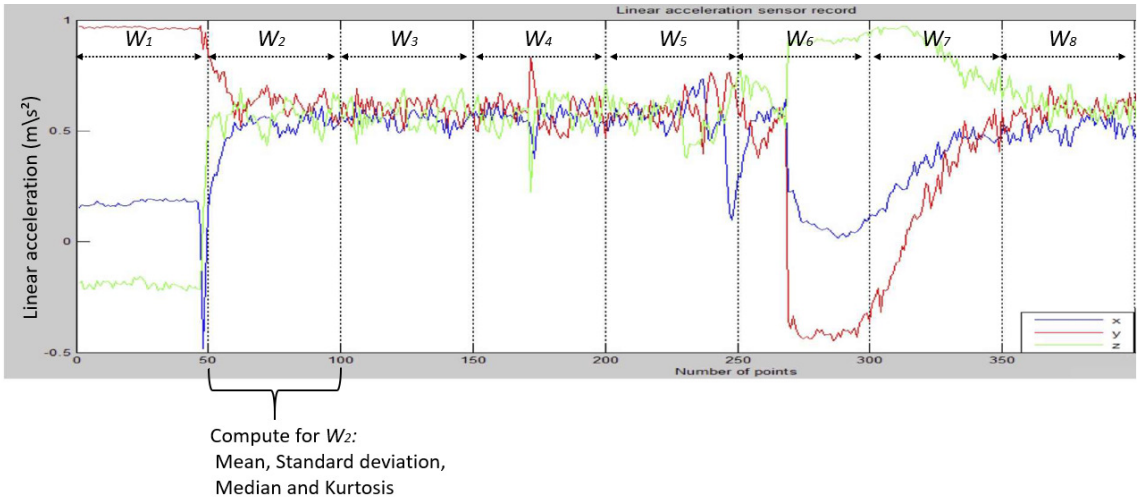

Fig. 4 Sample data of 500 points with a window of 1s.

3.2.3 Arm gesture classification

For the purpose of user recognition based on his/her arm gesture, we use Random Forest (RF) classifier. The fundamental idea behind a random forest is to combine many decision trees into a single model which makes it more robust overall predictions. The main reason for the selection of this classifier is its accuracy shown in the previous study [43]. Moreover, RF classifier is extremely quick and does not overfit and it is equally good for small and large databases.

3.3 Fusion

The combination of several biometric systems can be done at four different levels [44]: raw data-level fusion, feature-level fusion, score-level fusion, and decision-level fusion. Raw data and feature level fusion pose a number of constraints, such as data homogeneity that can only be met in very specific applications [44]. In our proposed system, we apply fusion at the score-level that can be applied to all types of systems. It is a simple and effective method. So, after

both ear and arm distances, $score_{ear}$ and $score_{arm}$ have been computed, we perform score normalization to transform the scores of the individual matchers into a common domain before combining them. Then, a weighted score $MULTI_{fusion}$ is estimated as follows:

$$MULTI_{fusion} = \alpha \cdot score_{ear} + \beta \cdot score_{arm}, \tag{3}$$

where α and β are the weights that characterize the contribution of each unimodal biometric system, ear and arm gesture in the performances of the overall multimodal system. Their values are estimated according to the experiment results expounded in Section 4.

The pseudo code 1 (**Algorithm 1**) summarizes all steps of our proposition.

Algorithm 1 Biometric authentication using ear and arm gesture

input: Ear image, Arm gesture signal
// I- Ear subsystem
// I-1- Ear image pre-processing
Image normalization;
Convert the image into gray level image;
Apply the median filter;
Histogram equalization.
// I-2- feature extraction LPQ
Divide the image into 4 verticale segments;
Divide the image into 3 horizontal segments;
Apply LPQ for each segment;
Fuse all segments.
// I-3- classification
Features classification using the Euclidean distance.
// II- Arm gesture subsystem
Data collection from accelerometer $\overrightarrow{M}$, gyroscope $\overrightarrow{G}$ in X, Y and Z directions; the collected data are of the form $\{m_x, m_y, m_z, g_x, g_y, g_z\} \in \mathbb{R}^6$.
// II-1- Arm gesture signal pre-processing
Add the magnitude $\|\overrightarrow{U}\|$ as a fourth dimension to all of these sensors;
Apply average smoothing filter;
For each dimension X, Y, Z and magnitude, segment the collected signal into fixed-size sliding window of size h.
// II-2- Arm gesture signal feature extraction
For each window, compute separately the Mean, Standard deviation, Median and Kurtosis.
// II-3- Arm gesture signal classification
Features classification using the Random Forest classifier.
// III- Fusion subsystem
$MULTI_{fusion} = \alpha \cdot score_{ear} + \beta \cdot score_{arm}$.

4 Experiments and results

In this section we describe the experiments accomplished to evaluate and measure the effectiveness of the proposed bimodal biometric authentication system. First, we present, separately, the evaluations of each biometric trait (arm gesture and ear), then we present the results of the fusion of these two traits to show and highlight the contribution of the multimodality in improving recognition rates. In our work, we measure the performance of our proposition using mainly Equal Error Rate (EER) which indicates that the proportion of false acceptances rate (FAR) is equal to the proportion of false rejections rate (FRR). The lower the EER value, the higher the accuracy of the biometric system.

4.1 Ear

To conduct our ear experimental tests we use our collected database in unconstrained conditions where we collect 10 images of 7 volunteers within an age ranging from 19 to 56 years. We also use the freely available unconstrained database of the Annotated Web Ears (AWE) proposed in [8] and downloaded from *http://awe.fri.uni-lj.si.* AWE contains a total of 1,000 annotated images from 100 distinct subjects. To demonstrate the robustness of the proposed method against occlusions we performed two kinds of experiments. In the first experiment we have split the database into two sub-sets the first sub-set (sub-set 1) contains only images that include occlusion either by accessories, hair, hearing aids etc. The second sub-set (sub-set 2) contains only images that do not have any occlusion, but the images contain blur, variations in rotation and illumination. Sample images of these subsets are presented in Figure 5. This experiment allows us to evaluate the real performances of the proposed system and know how good the recognition performance is when occlusions occur on ear images. The performances presented by the state of the art works are obtained on databases which are composed of images collected under controlled conditions and few images have occlusion, therefore the results obtained do not evaluate the true performances of these works and their robustness in the case of occlusions and real world applications.

The second experiment consisted of evaluating our proposed method, based on image fragmentation, on the two sub-sets (sub-set 1 and sub-set 2). In this experiment, we compared the performance achieved by LPQ that we adopted for feature extraction and LBP used by [10], in both cases: case 1: the extraction of the features directly from the ear image without applying the proposed image fragmentation. case 2: the application of the proposed image fragmentation method before feature extraction. We will first present the experimental results carried out in sub-set 1. As reported in Figure 6, LPQ achieves lower EERs than those achieved by LBP where we note in case 1, i.e. the case where we do not apply our image fragmentation method, an EER = 78% achieved by LPQ (illustrated in Figure 6(a)) and EER = 80% achieved by LBP (illus-

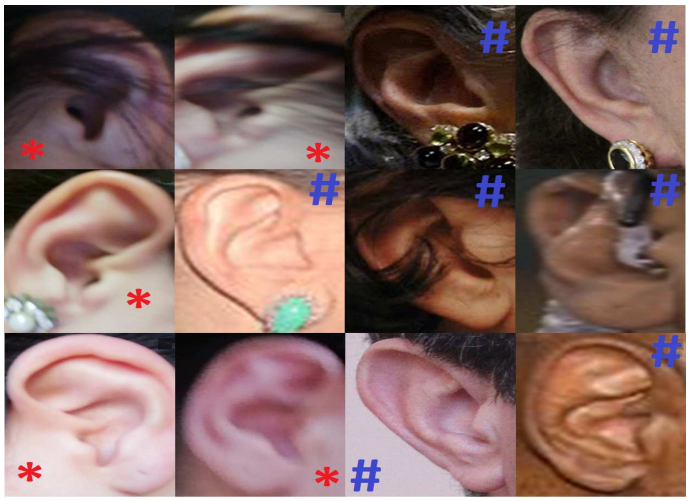

Fig. 5 Images from our database marked with (*) and AWE marked with (#). The first two lines show ear images with occlusion (from sub-set 1), the last line shows ear images without occlusion (from sub-set 2).

trated in Figure 6(c)). As shown in Figure 6(b) and 6(d), the application of our image fragmentation method significantly improves the performances of the ear recognition system. Indeed, the EER rate achieved by LPQ decreased from 78% to 71.2%, i.e., an improvement of 6.8%. For LBP EER goes from 80% to 69.1%, so there is an improvement of 10.9%. We explain this by the fact that the proposed method takes into account only the parts that are not occluded, thanks to the formula (1) described in the Section 3.1.1.

It should be noted that since the sub-set 1 contains images that all have occlusions, blur, variations in illumination and rotation, EERs are quite high and reflect the true performance of feature extraction methods when used on images collected under uncontrolled conditions.

Figure 7 shows the results obtained when applying LPQ and LBP on the sub-set 2. In this case too, LPQ (EER = 28.8%, pictured in Figure 6(a)) achieves better results than LBP (EER = 30.5%, pictured in Figure 6(c)). The proposed segmentation method also improves the results when it is applied to the sub-set 2. As shown in Figure 7(b), EER = 20.8% when applying LPQ for feature extraction and EER = 25.3 % for LBP (pictured in Figure 6(d)). It can be seen in this case that the obtained EERs are lower than those obtained in sub-set 1. It is therefore concluded how much the occlusion deteriorates the performance of the authentication system.

4.2 Arm gesture

In order to evaluate the performances of our proposed arm gesture authentication system we implemented a real time Android application on a smartphone

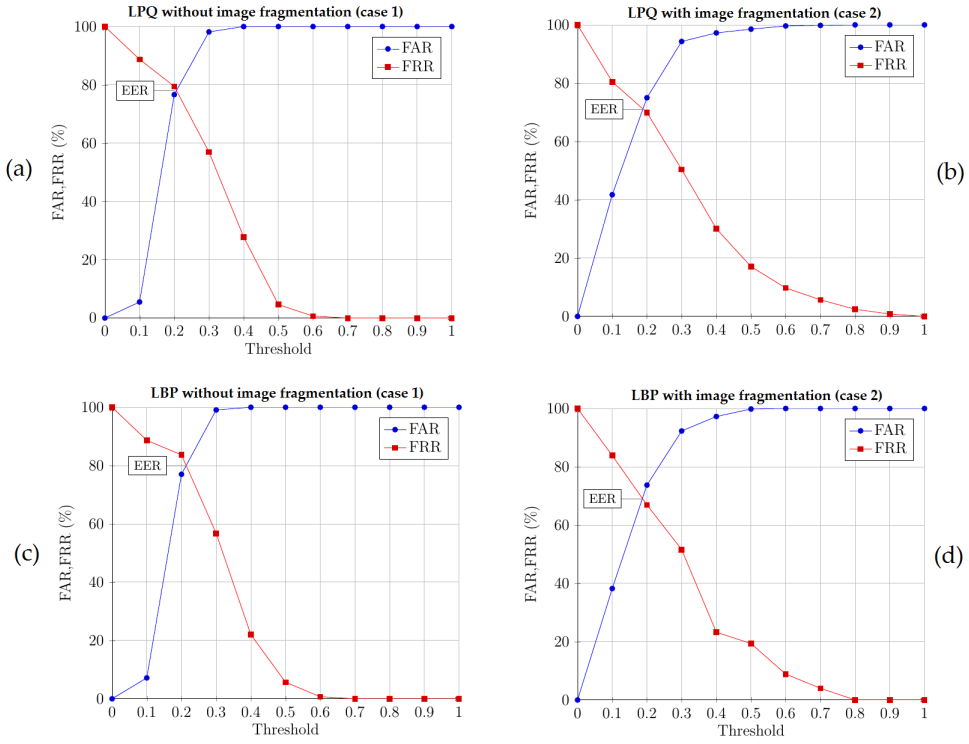

Fig. 6 FAR/FRR curve performed by LPQ and LBP applied to ear images with occlusions (sub-set 1) in both cases without/with application of the proposed image fragmentation.

to collect the arm gesture data simultaneously from the accelerometer and gyroscope. The data collection was done in unconstrained conditions without any user interaction. We have set the sampling frequency to 50Hz and we did not collect any information that can be used to identify the volunteers to preserve their privacy. The data were collected from 7 volunteers within an age ranging from 19 to 56 years recorded from three different sessions spanning over one week. To extend the data set for the tests we used the public database HMOG (downloadable on: http://www.cs.wm.edu/qyang/hmog.html) that contains the record of touch, sensor and key press data invoked by 100 users during document reading, text production and navigation on a map to locate a destination.

In our implementation, we set the number of trees in the forest to 100. To evaluate the performance of our system, we divided the initial database into three sub-sections: 60% for training, 20% for testing and 20% for verification. To have credible results, the three subparts are disjoined and for the verification, we included data from users who did not participate in the classifier's learning to test its robustness. Table 4 shows the results obtained in terms

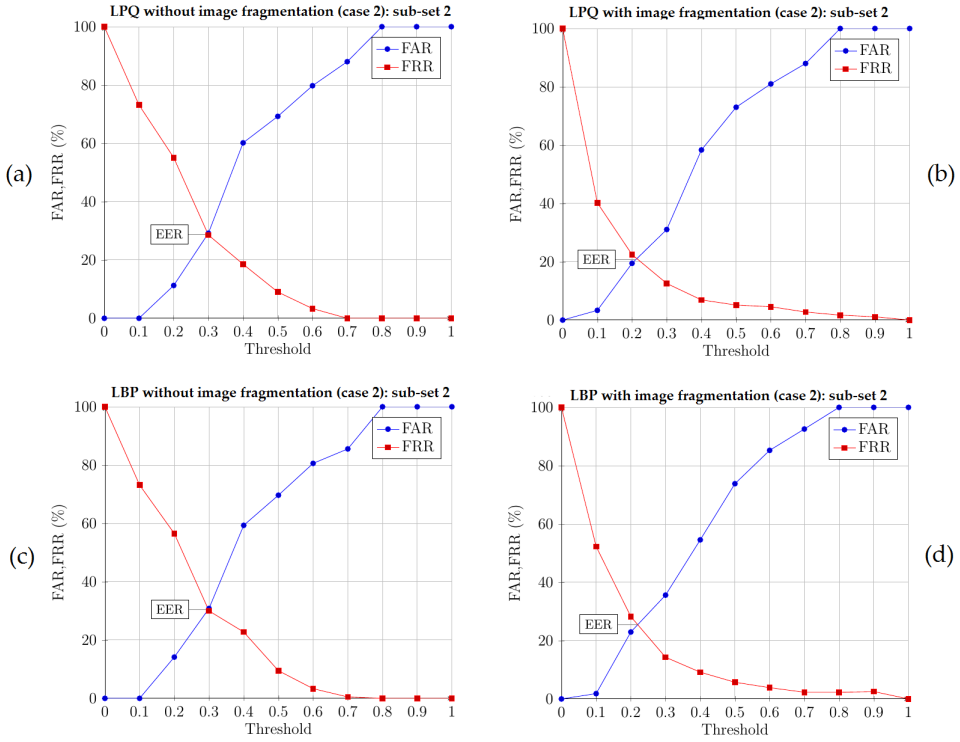

Fig. 7 FAR/FRR curve performed by LPQ and LBP applied to ear images without occlusions (sub-set 2) in both cases without/with application of the proposed image fragmentation.

of EER and FAR in the case where 1) the verification section contains only data belonging to users who have participated in the classifier's learning but belonging to different and separate sessions (case 1 in Table 4); and 2) the verification section contains only data belonging to users who have not participated in the classifier's learning (case 2 in Table 4). This test allows us to test the robustness and effectiveness of the system in recognizing impostors. As shown in Table 4, FAR and EER are higher in case 2. This is because the set used for the verification contains only users who have not participated in classifier learning, however the obtained rates remain acceptable.

Table 4 Summary of EER and FAR achieved by arm gesture classifier.

Case	FAR (%)	EER (%)
1	6	9.66
2	28.33	20.83

To behold the impact of the fusion of the accelerometer and gyroscope data we have trained our classifier according to several cases:

1. The application of raw feature vector $V1$, where each element of this vector represents one of the features considered and presented in Section 3.2.2 calculated for each sensor:

$$V1 = (\overline{X}_{acc}, \sigma_{acc}, Med_{acc}, K_{acc}, \overline{X}_{gyro}, \sigma_{gyro}, Med_{gyro}, K_{gyro}).$$

2. The fusion of the features obtained from the accelerometer $(\overline{X}_{acc}, \sigma_{acc}, Med_{acc}, K_{acc})$ by applying the weighted sum with equal weights for each feature:

$$V2 = (0.25{\cdot}\overline{X}_{acc}{+}0.25{\cdot}\sigma_{acc}{+}0.25{\cdot}Med_{acc}{+}0.25{\cdot}K_{acc}, \overline{X}_{gyro}, \sigma_{gyro}, Med_{gyro}, K_{gyro}).$$

3. The fusion of the features obtained from the gyroscope $(\overline{X}_{gyro}, \sigma_{gyro}, Med_{gyro}, K_{gyro})$ by applying the weighted sum with equal weights for each feature:

$$V3 = (\overline{X}_{acc}, \sigma_{acc}, Med_{acc}, K_{acc}, 0.25{\cdot}\overline{X}_{gyro}{+}0.25{\cdot}\sigma_{gyro}{+}0.25{\cdot}Med_{gyro}{+}0.25{\cdot}K_{gyro}).$$

4. The fusion of both of the accelerometer and gyroscope features

$$V4 = 0.25{\cdot}(\overline{X}_{acc}{+}\sigma_{acc}{+}Med_{acc}{+}K_{acc}, \overline{X}_{gyro}{+}\sigma_{gyro}{+}Med_{gyro}{+}K_{gyro}).$$

Table 5 summarizes the accuracy results achieved according to the different cases. The accuracy C is calculated as described in equation 4.

$$C = \frac{\sum_{i=1}^{n} a_i}{n}, \tag{4}$$

where n is the user number and a_i is the accuracy obtained for each user. As can be seen on the table, the highest accuracy is $C = 92.65$, which is obtained in the case of gyroscope data fusion.

Table 5 Accuracy according to different features vectors.

Feature vector	Accuracy (%)
Raw feature vector ($V1$)	89.92
Fusion of accelerometer features ($V2$)	86.40
Fusion of gyroscope features ($V3$)	**92.65**
Fusion of accelerometer and gyroscope features ($V4$)	83.92

We also evaluated our system in terms of EER. Table 6 presents the obtained results. As can be seen in the table, the fusion of the gyroscope features produces the best EERs, which is consistent with the previous obtained results.

Based on the obtained results, we have adopted the case 3 for our arm gesture authentication system, i.e., the classifier training is done using the fused data from the gyroscope, as described in V3.

Table 6 EER according to different features vectors.

Feature vector	EER (%)
Raw feature vector ($V1$)	23.74
Fusion of accelerometer features ($V2$)	14.72
Fusion of gyroscope features ($V3$)	**10.60**
Fusion of accelerometer and gyroscope features ($V4$)	15.16

4.3 Fusion

To investigate the advantage in fusing ear and arm gesture biometrics we implement a score-level fusion. The score weights vary from Ear·0.9/Arm·0.1 to Ear·0.1/Arm·0.9. The results are listed in Table 7. The weights Ear·0.1/Arm·0.9 achieve the best results in both sub-set 1 and 2. We therefore conclude that arm gesture is the feature that contributes most to the performance of the multimodal system. It comes down to the fact that ear images can be affected with high rotation and illumination variation and sometimes the ear is completely hidden with accessories such as caps, scarves, hair bands or long hair. In this case where the ear is completely hidden, the system makes decisions based only on arm gesture. However, the ear still provides additional information that significantly improves recognition rates and remains a very promising new biometric feature that could replace facial authentication as it does not suffer from the problem of aging and expression variation.

Table 7 ear and arm gesture biometrics fusion results.

Fusion	EER (%) for subset 1	EER (%) for subset 2
score-level fusion (ear 0.1 / arm 0.9)	**7.77**	**5.15**
score-level fusion (ear 0.2 / arm 0.8)	15.6	10.3
score-level fusion (ear 0.3 / arm 0.7)	23.2	15.4
score-level fusion (ear 0.4 / arm 0.6)	31.1	20.7
score-level fusion (ear 0.5 / arm 0.5)	38.8	25.6
score-level fusion (ear 0.6 / arm 0.4)	46.6	31
score-level fusion (ear 0.7 / arm 0.3)	54.9	36.4
score-level fusion (ear 0.8 / arm 0.2)	62.1	40.8
score-level fusion (ear 0.9 / arm 0.1)	70.2	47

Table 8 lists the average EERs and testing times from the unimodal and multimodal schemes. As the results show, multimodal systems achieved significantly higher authentication accuracy than unimodal systems.

4.3.1 Comparison results with Abate et al. [10]

By using the same dataset, Table 9 reports the final results obtained of the EERs achieved by our scheme, by our competitor scheme proposed by Abate

Table 8 EER and testing time results.

Modality	EER (%)	Testing time (sec.)
Ear	20.8	1.713
Arm gesture	10.60	0.165
Score-level fusion	5.15	0.858

et al. [10] and the rates with which our proposal improves those of [10] in both cases of the use of images with and without occlusion. As can be seen on the table, our proposal improves the performance of arm gesture authentication by 2.55% and for the ear we achieve an improvement of 10.9 %.

Table 9 Comparative results with Abate et al. [10] in terms of EER.

	Our scheme			Abate et al. [10]			enhancement		
	ear	arm gesture	fusion	ear	arm gesture	fusion	ear	arm gesture	fusion
subset 2	20.8	10.60	5.15	25.3	13.15	12.29	4.5	2.55	7.14
subset 1	69.1	10.60	7.77	80	13.15	14.05	10.9	2.55	6.28

5 Conclusion

In this work, we have developed an efficient, unobtrusive and robust multi-biometric system to authenticate smartphone user when he/she is responding a call. Our system combines ear shape features-set and arm gesture features-set based on score-level fusion strategy. The arm gesture achieved EER values of 0.1060. To address the problem of occlusion in ear images, we have proposed an image fragmentation method for calculating ear features that significantly improve the performance of the authentication system. Experiments carried out on images collected under totally uncontrolled conditions which all contain occlusion show very encouraging results, where an improvement in EER of almost 11% is achieved. The multimodal biometric system achieves an EER of 5.15%. For future work, we plan to explore new architectures of deep learning to improve our method. We also plan to exploit other databases.

Acknowledgements This work was carried out in the framework of research activities of the laboratory LIMED, which is affiliated to the Faculty of Exact Sciences of the University of Bejaia, with collaboration with LIGM, University of Gustave-Eiffel.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Paul P. P., Gavrilova M. L., and Alhajj R. (2014). *Decision Fusion for Multimodal Biometrics Using Social Network Analysis*, IEEE Transactions On Systems, Man, And Cybernetics: Systems, Vol. 44(11), pp. 1522–1533, DOI: 10.1109 / TSMC.2014.2331920.
2. Rodrigues R. N., Lee L. L. and Venu G . (2009). *Robustness of multimodal biometric fusion methods against spoof attacks*, Journal of Visual Languages & Computing, Vol. 20(3), pp. 169–179, DOI: 10.1016/j.jvlc.2009.01.010.
3. Rodrigues R. N., Niranjan K. and Venu G. (2010). *Evaluation of biometric spoofing in a multimodal system*, 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), DOI: 10.1109/BTAS.2010.5634531
4. Akhtar Z., Sandeep K., and Nasir A. (2011). *Spoof attacks on multimodal biometric systems*, International Conference on Information and Network Technology. Vol. 4, pp 46–51.
5. Zhang Y., Mu Z., Yuan L. and Yu C. (2018). *Ear verification under uncontrolled conditions with convolutional neural networks*, IET Biometrics, Vol. 7(3), pp. 185–198, DOI: 10.1049/iet-bmt.2017.0176.
6. Ojansivu V. and Heikkila J. (2008). *Blur insensitive texture classification using local phase quantization*, Image and Signal Processing: 3rd International Conference, ICISP 2008, pp. 236–243, DOI: 10.1007/978-3-540-69905-7_27.
7. Kandgaonkar T. V., Mente R. S., Shinde A. R. and Raut S. D. (2015). *Ear biometrics: A survey on ear image databases and techniques for ear detection and recognition*, IBMRD's Journal of Management & Research, Vol. 4(1), pp. 88–103, DOI: 10.17697/ibmrd/2015/v4i1/60357.
8. Emersic Z., Struc V. and Peer P. (2017). *Ear Recognition: More Than a Survey*, Neurocomputing, Vol. 255, pp. 26–39, DOI: 10.1016/j.neucom.2016.08.139.
9. Sitova Z., Sedenka J., Yang Q., Peng G., Zhou G., Gasti P. and Balagani K. S. (2016). *HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users*, IEEE Transactions on Information Forensics and Security, Vol. 11(5), pp. 877–892, DOI: 10.1109/TIFS.2015.2506542.
10. Abate A. F., Nappi M. and Ricciardi S. (2019). *I-Am: Implicitly Authenticate Me Person Authentication on Mobile Devices Through Ear Shape and Arm Gesture*, IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 49(3), pp. 469–481, DOI: 10.1109/TSMC.2017.2698258.
11. Akhtar Z., Buriro A., Crispo B. and Falk T. H. (2017). *Multimodal smartphone user authentication using touchstroke, phone-movement and face patterns*, Global Conference on Signal and Information Processing (GlobalSIP), pp. 1368–1372, DOI: 10.1109 / GlobalSIP.2017.8309185.
12. Arteaga-Falconi J. S., Al Osman H., El Saddik A. (2018). *ECG and fingerprint bimodal authentication*, Sustainable Cities and Society, Vol. 40, pp. 274–283, DOI: 10.1016/j.scs.2017.12.023.
13. Gofman M. I., Mitra S. and Smith N. (2016). *Hidden Markov Models for feature-level fusion of biometrics on mobile devices*, IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), pp. 1–2, DOI: 10.1109/AICCSA.2016.7945755.
14. Zhang Q., Li H., Sun Z. and Tan T. (2018). *Deep Feature Fusion for Iris and Periocular Biometrics on Mobile Devices*, IEEE Transactions on Information Forensics and Security, Vol. 13(11), pp. 2897–2912, DOI: 10.1109 / TIFS.2018.2833033.
15. Sequeira A. F., Monteiro J. C., Rebelo A. and Oliveira H. P. (2014). *MobBIO: A multimodal database captured with a portable handheld device*, 2014 International Conference on Computer Vision Theory and Applications (VISAPP), Vol. 3.

16. Gofman M., Mitra S., Cheng K., and Smith N. (2015). *Quality-Based Score-level Fusion for Secure and Robust Multimodal Biometrics-based Authentication on Consumer Mobile Devices*, ICSEA 2015 : The Tenth International Conference on Software Engineering Advances, pp. 274–276.
17. Abozaid A, Haggag A., Kasban H. and Eltokhy M. (2018). *Multimodal biometric scheme for human authentication technique based on voice and face recognition fusion*, Multimedia Tools and Applications, DOI: 10.1007/s11042-018-7012-3.
18. Mahmoud R. O., Selim M. M. and Muhi O. A. (2020). *Fusion Time Reduction of a Feature Level Based Multimodal Biometric Authentication System*, International Journal of Sociotechnology and Knowledge Development, Vol. 12(1), pp. 67-83, DOI: 10.4018/IJSKD.2020010104.
19. Wu L., Yang J., Zhou M., Chen Y. and Wang Q. (2020). *LVID: A Multimodal Biometrics AuthenticationSystem on Smartphones*, IEEE Transactions on Information Forensics and Security, Vol. 15, pp.1–6, DOI: 10.1109/TIFS.2019.2944058.
20. Dhvani S. and Vinayak H. (2016). *IoT Based Biometrics Implementation on Raspberry Pi*, Procedia Computer Science, Vol. 79, pp. 328–336, DOI: 10.1016/j.procs.2016.03.043.
21. Habib K., Arild T. and Leister W. (2014). *A novel authentication framework based on biometric and radio fingerprinting for the IoT in eHealth*, Proceedings of International Conference on Smart Systems, Devices and Technologies (SMART), pp. 32–37, DOI: 10.13140/2.1.3944.1286.
22. Macek N., Franc I., Bogdanoski M. and Mirkovic A. (2016). *Multimodal biometric authentication in IoT: Single camera case study*, 8th International Conference on Business Information Security (BISEC'2016), pp. 33–37.
23. Olazabal O., Gofman M., Bai Y., Choi Y., Sandico N., Mitra S. and Pham K. (2019). *Multimodal Biometrics for Enhanced IoT Security*, 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), pp. 886–893, DOI: 10.1109/CCWC.2019.8666599.
24. Gofman M. I., Sandico N., Mitra S., Suo E., Muhi S. and Vu. T. (2018). *Multimodal Biometrics via Discriminant Correlation Analysis on Mobile Devices*, Proceedings of the International Conference on Security and Management (SAM). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
25. Kim D.J., Kil-Ram H. and Kwang-Seok H. (2008). *An Implementation of Multimodal Speaker Verification System using Teeth Image and Voice on Mobile Environment.* Journal of the Institute of Electronics Engineers of Korea CI, Vol. 45(5), pp. 162–172.
26. Aronowitz H., Li M., Toledo-Ronen O., Harary S., Geva A., Ben-David S., Rendel A., Hoory R., Ratha N., Pankanti S. and Nahamoo D. (2014). *Multi-modal biometrics for mobile authentication*, IEEE International Joint Conference on Biometrics, pp. 1–8, DOI: 10.1109/BTAS.2014.6996269.
27. Gofman M. I. and Mitra S. (2016). *Multimodal biometrics for enhanced mobile device security*, Communications of the ACM, Vol. 59(4), pp. 58-65, DOI: 10.1145/2818990.
28. Mohanta T. K. and Mohapatra S. (2014). *Development of Multimodal Biometric Framework for Smartphone Authentication System*, International Journal of Computer Applications, Vol. 102(7), pp. 6–11, DOI: 10.5120/17825-8597.
29. Ojala T., Pietikainen M., and Harwood D. (1996). *A comparative study of texture measures with classification based on feature distributions*, Pattern Recognition, Vol. 29(1), pp. 51–59, DOI: 10.1016/0031-3203(95)00067-4.
30. Kannala J. and Rahtu E. (2012). *BSIF: binarized statistical image features*, In Proceedings of the International Conference on Pattern Recognition (ICPR), pp.1363-1366.
31. Vu N. S. and Caplier A. (2010). *Face Recognition with Patterns of Oriented Edge Magnitudes*, Computer Vision, pp. 313–326, DOI: 10.1007/978-3-642-15549-9_23.
32. Pflug A., Paul P. N. and Busch C. (2014). *A comparative Study on Texture and SurfaceDescriptors for Ear Biometrics*, 2014 International Carnahan Conference on Security Technology (ICCST), pp. 1–8, DOI: 10.1109/BTAS.2014.6996239.
33. Krizaj J., Struc V. and Pavesic N. (2010). *Adaptation of SIFT Features for Robust Face Recognition*, International Conference Image Analysis and Recognition, pp. 394–404, DOI: 10.1007/978-3-642-13772-3_40.

34. Ojansivu V., Rahtu E. and Heikkila J. (2008). *Rotation invariant local phase quantization for blur insensitive texture analysis*, 19th International Conference on Pattern Recognition, pp. 236–243, DOI: 10.1109/ICPR.2008.4761377.
35. Manjunath B. S. and Ma W. Y. (1996). *Texture features for browsing and retrieval of image data*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.18(8), pp. 837–848, DOI: 10.1109/34.531803.
36. Kumar A. and Wu C. (2012). *Automated human identification using ear imaging*, Pattern Recognit. Vol. 45(3), pp. 956–968, DOI: 10.1016/j.patcog.2011.06.005.
37. Introduction to USTB ear image databases (2002), University of Science and Technology of Beijing, http://www1.ustb.edu.cn/resb/en/doc/Imagedb_123_intro_en.pdf.
38. Pflug A., Busch C. and Ross A. (2014). *2D ear classification based on unsupervised clustering*, Proceedings of the International Joint Conference on Biometrics, pp. 1–8, DOI: 10.1109/BTAS.2014.6996239.
39. Malhotra A., Sankaran A., Mittal A., Vatsa M. and Singh R. (2017). *Chapter 6 - Fingerphoto Authentication Using Smartphone Camera Captured Under Varying Environmental Conditions*, Human Recognition in Unconstrained Environments, pp. 119–144. DOI: 10.1016/B978-0-08-100705-1.00006-3.
40. Ahonen T., Hadid A. and Pietikainen M. (2006). *Face Description with Local Binary Patterns: Application to Face Recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28(12), pp. 2037–2041, DOI: 10.1109/TPAMI.2006.244.
41. Dornaika F., Moujahid A., El Merabet Y. and Ruichek Y. (2017). *A Comparative Study of Image Segmentation Algorithms and Descriptors for Building Detection*, Handbook of Neural Computation, pp. 591–606, DOI: 10.1016/B978-0-12-811318-9.00032-6.
42. Barra S., Fenu G., De Marsico M., Castiglione A., Nappi M. (2018). *Have you permission to answer this phone?*, 2018 International Workshop on Biometrics and Forensics (IWBF), pp. 1–7, DOI: 10.1109 / IWBF.2018.8401563.
43. Buriro A., Crispo B. and Zhauniarovich Y. (2017). *Please hold on: Unobtrusive user authentication using smartphone's built-in sensors*, IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), New Delhi, India, pp. 1–8, DOI: 10.1109/ISBA.2017.7947684.
44. Chhabria S. A., Dharaskar R. V., Thakare V. M. (2013). *Survey of Fusion Techniques for Design of Efficient Multimodal Systems*, International Conference on Machine Intelligence Research and Advancement, pp. 486–492, DOI: 10.1109/ICMIRA.2013.103.