Published in: IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, VOL. 49, NO. 3, MARCH 2019

The Version of Record is available online at: 10.1109/TSMC.2017.2698258.

I-Am: Implicitly Authenticate Me Person Authentication on Mobile Devices Through Ear Shape and Arm Gesture

Andrea F. Abate, *Member, IEEE*, Michele Nappi, *Senior Member, IEEE*, and Stefano Ricciardi, *Member, IEEE*

***Abstract*—Today, identity verification is required in many common activities, and it is arguably true that most people would like to be authenticated in the easiest and most transparent way, without having to remember a personal identification number. To this regard, this paper presents a multibiometric system based on the observation that the instinctive gesture of responding to a phone call can be used to capture two different biometrics, namely ear and arm gesture, which are complementary due to their, respectively, physical and behavioral nature. We conducted a comprehensive set of experiments aimed at assessing the contribution of each of the two biometrics as well as the advantage in their fusion to the system's overall performance. Experiments also provide objective measurement of both saliency and correlation of data captured by each sensor involved (accelerometer, gyroscope, and camera) according to various features extraction, features matching, and data-fusion techniques. The reports provide evidences about the potential of the proposed system and method for user authentication "in-the-wild," whilst its eventual usage for person identification is also investigated. All of the experiments have been carried out on a specifically built, publicly available ear–arm database, including multibiometric captures of more than 100 subjects performed during different sessions, that represents an additional contribution of this paper.**

***Index Terms*—Arm gesture, biometrics, ear, multibiometric database, person authentication, smartphones.**

I. Introduction

BIOMETRICS-based person authentication and identification have become common practices in many contexts, and their diffusion is expected to steadily grow in the next years also thanks to the diffusion of the latest generation of mobile devices equipped with a plethora of accurate and reliable sensors along with more and more powerful processors. The worldwide diffusion of mobile communication has indeed made available two billion smartphones embedding in a single ubiquitous device high-resolution cameras, digital compass, gyroscope, accelerometer, positioning system, etc. Such characteristics make these devices suited to operate as biometric terminals, capable of capturing, processing, and comparing multiple biometrics for both identity verification or recognition in a wide range of application scenarios [1]. The still limited raw computing power available in these devices forces biometric application designers and developers to some kind of compromise in terms of selecting and adapting the right algorithm to these platforms and the related operating systems.

However, the ubiquity of these devices, coupled with the familiarity the users have with them, represent a key advantage in the light of their usage for security-related procedures which are becoming more and more frequent in the everyday life of many of us. In any biometrics, indeed, its acceptability characteristic puts an upper limit to its diffusion and, in the end, to its usefulness in a given context. This is particularly true for the average user, which arguably would prefer to be checked in the most transparent possible way instead than be forced to undergo a rigid acquisition procedure.

Starting from these premises, this paper describes a multibiometric system for personal authentication based on smartphone as hardware platform and exploiting the observation that the instinctive gesture of responding to or placing a phone call can be used to capture two different biometrics, namely ear and arm dynamics, which are complementary due to their, respectively, physical and behavioral nature.

This paper stems from the preliminary proposal of a person-authentication approach based on ear and arm-gesture presented in [2], expanding it through a comprehensive set of experiments designed to assess the contribution of each of the two biometrics as well as the advantage in their fusion and the saliency and correlation of data captured by each sensor involved (accelerometer, gyroscope, and camera) according to various features extraction, features matching, and data-fusion techniques.

The contribution of this paper is twofold. First, a novel multimodal biometric is presented, by combining a physical biometric (ear) with a behavioral one (arm gesture) both captured through the same device (smartphone). Second, by exploiting the first statistically significant ear–arm database available, has been possible to validate arm dynamics as a viable biometric, also proving that its contribution in a smartphone-based multimodal approach to identity verification is relevant and beneficial in terms of both robustness (ear acquisition could possibly fail for incorrect framing or excessive motion-blur) and level of accuracy achieved, for a potentially wide range of low to medium-security applications requiring user authentication "in-the-wild" through commonly available mobile devices.

The rest of this paper is organized as follows. Section II resumes main works related to this research, Section III describes the proposed system in detail, while Section IV presents the results of the experiments conducted on the multibiometric database. Finally, Section V draws conclusions, along with directions for future research.

II. Related Work

The proposed system represent an example of smartphone empowered multibiometric system.

The first examples of such systems go back to a decade ago, when the first suitable devices emerged on the market. In 2006, Clarke and Furnell [3] were among the first ones to propose the use of keystroke analysis within a composite authentication algorithm to enable transparent user authentication by mobile devices. Since all smartphones feature one or more imaging sensors, a number of proposals concern the use of these hardware resources for face recognition.

In [4], smartphone technology and wireless networking are exploited to cope with the limitations of blind people in identifying other persons, through a face recognition approach providing audio feedback of identification results. In the effort of optimizing mobile devices performance for computing intensive task, such as face recognition, Cheng and Wang [5] described a GPU-based implementation achieving improved speed in feature extraction and matching along with a significant reduction of energy consumption, a critical factor of any mobile platform.

More recently, Shen *et al.* [6] proved that even the best performing face recognition algorithms, like the sparse representation classfication can be effectively implemented on smartphones with a reasonably short computing time by means of a platform-specific code optimization.

Face has also been combined with voice in the MoBio multimodal biometric system proposed by Tresadern *et al.* [7], by fusing these two biometrics either at the score level or at the feature level for improved identity verification.

A similar multimodal strategy has been adopted in [8], but in this case exploiting ear and voice instead, with data-fusion performed at feature level.

A different approach to multimodal authentication on smartphones is presented in [9], since the authors do not propose to combine two biometric identifiers but rather something the user is (iris biometrics) plus something the user owns (the smartphone characterized by its imaging sensor pattern noise).

Gait recognition has been approached on mobile devices by exploiting their embedded accelerometers. Nickel *et al.* [10] presented a method for extracting gait features to be classified by means of k-nearest neighbor algorithm, demonstrating its practical feasibility on smartphones, while in [11] secondary gait motion affecting arm swing has been proposed as a weak biometric.

Finally, touch dynamic images have been proposed in [12] for mobile user verification as well as for continuous authentication [13] and implicit identification [14].

To the best of our knowledge, this is the first proposal of a system combining ear features and arm gestures for authentication purposes, with both biometrics collected through a mobile device whilst the subject is responding to (or placing) a phone call.

The discriminant power of gestures is known well before the technology for capturing and processing this dynamic info on a portable device was available [15] while the advent of inexpensive sensors for measuring acceleration and orientation, around one decade ago, has greatly contributed to stimulate the exploitation of body dynamics in a wide range of applications [16], [17].

However, the idea of exploiting smartphone embedded motion sensors (mainly the accelerometers and the embedded camera) for biometric applications is not exactly new, though only a few papers can be found in literature on this topic so far. With regard to arm gestures, in 2009, Liu *et al*. [18] presented uWave, a gesture recognition methodology based on quantization of accelerometer readings, dynamic time warping (DTW) and template adaptation allowing the user to define custom gestures via a single training sample. According to this approach a gesture can be detected and recognized by a characteristic time series of acceleration. DTW and multiple sensors fusion is exploited in [19] to detect and recognize "aggressive" driving patterns by means of driver's smartphone, used as a mobile monitoring platform. The proposed method is characterized by fusing interaxial data from different sensors into a single classifier while Euler representation of device attitude is used to improve classification performance. The first attempt of using the arm motion related to the act of answering or placing a phone call as a discriminant feature for authenticating the smartphone user is proposed by Conti *et al*. [20], that propose this particular gesture as a new biometric. The authors presented two variations of DTW algorithm referred to as DTW distance (DTW-D) and DTW similarity that applied to two different sensors (accelerometer and gyroscopes) provided an authentication accuracy on ten subjects comparable to well established biometrics like face or voice. Similar results, though on only six subjects, have been achieved in [21], while Feng *et al*. [22] proposed two different methods, respectively, based on motion statistics and trajectory reconstruction to extract and compare dynamic features resulting from 9-D motion samples, thus confirming the potential of arm's dynamics as a biometric.

Concerning ear, that has been exploited in host-based biometric systems both in 2-D [23], [24] and 3-D [25], [26], its usage for "implicit" person authentication by means of a smartphone has been proposed by Fahmi *et al*. [27] which considered both shape and texture information to represent ear image as a concatenated descriptor. Features matching is then performed using a nearest neighbor classifier in the computed feature space with Euclidean distance as a similarity measure. The authors reported a recognition rate of 92.5% on 20 subjects involved in the experiments.

In the light of the aforementioned works, the main focus of this paper, that is the combination of gestural features and ear shape captured at the same time during the usage of smartphones (or other mobile devices), seems reasonable in terms both of the embedded sensors capabilities and of the complementarity of the two identifiers considered, coupling behavioral and physical characteristics.

III. Description of Proposed Method

The proposed identity authentication method stems from the observation that whenever a smartphone user responds to or places a call, the mobile device's motion sensors could record the motion associated to the phone-holding hand that can be considered as the "end effector" of the cinematic chain composed by clavicle, upper-arm and lower-arm plus hand, hereafter simply referred to as "arm." At the same time, the smartphone's front (secondary) camera could be in a favorable position to capture a sequence of ear images, one or more of which could possibly be used for extracting discriminant features (see Fig. 1). The scores resulting by the two authentication techniques may then be combined together. This approach, described in the following Sections III-A–III-H is implemented through a multimodal biometric system basically composed by three components: 1) the ear subsystem; 2) the arm-motion subsystem; and 3) the fusion-decision subsystem as schematically represented in Fig. 1.

A. Arm Gesture Acquisition

Regardless to being captured contextually or deliberately, arm gesture acquisition is triggered by the smartphone interface that starts recording motion-data from both accelerometer and gyroscope until these readings show that the gesture is over. This happens whenever the acceleration readout is mainly due to gravity (i.e., the other components are below a specific threshold), whereas the gyroscope is also exploited to confirm the completion of the gestures (by considering terminal orientation and angular velocity) as well as to exclude that involuntary movements are considered as a relevant gesture. Acceleration data are captured at a rate of 50 samples/s, though slight variations on the average reading rate are possible indeed. Acceleration data are captured at a rate of 50 samples/s, though slight variations on the average reading rate are possible indeed.

Each sample contains four values (x, y, z, t) including the instant acceleration on three axes and the time elapsed from the start of recording. The resulting 4-D vector is therefore used for features extraction. It has to be remarked that to the aim of improving the reliability of this biometric template, the acquisition process is repeated five times only for the enrollment, and the average of the five vectors is saved.

B. Ear Acquisition

Ear image capture can be performed either contextually to a call, or deliberately for authentication purposes. Whatever the modality adopted, the acquisition process involves the recording of a short video sequence at a frame rate of at least 30 frames/s (the higher, the better) that represents the input for the ear detector based on the Viola–Jones [28] algorithm and specifically trained to recognize left or right ear. The choice of an optimal ear-size threshold to stop the detection process combined with proper capture resolution typically allow for real-time performance of the detector on latest generation hardware, though suboptimal performances could occur in case of insufficient illumination. The candidate ear crops are therefore ordered according to their sharpness, to the aim of selecting the most feature rich image. This is accomplished by measuring the difference (pixel-wise) between each frame and a copy of it blurred by means of a Gaussian filter (see Fig. 2). This measure results to be the lesser, the softer or more blurred the original image is (possibly due to focusing or motion-blur issues), so the frame that maximizes this measure is the best candidate for the subsequent feature extraction process.

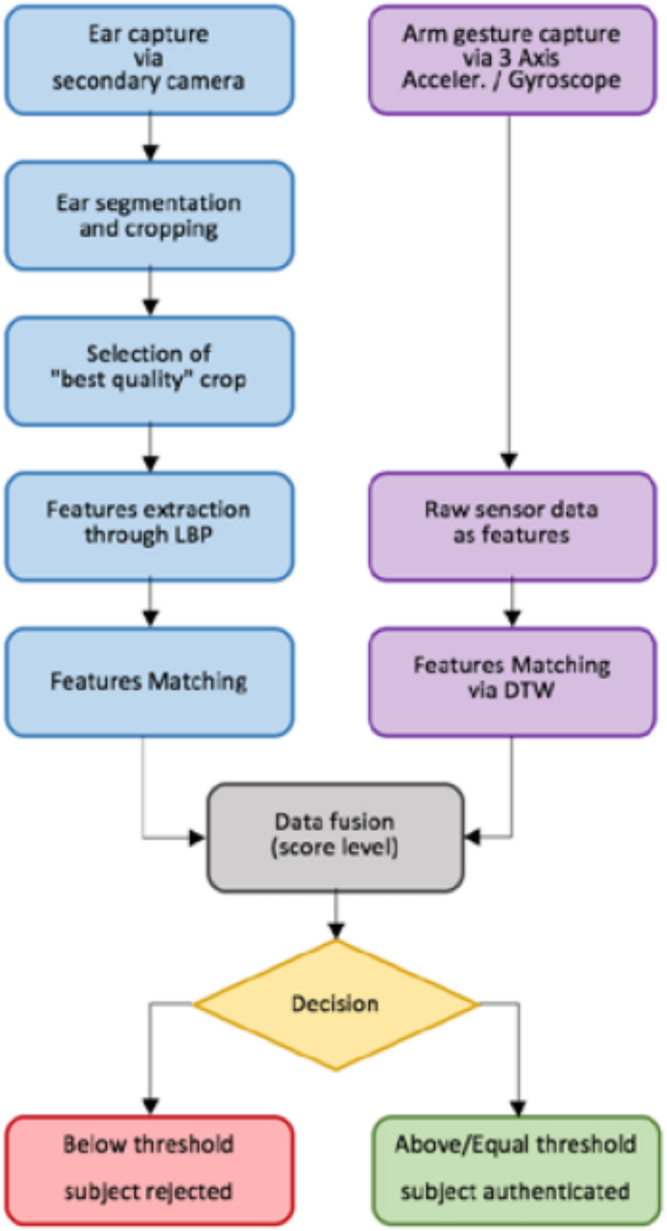
Ear capture via secondary camera
Arm gesture capture via 3 Axis Acceler. / Gyroscope
Ear segmentation and cropping
Selection of "best quality" crop
Features extraction through LBP
Raw sensor data as features
Features Matching
Features Matching via DTW
Data fusion (score level)
Decision
Below threshold subject rejected
Above/Equal threshold subject authenticated

Fig. 1. Schematic of proposed system. The system and the workflow depicted above relates to the most performing configuration tested.

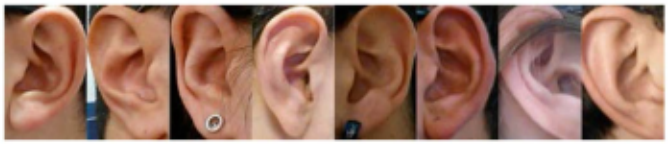

Fig. 2. Examples of ear crops contained in the multimodal database. Original frames captured by the secondary camera of a Samsung Galaxy S4 phone.

C. Extraction of Arm Dynamics

With regard to the arm motion associated to the gesture of responding/placing a call, we evaluated different methods for extracting discriminant info, for instance by representing the captured signal through the interpolating spline or even by using the coefficients of the signal's fast Fourier transform (FFT). In the first approach, the interpolating spline has been computed from the raw data while the feature vector representing the spline contains an ordered sequence of 10, 20, or 30 spline's key-points sampled at regular intervals from the total acquisition time. In the second approach, once the raw data have been converted in the frequency domain by computing the FFT, a low-pass filter has been applied to the frequency spectrum to isolate salient info contained in low frequencies from noise and signal discontinuity typically present in high frequencies. The adopted low-pass filter was designed to cut the higher 3/4 or 7/8 of the whole frequency spectrum. However, as we report in experiment #1 we found that in both cases the advantage of a more compact descriptor was overcome by the performance drop. Consequently, we decided to exploit the whole set of sampled data (typically 60–80 four-tuples) that given an average duration of the captured motion is below 2 s, lead to an average descriptor made of 240–320 values. It is worth noting that no filtering or data cleaning has been applied to the samples, which are raw data indeed, since we wanted to keep the preprocessing load to a minimum. The variable dimensions of these descriptors are inherently addressed by the metric used in the matching stage. This choice delivered the best results, as described later in Section IV.

D. Extraction of Ear Features

Features extraction of ear shape is performed by means of the local binary patterns (LBPs) algorithm [29] that is well known in computer vision and has been widely used for biometric applications involving face [30] and facial expression recognition [31], and palmprint [32] and ear [33] recognition. In this particular implementation, a fixed number of 25 contiguous blocks partitioning the input image has been adopted rather than the more commonly used fixed block size. The reason behind this choice is related to the variable size of the crops, depending by the aforementioned process of ear detection and cropping. As a result, 25 histograms are computed (one for each block) leading to a final concatenated features descriptor comprising 6400 values.

E. Features Matching

The comparison of a probe ear descriptor to a gallery template corresponding to the claimed identity, is performed by means of Euclidean distance between two n-dimensional features $p = (p_1, p_2, \ldots, p_n)$ and $q = (q_1, q_2, \ldots, q_n)$ as given by

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}. \tag{1}$$

In this case, the distance between each couple of corresponding histograms in the probe/gallery vectors is computed. Then, the overall distance between two descriptors is obtained as the sum of all the Euclidean-distances computed between 25 couples of corresponding histograms, normalized in the range [0, 1]. If subject identification is required, a one-to-many comparison between the probe and each of the gallery's templates is performed, resulting in a score vector of the same size of the gallery, instead than a single score as per identity verification.

With regard to arm-motion, since captured samples are typically of different lengths, to the aim of effectively compare them we exploit the DTW algorithm [34], [35] that is particularly suited to find the best alignment between two signal curves by means of a nonlinear transformation with respect to the independent variable (time, in this case), thus implicitly providing a reliable measure of their similarity. In the next lines, the basic formulation of DTW metric is recalled: given two generic time series $R = r_1, r_2, \ldots, r_i, \ldots, r_n$ and $S = s_1, s_2, \ldots, s_j, \ldots, s_m$, of length n and m, respectively, to align two sequences using DTW an n-by-m matrix is built, where the $(i_{\text{th}}, j_{\text{th}})$ element of the matrix contains the (typically Euclidean) distance $d(r_i, s_j)$ between the two points r_i and s_j. Each element (i, j) in the matrix corresponds to the alignment between the points r_i and s_j. A warping path W, is a contiguous (in the sense stated below) set of matrix elements that defines a mapping between R and S. The kth element of W is defined as $w_k = (i, j)_k$, so, we have

$$W = w_1, w_2, \ldots, w_k, \ldots, w_k \quad \max(m, n) \le K < m + n - 1. \tag{2}$$

A few constraints W is typically subject to, include *boundary conditions*, *continuity*, and *monotonicity*.

The first constraint requires the warping path to start and finish in diagonally opposite corner cells of the matrix [i.e., $w_1 = (1, 1)$ and $w_K = (m, n)$]. The second constraint restricts the allowable steps in the warping path to adjacent cells, including diagonally adjacent cells [i.e., given $w_k = (a, b)$ then $w_{k-1} = (a', b')$, where $a - a' \le 1$ and $b - b' \le 1$]. The third constraint forces the points in W to be monotonically spaced in time [i.e., $w_k = (a, b)$ then $w_{k-1} = (a', b')$ where $a - a' \ge 0$ and $b - b' \ge 0$]. There are exponentially many warping paths that satisfy the above conditions, however, we are interested only in the path which minimizes the warping cost

$$\text{DTW}(R, S) = \min\left\{\sqrt{\sum\nolimits_{k=1}^{K} w_k / K}\right\}. \tag{3}$$

The K in the denominator is used to compensate for the fact that warping paths may have different lengths. This path can be found very efficiently using dynamic programming to evaluate the following recurrence which defines the cumulative distance $\gamma(i,j)$ as the distance $d(i,j)$ found in the current cell and the minimum of the cumulative distances of the adjacent elements

$$\gamma(i,j) = d(r_i, s_j) + \min\{\gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1)\}. \quad (4)$$

As an example, different acquisitions of arm motion often highlight different arm speed and motion duration. In these cases, motion curve alignment via overlapping or curve offsetting would be affected by their differences, whilst DTW is able to find their similarities.

Various metrics have been considered and evaluated against the baseline provided by 3-D Euclidean distance: the monodimensional DTW-D (matching only the curves related to the same axis under the hypothesis of a particular motion-specific relevance of one axis), the average of the three mono-dimensional DTW distances, and the multidimensional DTW distance (MD-DTW) [36] briefly resumed in the following lines. Let R and S be two series of dimension K and length n and m, respectively. The first step of MD-DTW algorithm is to normalize each dimension of R and S separately to a zero mean and unit variance, eventually applying a Gaussian smoothing to each dimension. Fill the $n \times m$ distance matrix D according to

$$D(i,j) = \sum_{k=1}^{K} |R(i,k) - S(j,k)| \quad (5)$$

and use this distance matrix to find the best synchronization with the standard DTW algorithm.

F. Fusion Rules

Since we found the ear–arm Pearson's correlation coefficient [37] computed on the multimodal dataset collected is 0.3207, there is a potential advantage in fusing the two biometrics, although a relatively small correlation exists. Main techniques for fusing info from multiple sensors [38] include feature-level fusion, score-level fusion [39], and decision-level fusion [40], [41]. In the proposed system, after both ear and arm distances have been computed, a score-level weighted data fusion is performed according to the following rules specific to the verification and identification scenarios:

$$\text{score}_{\text{ver}} = \text{score}_{\text{ear}} * 0.1 + \text{score}_{\text{arm}} * 0.9 \quad (6)$$

$$\text{score}_{\text{id}} = \text{score}_{\text{ear}} * 0.9 + \text{score}_{\text{arm}} * 0.1. \quad (7)$$

The inversion of the weights that characterizes the identification fusion rule, stems from the results of experiments described in Section IV according to which in the verification scenario the arm gesture biometric performs considerably better than ear biometric, while the opposite is true for the identification scenario. In both cases, the fusion of the two biometrics leads to a measurable improvement of performances, a result in line with the measured Pearson's correlation coefficient.

G. Decision Stage

Finally, the final score resulting from the adopted fusion rule is compared to a decision threshold [corresponding to the system's equal error rate (EER)] to evaluate whether the subject can be authenticated or not. Data fusion at decision level has been experimented also. In this case, two decision schemes have been considered: 1) authenticating the user only if both ear AND arm probes match the templates corresponding to the claimed identity or 2) authenticating the user if ear probe results genuine OR arm probe results genuine. A feature-level data-fusion scheme would not be beneficial in this case, mainly due to the high dishomogeneity between the two types of feature vectors to be combined.

H. Computational Load and Performance Issues

The main design requirement of the system based on proposed approach, is to enable user authentication exploiting mobile devices instead of larger more performing computing platforms. This goal translates in an operational constraint requiring that all the aforementioned stages have to be performed in real-time or near real-time. In this sense, the choice of LBP (ear features) and raw dynamics data (arm features) reduce the computing load of feature-extraction stage to a minimum, while Euclidean distance (ear) and DTW (arm) for feature-matching can be performed rapidly enough to provide a decision typically in under 1 s (for the authentication scenario on the hardware used for the experiments). Since the technology trend for mobile devices implies higher and higher image capture frame rate, the advantage of having both more frames to choose from and less blurred ear images should be balanced by more computing power or more efficient methods to find the best ear crop.

IV. Experiments

This section describes in detail the experiments designed and accomplished to measure the effectiveness of the two biometrics considered, both evaluated separately or combined together within the proposed smartphone-based biometric system described before. First, the methodology behind the acquisition process for both the modalities and the resulting unprecedented database will be described. Second, three experiments aimed at quantitatively assessing arm gesture (experiment #1), ear (experiment #2), and ear–arm (experiment #3) in the authentication scenario, will be presented. Finally, performance of the integrated system in an identification scenario will be reported and discussed in experiment #4.

A. Building the Ear–Arm Database

As anticipated above, one of the contribution of this paper is the building of a public multibiometric database available at http://www.biplab.unisa.it/portal/index.php/dataset/e-ag-database, specifically designed to be used for identity verification based on ear, arm gesture, or exploiting the combination of both these biometrics. The database has been built by collecting ear images and arm dynamics through the built-in sensors of a Samsung Galaxy S4 smartphone, a model of smartphone whose main features are representative of a large range of mobile phones currently present on the market.

In order to achieve statistically meaningful results from the experiments, a total of 100 different subjects has been involved in the capture process. The whole acquisition has been intentionally performed in the course of three different sessions spanning over two weeks. About a third (30) of these 100 subjects attended to all of the three sessions, whilst the other ones (70) participated only to the first one. Multiple acquisition over time also offered the opportunity to stress the stability of arm-motion (which has a strong behavioral component) as a biometric. The following three biometric templates have been captured for each of the 70 participants to the first session: 1) three ear images; 2) three accelerometer + gyroscope recordings (sitting); and 3) three accelerometer + gyroscope recordings (standing).

For each of the 30 participants to all of the three sessions, nine biometric templates (the three listed above for each of the three sessions) have been captured instead. Usually, the ear used to listen through a phone is always the same. For this reason, the ear acquired was the one normally used by each subject when using a phone. According to this choice, the database includes both right and left samples, the latter being 16% of the whole dataset (see Fig. 2). The experiments described hereafter have been conducted on a subset of the aforementioned database, containing a normalized amount of three templates for each subject, achieved by considering only one sample for each session for the subjects captured three times. The resulting dataset features 300 ear images, 600 (300 sitting + 300 standing) accelerometer recordings, and 600 gyroscope recordings (300 sitting + 300 standing). We remark again that arm gesture samples have not undergone any noise reduction/cleaning. In the verification scenario, this database is organized into two logical units: 1) the gallery-set, comprising one template for each subject and 2) the probe-set containing the two remaining templates that represent access trials subsequent to subject's enrollment. The system has been designed so that 50% of the simulated accesses were from genuine users and the remaining 50% from impostors. All the results presented in the following sections refer to the average computed on 100 iterations of each experiment achieved through 100 permutations of probe's identity (with a genuine/impostor ratio equal to 1) over 200 verification trials, one for each element of the probe-set.

B. Experiment #1—Arm Gesture

In experiment #1, an objective assessment of arm gesture as a biometric is carried out exploiting two well established performance metrics: 1) area under the ROC curve (AUROC) and 2) EER. More in detail, this experiment aims at measuring: 1) performance of various features extraction/matching methods applied to accelerometer data; 2) performance of various features extraction/matching methods applied to gyroscope data; 3) the eventual benefit in fusing accelerometer and gyroscope data; and 4) the robustness of arm gesture to pose variations (with subjects sitting and standing). With regard to features extraction, three methods have been considered, as anticipated in Section III-D: 1) *raw data vector* composed (in average) by 60 four-tuples for a total of 240 elements; 2) *compressed features vector composed by 10, 20, or 30 polynomial key-points* of the spline interpolating the motion curve; and 3) *compressed features vector composed by FFT coefficients* related to the low frequencies (low-pass filter applied to preserve 12.5% or 25% of the full frequencies spectrum). With regard to features matching, the following metrics have been considered, as anticipated in Section III-E: 1) 3-D Euclidean distance (baseline); 2) DTW between *x*-axis, DTW between *y*-axis, DTW between *z*-axis; 3) mean value of DTW distances on *xyz* axes; and 4) multidimensional DTW.

TABLE I

COMPARISON OF DIFFERENT FEATURES MATCHING METHODS APPLIED TO RAW ACCELEROMETER DATA

Features matching methods (raw data, sitting)	**AUROC**	**EER**
Average of DTW(X); DTW(Y); DTW(Z)	0.9382	0.1315
DTW(Y)	0.8727	0.2066
Multi-Dimensional DTW	0.8552	0.2386
DTW(X)	0.8530	0.1711
DTW(Z)	0.8339	0,2298
3D Euclidean distance	0.7905	0,2776

TABLE II

COMPARISON OF DIFFERENT FEATURES EXTRACTION/MATCHING METHODS APPLIED TO ACCELEROMETER DATA

Features extraction methods (accelerometer data, sitting)	**AUROC**	**EER**
DTW average on 30 spline key-points	0.9272	0.1506
DTW average on 20 spline key-points	0.9142	0.1688
DTW average on 10 spline key-points	0.8809	0.2091
Euclidean Dist. on 30 spline key-points	0.8576	0.2374
Euclidean Dist. on 20 spline key-points	0.8555	0.2361
Euclidean Dist. on 10 spline key-points	0.8553	0.2300
Euclidean Dist. on FFT (0.125 low pass filter)	0.7791	0.3016
Euclidean Dist. on FFT (0.25 low pass filter)	0.7700	0.3079

1) Results From Accelerometer Data: Table I resumes the quantitative results achieved according to the above metrics applied to raw data, ordered by decreasing AUROC values. As anticipated in Section III, the best performing metric is the mean of the DTW distances computed separately on each axis, that improves considerably over the DTW-D on the single *y* axis (which contains more salient info due to the specificity of the "responding to a call" motion pattern). MD-DTW performs slightly worse than the latter, but still measurably better than DTW(*X*) and DTW(*Z*). The poor performance of 3-D Euclidean distance here, is easily explained considering the temporal misalignments characterizing the motion recordings that affect severely this metric. In Table II, features representations based on interpolating spline key-points and FFT coefficients are compared and reported according to the decreasing AUROC results. It is easy to spot the superiority

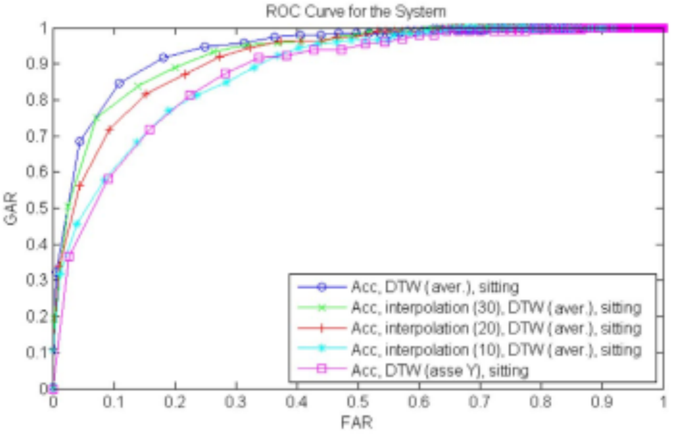

Fig. 3. ROC curves for arm gesture (accelerometer data) related to the five best performing combination of features extraction/matching metrics.

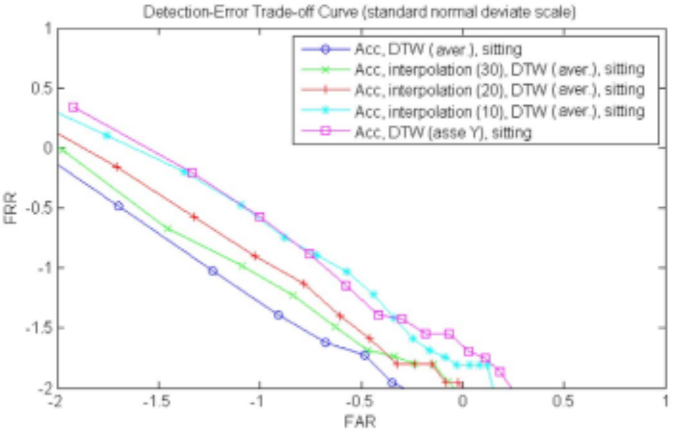

Fig. 4. DET curves for arm gesture (accelerometer data) related to the five best performing combination of features extraction/matching metrics.

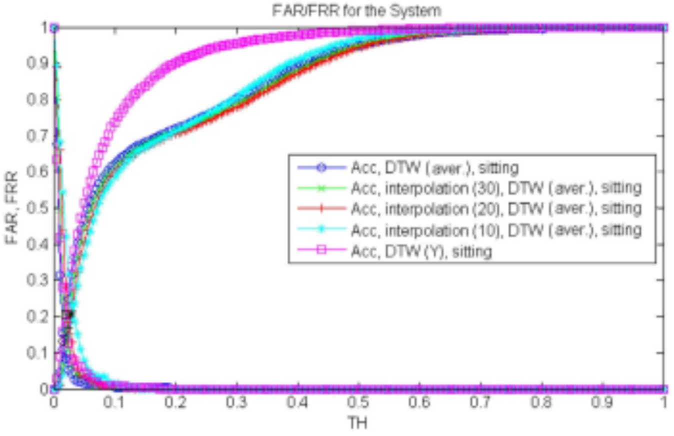

Fig. 5. FAR/FRR curves for arm gesture (accelerometer data) related to the five best performing combination of features extraction/matching metrics.

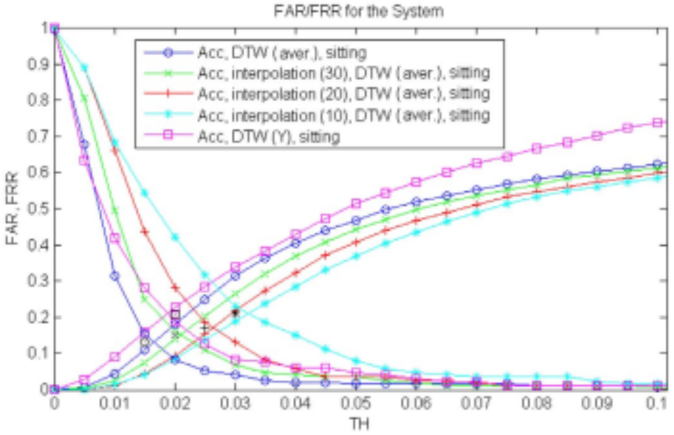

Fig. 6. Close-up of intersecting FAR/FRR curves for arm gesture (accelerometer data) related to the five best performing combination of features extraction/matching metrics.

TABLE III
COMPARISON OF DIFFERENT FEATURES EXTRACTION/MATCHING METHODS APPLIED TO GYROSCOPE DATA

Features extraction/matching methods (gyroscope data, sitting)	**AUROC**	**EER**
DTW average	0,8306	0,2611
DTW average on 30 spline key-points	0,8173	0,2752
DTW average on 20 spline key-points	0.8162	0.2658
DTW average on 10 spline key-points	0.7920	0.2658
DTW (*Z* axis)	0.7858	0.2672
Euclidean Dist. on 10 spline key-points	0.7817	0.3151
Multi-dimensional DTW	0.7788	0.2988
DTW (*Y* axis)	0.7753	0.2944
Euclidean dist. on FFT (0.125 low pass filter)	0.7721	0.3131
Euclidean dist. on 30 spline key-points	0.7695	0.3112
Euclidean dist. on 20 spline key-points	0.7682	0.3205
Euclidean dist. on 10 spline key-points	0.7600	0.3041
Euclidean distance	0.7153	0.3386
DTW (*X* axis)	0.6814	0.3759

of DTW (mean) metric, while the number of spline key-points seems to directly affect the verification performance. In general, FFT coefficients perform worse than spline key-points, and even worse if the low-pass filter preserve part of the higher frequencies.

This observation seems consistent to the hypothesis that the most salient content in arm gesture acceleration data is found in the low-end of the frequencies spectrum, less affected by motion discontinuity. In the following Figs. 3–6 depict (for the five best combinations of features representations and matching methods), respectively, the ROC curves, the DET curves, the FAR/FRR curves, and a close-up view of the FAR/FRR curves intersection, showing more in detail the EER zone.

2) Results From Gyroscope Data: In Table III, the features extraction and matching methods together, already considered in Tables I and II are applied and evaluated for gyroscope data, to understand how salient this information is in the context of arm gesture biometric.

The numeric results and the related ROC, DET, and FAR/FRR (see Figs. 7–9) curves highlight the inferior discriminating power of gyroscope compared to the accelerometer, that translates in lower verification performance as confirmed by the best EER value (0.26) versus the corresponding best value in Table I (0.15). With regard to the matching methods, the DTW (average) confirms its advantage over the other metrics, though this edge reduces if spline-based interpolation

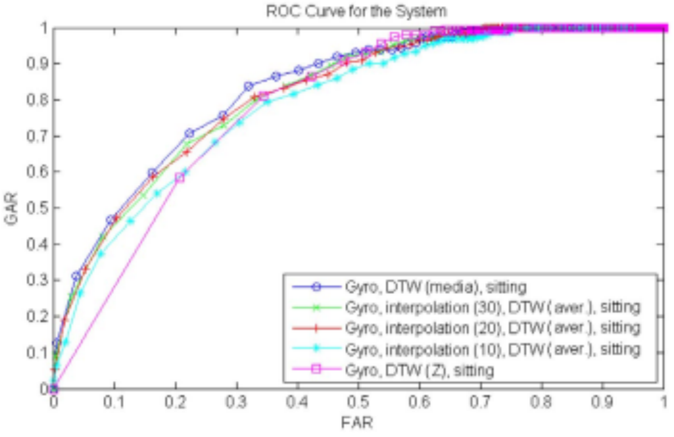

Fig. 7. ROC curves for arm gesture (gyroscope data) related to the five best performing combination of features extraction/matching metrics.

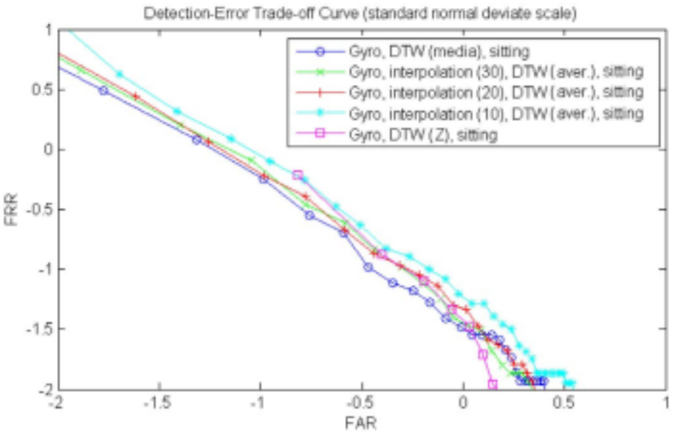

Fig. 8. DET curves for arm gesture (gyroscope data) related to the five best performing combination of features extraction/matching metrics.

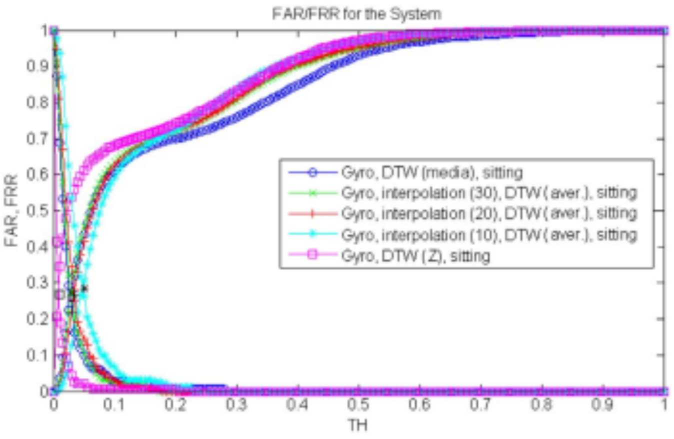

Fig. 9. FAR/FRR curves for arm gesture (gyroscope data) related to the five best performing combination of features extraction/matching metrics.

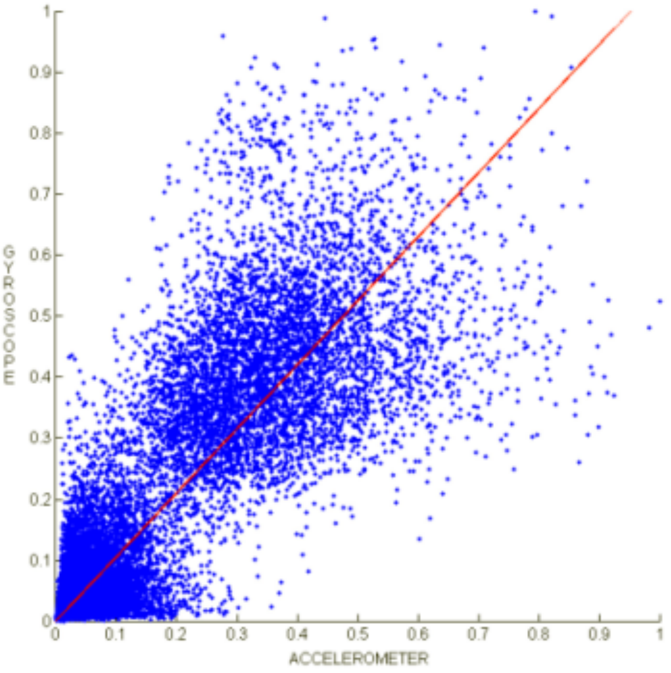

Fig. 10. Scatterplot of correlation among accelerometer and data and gyroscope data for arm gesture biometrics. In red the linear regression. The Pearson's correlation coefficient is 0.8556, so the two types of data are strongly interdependant.

TABLE IV
COMPARISON OF RESULTS FOR DIFFERENT (SCORE-LEVEL AND FEATURE LEVEL) ARM GESTURE DATA-FUSION STRATEGIES

Data-fusion methods (Acc+gyro data, DTW average, sitting)	**AUROC**	**EER**
score-level fusion (acc 0.9 / gyro 0.1)	0.9369	0.1384
score-level fusion (acc 0.8 / gyro 0.2)	0.9302	0.1392
feature-level fusion (concatenation)	0.9277	0.1548
score-level fusion (acc 0.7 / gyro 0.3)	0.9921	0.1407
score-level fusion (acc 0.8 / gyro 0.2)	0.9177	0.1607
feature-level fusion (sum)	0.9130	0.1670
score-level fusion (acc 0.5 / gyro 0.5)	0.9037	0.1701
score-level fusion (acc 0.4 / gyro 0.6)	0.8920	0.1865
score-level fusion (acc 0.3 / gyro 0.7)	0.8788	0.2106
score-level fusion (acc 0.2 / gyro 0.8)	0.8639	0.2199
score-level fusion (acc 0.1 / gyro 0.9)	0.8460	0.2410

is adopted (the higher the number of feature key-points, and the higher the performance). It is worth noting that, for the gyroscope, the z-axis is the more salient, whereas, for the accelerometer, the y-axis is the more salient. FFT-based features extraction results more performing in this case.

In this trial, the matching method used is invariably the DTW average, the best performing among the ones considered so far. The results are summarized in Table IV and graphically depicted by Figs. 11–13.

3) Accelerometer + Gyroscope Fusion: We also wanted to investigate the possibility of combining the contribution of accelerometer and gyroscope by fusing the corresponding captured data through feature-level and score-level strategies. A preliminary correlation analysis for the two types of data reports a Pearson's correlation coefficient of 0.8556 (see Fig. 10), but we wanted to assess if a (marginal) advantage would still be possible. The feature-level fusion scheme adopted is based on the sum and the concatenation of the two features vectors, while the score-level fusion scheme exploits variable weights for the accelerometer and the gyroscope components. The weights vary in the range 0.1–0.9 (e.g., Acc.*0.1 + Gyro*0.9) and their sum is always 1.

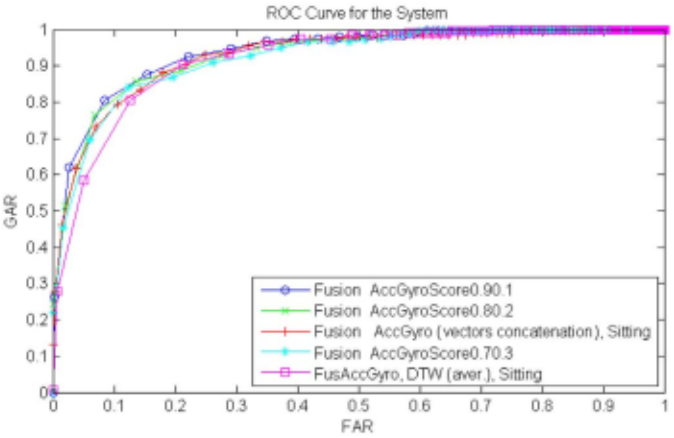

Fig. 11. ROC curves for arm gesture (accelerometer+gyroscope data related to the five best performing fusion schemes reported in Table IV.

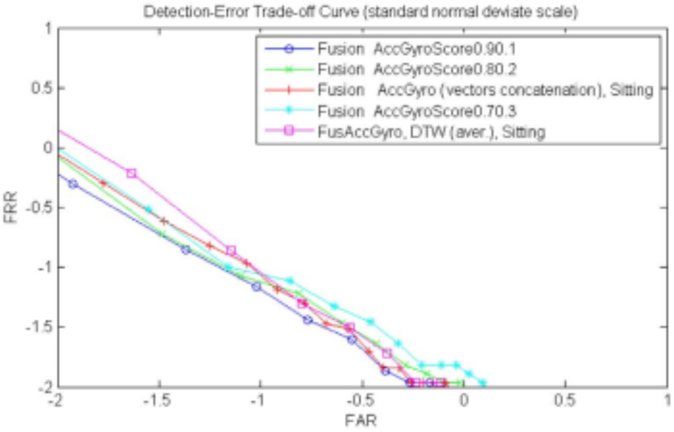

Fig. 12. DET curves for arm gesture (accelerometer+gyroscope data related to the five best performing fusion schemes reported in Table IV.

They show that feature-level fusion is less performing, with features-vector concatenation leading to a slightly better result (EER $=$ 0.1548) compared to features-vectors sum (EER $=$ 0.1607). Score-level fusion achieves the best result (EER $=$ 0.1384) when accelerometer weight is 0.9. However, even in this case it performs slightly worse than the accelerometer alone, so gyroscope data should be disregarded.

4) Robustness to Posing Variations: Since the arm gesture associated to the action of placing/responding to a call can be performed in different ways, for instance depending on the sitting or standing condition of a given subject, it is interesting to assess how these two main posing variations may possibly affect the authentication accuracy. To this aim, the best performing features matching metrics considered in previous evaluations (DTW average) is applied to the following three datasets.

1) dataset#1 containing all the raw accelerometer data captured in sitting condition.
2) dataset#2 containing all the raw accelerometer data captured in standing condition.
3) dataset#3 containing 50% of elements randomly extracted from dataset#1 (sitting) and the remaining 50% of elements randomly extracted from dataset#2 (standing).

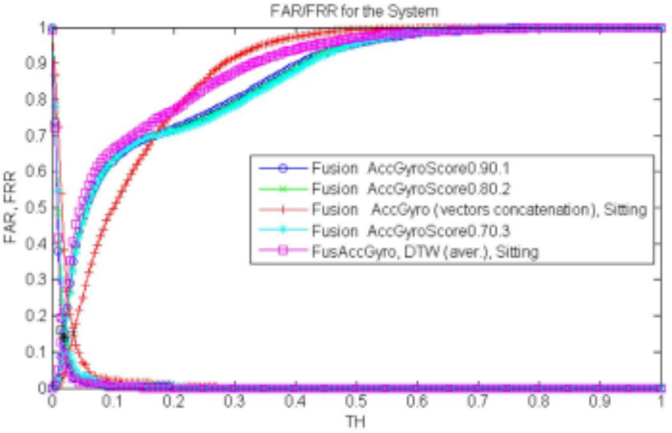

Fig. 13. FAR/FRR curves for arm gesture (accelerometer+gyroscope data related to the five best performing fusion schemes reported in Table IV.

TABLE V
COMPARISON OF DIFFERENT POSING CONDITIONS DURING ARM GESTURE CAPTURE

Posing variations (raw accelerometer data)	**AUROC**	**EER**
DTW average (standing)	0.9477	0.1163
DTW average (sitting)	0.9395	0.1305
DTW average (all)	0.8957	0.1872

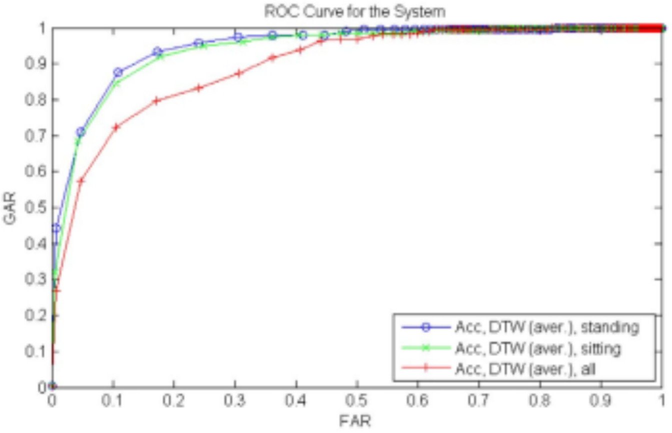

Fig. 14. ROC curves for arm gesture (raw accelerometer data) related to three different posing conditions during capture.

As reported in Table V and visually confirmed by Figs. 14–16, the system provides the best performance when the samples related to standing subjects are considered (EER $=$ 0.1163), arguably due to the greater freedom of motion leading to more salient data. This performance is not far from the best EER values achieved by [42] exploiting a 3-D method for gesture capture. The results related to dataset#3 (the most challenging), are still exploitable (EER $=$ 0.1872) thanks to a high correlation between the two poses.

C. Experiment #2—Ear

The second experiment is aimed at measuring the individual performance of the ear biometrics in the context of smartphone operated capture. The average EER achieved on 100 iterations of the experiment is 0.1774, with an AUROC of 0.8856

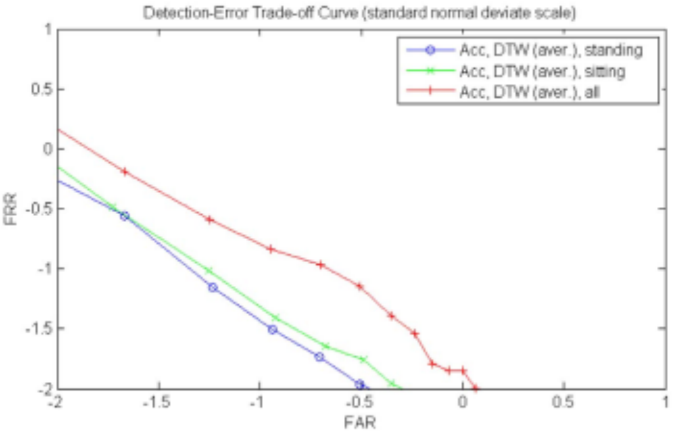

Fig. 15. DET curves for arm gesture (raw accelerometer data) related to three different posing conditions during capture.

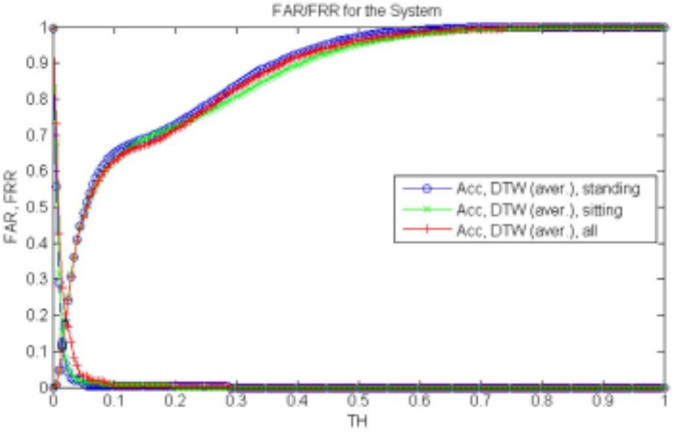

Fig. 16. FAR/FRR curves for arm gesture (raw accelerometer data) related to three different posing conditions during capture.

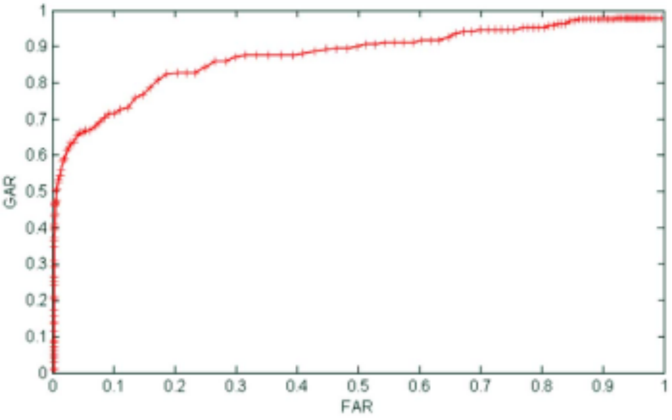

Fig. 17. ROC curve for ear captured through a smartphone. Ear features are represented by LBP descriptors and matched through Euclidean distance.

(see Figs. 17–19). These values, that can be considered somewhat suboptimal for typical ear recognition systems based on LBP descriptors, can be explained considering the rather challenging conditions in which ear images has been captured, sometimes leading to imperfect detection and cropping, or even to softer images due to residual motion-blur even in the best frame selected.

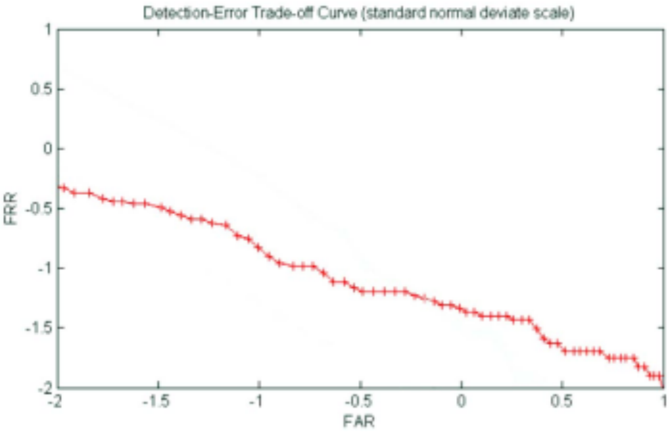

Fig. 18. DET curve for ear captured through a smartphone. Ear features are represented by LBP descriptors and matched through Euclidean distance.

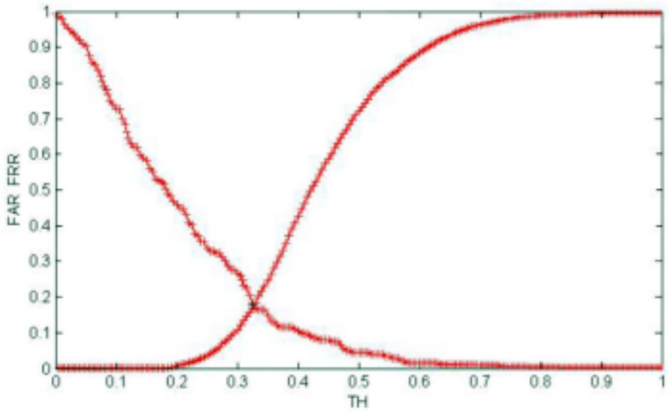

Fig. 19. FAR/FRR curve for ear captured through a smartphone. Ear features are represented by LBP descriptors and matched through Euclidean distance.

D. Experiment #3—Combining Arm Gesture and Ear

The third experiment is aimed at assessing the potential advantage in combining ear and arm gesture biometrics. We found that the Pearson's correlation index for these two biometrics is 0.3207 (see Fig. 20) implying a potential advantage in combining them. Two types of fusion schemes have been tested, with results shown in Table IV: 1) score-level fusion and 2) decision-level fusion.

Score weights vary from Ear*0.9/Arm*0.1 to Ear*0.1/Arm*0.9, while decision-level fusion may implement a logic AND strategy in which system authorizes the user only if both ear and arm subsystems authorize it, or a logic OR strategy, where the system authorizes the user if at least one of the two subsystems authorize it.

Features-level fusion has not been considered as a valid option, due to the strong nonhomogeneity of the two features-vectors (LBP is a 6400 8bit elements long vector, while accelerometer data are typically packed in a vector with average length of 240).

The results listed in Table VII and both the ROC, and FAR/FRR curves (see Figs. 21 and 22), confirm the validity of the multimodal approach, with score-level fusion providing the best EER of 0.1004 when the weight for arm-motion is 0.9 and

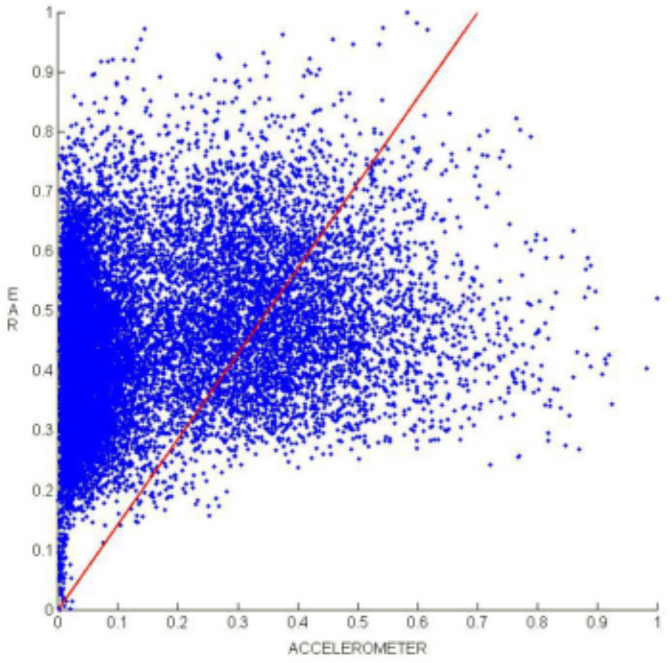

Fig. 20. Scatterplot of correlation among ear and arm gesture biometrics (accelerometer data). In red the linear regression. Since the Pearson's correlation coefficient is 0.3207, the two biometrics cannot be considered independent one from the other, but there is an advantage in combining them.

TABLE VI
COMPARISON OF RESULTS FOR DIFFERENT FUSION STRATEGIES

Fusion strategies (ear-arm)	**AUROC**	**EER**
ear-arm, score level fusion, weights (0.1/0.9)	0.9560	0.1004
ear-arm, score level fusion, weights (0.2/0.8)	0.9454	0.1204
ear-arm, decision-level fusion "OR"	0.9412	0.1279
ear-arm, score level fusion, weights (0.3/0.7)	0.9370	0.1241
ear-arm, score level fusion, weights (0.4/0.6)	0.9294	0.1378
ear-arm, score level fusion, weights (0.5/0.5)	0.9218	0.1493
ear-arm, score level fusion, weights (0.6/0.4)	0.9148	0.1552
ear-arm, score level fusion, weights (0.7/0.3)	0.9082	0.1603
ear-arm, score level fusion, weights (0.8/0.2)	0.9008	0.1627
ear-arm, score level fusion, weights (0.9/0.1)	0.8937	0.1691
ear-arm, decision-level fusion "AND"	0.8897	0.1717

TABLE VII
FINAL COMPARISON OF FIVE DIFFERENT APPROACHES

Approaches	**AUROC**	**EER**
score-level fusion (ear 0.1 / arm 0.9)	0.9560	0.1004
arm gesture accelerometer DTW average	0.9382	0.1315
score-level fusion (acc. 0.9 / gyro 0.1)	0.9369	0.1384
ear (LBP)	0.8856	0.1774
arm gesture gyroscope DTW average	0.8306	0.2611

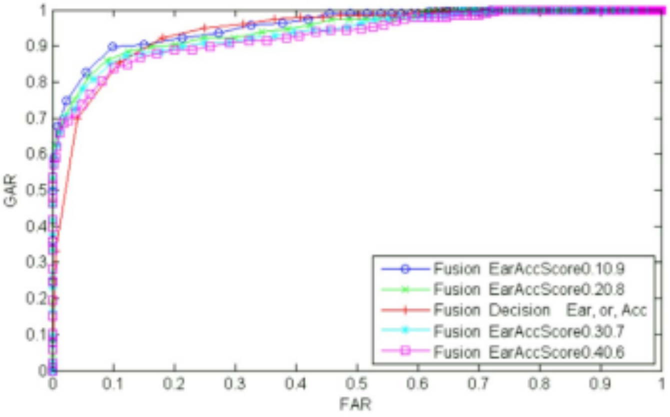

Fig. 21. ROC curves for the five best performing ear–arm fusion strategies.

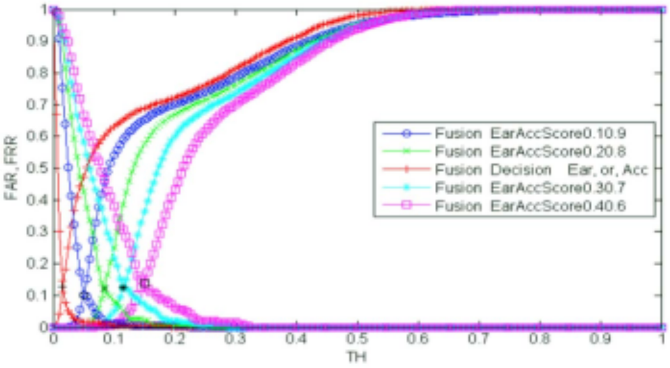

Fig. 22. FAR/FRR curves for the five best performing ear–arm fusion strategies.

0.1 for ear. Progressively reverting the weighting (i.e., increasing the weight of ear and decreasing the weight of arm-motion) produce a corresponding progressive decrease of performance.

This can be explained with the slightly inferior performance of the ear subsystem and also by comparing the FAR/FRR curves of both biometrics. The decision-level fusion schemes resulted less performing, though the OR version, with an EER of 0.1279, still improves over the best single biometrics alone (arm gesture). On the contrary, the AND scheme provides performance similar to the ear alone.

E. Experiment #4—System Applied to Identification Scenario

The last experiment we present, is aimed at assessing performance in the identification scenario (one-to-many comparison). Person identification does not represent the ideal application context for the proposed biometric system that is more suited to low/medium-security authentication according to the results described before. Nevertheless, we wanted to stress the system to eventually find its lower limit, so we compared the most effective fusion-schemes to each single biometrics by matching any element of the probe set to each of the elements of the gallery set. The ROC curve for this experiment is shown in Fig. 23 while, Table VIII resumes the results achieved, including an additional column reporting the rank-1 cumulative match score (CMS) to provide an intuitive figure of the system's recognition capability. The approaches are ordered according to decreasing value of $\text{CMS}_{\text{rank1}}$. Overall, the recognition accuracy is lower than the figures achieved for authentication, but still usable considering the maximum recognition rate of 80.5% achieved. Not surprisingly, in identification scenario the ear component is much more discriminant than arm gesture as proved by the performance related to the ear alone (RR = 73%) compared to arm gesture alone (RR = 58%). Consequently, the weights in

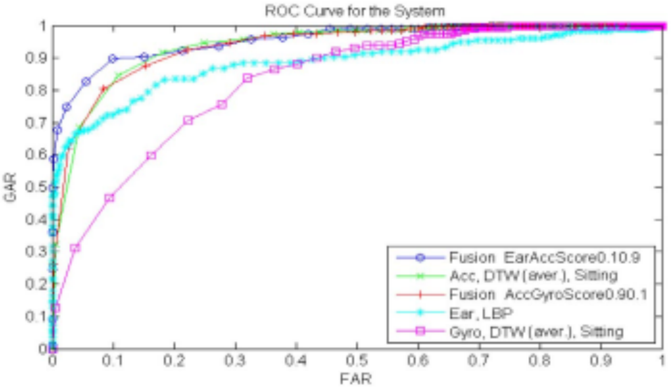

Fig. 23. ROC curves comparison for the five approaches resumed in Table VII.

TABLE VIII
IDENTIFICATION SCENARIO. COMPARISON OF FIVE APPROACHES

Identification Scenario	**AUROC**	**EER**	**$\text{CMS}_{\text{rank1}}$**
ear-arm, score-level fusion (ear 0.9 / acc. 0.1)	0.9571	0.0994	0.8050
ear (LBP)	0.8877	0.1747	0.7300
arm, score-level fusion (acc. 0.9 / gyro 0.1)	0.9371	0.1380	0.5950
arm, acc. DTW average	0.9398	0.1308	0.5800
arm, gyro DTW average	0.8322	0.2581	0.2200

the ear–arm score-level fusion are inverted (Ear 0.9/Acc. 0.1) with regard to those applied in the authentication scenario, where the arm gesture performed better indeed. This also implies that ear-matching metrics more performing than LBP would be likely to further increase overall system accuracy. It is also worth noting that for identification, like observed for authentication, the fusion of the two biometrics is beneficial, providing a measurable and not marginal performance improvement quantified in +0.069 for the AUROC, −0.075 for the EER, and +7.5% for the RR. Finally, since the use of DTW for matching arm-motion curves is robust and accurate but computationally expensive, it could represent a limiting factor in case of one-to-many comparisons as required by identification applications.

To this regard, it could be worth adopting more performing variations of the original algorithms like the Piecewise DTW [43] or the FastDTW [44] proposed for time series data mining.

V. CONCLUSION

In this paper, a multimodal biometric system aimed at ubiquitous person authentication by means of implicitly acquired biometrics has been described. The proposed system, exploits smartphone's sensors to capture both ear shape and arm motion when responding/placing a call, according to the hypothesis of a measurable advantage in combining a physical biometric identifier with a behavioral one.

The large number of experiments, conducted on a smartphone-captured multibiometrics database, was crucial to objectively assess the feasibility of arm gesture as a biometric and to measure the validity of its integration with ear not only in terms of convenience during the acquisition stage, but especially with regard to the accuracy achievable through their fusion. According to the best achieved EER values of 0.1 for the combined ear–arm and of 0.13 for the single arm gesture, the aforementioned assumption can be considered proved. This is even more significant considering that these results were all achieved on hardware surpassed by latest-generation mobile devices. However, there is still ample room for improvement in this system. Noise in arm gesture samples could be reduced through proper filtering (e.g., Kalman filter), hopefully resulting in improved verification accuracy, while smart context recognition and adaptive weighting strategies could dynamically modify the fusion weights according to the resting/walking user's status. Moreover, since we did not exploit orientation info provided by three axes magnetometer aboard most mobile devices, we are interested in assessing their possible contribution to arm gesture biometric in terms of both improved accuracy and robustness. Finally, a thorough experimentation of the whole system under totally uncontrolled conditions would definitively be one of the future directions of this paper.

ACKNOWLEDGMENT

The authors would like to thank the work done by C. Galdi and A. Ceruso during the experiments conducted in this paper.

REFERENCES

[1] C. Vivaracho-Pascual and J. Pascual-Gaspar, "On the use of mobile phones and biometrics for accessing restricted Web services," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 2, pp. 213–222, Mar. 2012.

[2] A. F. Abate, M. Nappi, and S. Ricciardi, "Smartphone enabled person authentication based on ear biometrics and arm gesture," in *Proc. IEEE Syst. Man Cybern. Conf. (SMC)*, Budapest, Hungary, 2016, pp. 3719–3724.

[3] N. L. Clarke and S. M. Furnell, "Advanced user authentication for mobile devices," *Comput. Security*, vol. 26, no. 2, pp. 109–119, 2007.

[4] K. M. Kramer, D. S. Hedin, and D. J. Rolkosky, "Smartphone based face recognition tool for the blind," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Buenos Aires, Argentina, 2010 pp. 4538–4541.

[5] K.-T. Cheng and Y.-C. Wang, "Using mobile GPU for general-purpose computing—A case study of face recognition on smartphones," in *Proc. Int. Symp. IEEE VLSI Design Autom. Test (VLSI DAT)*, Hsinchu, Taiwan, Apr. 2011, pp. 1–4.

[6] Y. Shen *et al.*, "Face recognition on smartphones via optimised sparse representation classification," in *Proc. 13th Int. Symp. IEEE Inf. Process. Sensor Netw.*, Berlin, Germany, 2014, pp. 237–248.

[7] P. A. Tresadern *et al.*, "Mobile biometrics (MoBio): Joint face and voice verification for a mobile platform," *IEEE Pervasive Comput.*, vol. 12, no. 1, pp. 79–87, Jan./Mar. 2013.

[8] T. K Mohanta and S. Mohapatra, "Development of multimodal biometric framework for smartphone authentication system," *Int. J. Comput. Appl.*, vol. 102, no. 7, pp. 6–11, 2014.

[9] C. Galdi, M. Nappi, and J.-L. Dugelay, "Multimodal authentication on smartphones: Combining iris and sensor recognition for a double check of user identity," *Pattern Recognit. Lett.*, vol. 82, pp. 144–153, Oct. 2016.

[10] C. Nickel, T. Wirtl, and C. Busch, "Authentication of smartphone users based on the way they walk using k-NN algorithm," in *Proc. 8th Int. Conf. IEEE Intell. Inf. Hiding Multimedia Signal Process. (IIH MSP)*, Pireas, Greece, 2012, pp. 16–20.

[11] D. Gafurov and E. Snekkkenes, "Arm swing as a weak biometric for unobtrusive user authentication," in *Proc. Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Harbin, China, Aug. 2008, pp. 1080–1087.

[12] X. Zhao, T. Feng, W. Shi, and I. A. Kakadiaris, "Mobile user authentication using statistical touch dynamics images," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 11, pp. 1780–1789, Nov. 2014.

[13] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 1, pp. 136–148, Jan. 2013.

[14] T. Feng, J. Yang, Z. Yan, E. M. Tapia, and W. Shi, "Tips: Context-aware implicit user identification using touch screen in uncontrolled environments," in *Proc. 15th Workshop Mobile Comput. Syst. Appl.*, Santa Barbara, CA, USA, Feb. 2014, Art. no. 9.

[15] J.-S. Kim, W. Jang, and Z. Bien, "A dynamic gesture recognition system for the Korean sign language (KSL)," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 26, no. 2, pp. 354–359, Apr. 1996.

[16] X. Zhang *et al.*, "Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors," in *Proc. 14th Int. Conf. Intell. User Interfaces*, Sanibel, FL, USA, 2009, pp. 401–406.

[17] X. Zhao, Z. Gao, T. Feng, S. Shah, and W. Shi, "Continuous fine-grained arm action recognition using motion spectrum mixture models," *Electron. Lett.*, vol. 50, no. 22, pp. 1633–1635, Oct. 2014.

[18] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan, "uWave: Accelerometer-based personalized gesture recognition and its applications," *Pervasive Mobile Comput.*, vol. 5, no. 6, pp. 657–675, 2009.

[19] D. A. Johnson and M. M. Trivedi, "Driving style recognition using a smartphone as a sensor platform," in *Proc. Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Washington, DC, USA, 2011, pp. 1609–1615.

[20] M. Conti, I. Zachia-Zlatea, and B. Crispo, "Mind how you answer me!: Transparently authenticating the user of a smartphone when answering or placing a call," in *Proc. ACM Symp. Inf. Comput. Commun. Security*, Hong Kong, 2011, pp. 249–259.

[21] A. F. P. Negara *et al.*, "Arm's flex when responding call for implicit user authentication in smartphone," *Int. J. Security Appl.*, vol. 6, no. 879, pp. 55–63, 2012.

[22] T. Feng, X. Zhao, and W. Shi, "Investigating mobile device picking-up motion as a novel biometric modality," in *Proc. Int. Conf. Biometr. Theory Appl. Syst. (BTAS)*, Arlington, TX, USA, 2013, pp. 1–6.

[23] J. D. Bustard and M. S. Nixon, "Toward unconstrained ear recognition from two-dimensional images," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 3, pp. 486–494, May 2010.

[24] M. De Marsico, M. Nappi, and D. Riccio, "HERO: Human ear recognition against occlusions," in *Proc. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, San Francisco, CA, USA, Jun. 2010, pp. 178–183.

[25] H. Chen and B. Bhanu, "Human ear recognition in 3D," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 718–737, Apr. 2007.

[26] J. Lei, X. You, and M. Abdel-Mottaleb, "Automatic ear landmark localization, segmentation, and pose classification in range images," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 2, pp. 165–176, Feb. 2016.

[27] P. N. A. Fahmi *et al.*, "Implicit authentication based on ear shape biometrics using smartphone camera during a call," in *Proc. IEEE Int. Conf. Syst. Man Cybern. (SMC)*, Seoul, South Korea, 2012, pp. 2272–2276.

[28] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. 2001, pp. I-511–I-518.

[29] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, "Local binary patterns for still images," in *Computer Vision Using Local Binary Patterns*. London, U.K.: Springer, 2011, pp. 13–47.

[30] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[31] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, Jun. 2007.

[32] X. Wang, H. Gong, H. Zhang, B. Li, and Z. Zhuang, "Palmprint identification using boosting local binary pattern," in *Proc. IEEE 18th Int. Conf. Pattern Recognit. (ICPR)*, vol. 3. Hong Kong, Aug. 2006, pp. 503–506.

[33] Y. Wang, Z.-C. Mu, and H. Zeng, "Block-based and multi-resolution methods for ear recognition using wavelet transform and uniform local binary patterns," in *Proc. 19th Int. Conf. Pattern Recognit. (ICPR)*, Tampa, FL, USA, Dec. 2008, pp. 1–4.

[34] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. KDD Workshop*, vol. 10. Seattle, WA, USA, 1994, pp. 359–370.

[35] M. Faundez-Zanuy, "On-line signature recognition based on VQ-DTW," *Pattern Recognit.*, vol. 40, no. 3, pp. 981–992, Mar. 2007.

[36] G. A. ten Holt, M. J. T. Reinders, and E. A. Hendriks, "Multi-dimensional dynamic time warping for gesture recognition," in *Proc. 13th Annu. Conf. Adv. School Comput. Imag.*, vol. 300. Jun. 2007.

[37] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach," *Biometrics*, vol. 44, no. 3, pp. 837–845, 1988.

[38] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proc. IEEE*, vol. 85, no. 1, pp. 6–23, Jan. 1997.

[39] A. Humm, J. Hennebert, and R. Ingold, "Combined handwriting and speech modalities for user authentication," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 39, no. 1, pp. 25–35, Jan. 2009.

[40] V. Chatzis, A. G. Bors, and I. Pitas, "Multimodal decision-level fusion for person authentication," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 29, no. 6, pp. 674–680, Nov. 1999.

[41] P. P. Paul, M. L. Gavrilova, and R. Alhajj, "Decision fusion for multimodal biometrics using social network analysis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 11, pp. 1522–1533, Nov. 2014.

[42] O. Mendels, H. Stern, and S. Berman, "User identification for home entertainment based on free-air hand motion signatures," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 11, pp. 1461–1473, Nov. 2014.

[43] E. J. Keogh and M. J. Pazzani, "Scaling up dynamic time warping for datamining applications," in *Proc. 6th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, Boston, MA, USA, Aug. 2000, pp. 285–289.

[44] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intell. Data Anal.*, vol. 11, no. 5, pp. 561–580, 2007.

Andrea F. Abate (M'12) received the Laurea (*cum laude*) degree in computer science from the University of Salerno, Salerno, Italy, in 1991, and the Ph.D. degree in applied mathematics and computer science from the University of Pisa, Pisa, Italy, in 1998.

He currently serves as an Associate Professor with the University of Salerno from 2006, where he is Team Leader of the Computer Graphics Laboratory. His current research interests include biometry, virtual/augmented/mixed reality, haptics, and human–computer interaction. He has authored many scientific papers published in scientific journals and proceedings of refereed international conferences and co-edited one book.

Dr. Abate is a member of the IEEE Haptics Technical Committee and a member of the the International Association for Pattern Recognition (GIRPR/IAPR).

Michele Nappi (M'05–SM'17) was born in Naples, Italy, in 1965. He received the Laurea *(cum laude)* degree in computer science from the University of Salerno, Salerno, Italy, in 1991, the M.Sc. degree in information and communication technology from I.I.A.S.S. "E.R. Caianiello," Salerno, and the Ph.D. degree in applied mathematics and computer science from the University of Padova, Padova, Italy.

He is currently an Associate Professor of Computer Science with the University of Salerno. He is a Team Leader of Biometric and Image Processing Laboratory, Fisciano, Italy. His current research interests include multibiometric systems, pattern recognition, image processing, compression and indexing, multimedia databases, human–computer interaction, and VR/AR. He has co-authored over 120 papers in international conference, peer review journals, and book chapters in the above fields.

Dr. Nappi was a recipient of several international awards for scientific and research activities. He is a President of the Italian Chapter of the IEEE Biometrics Council from 2015 to 2017, and a member of the International Association for Pattern Recognition.

Stefano Ricciardi (M'12) received the B.Sc. degree in computer science, the M.Sc. degree in informatics, and the Ph.D. degree in sciences and technologies of information, complex systems, and environment from the University of Salerno, Salerno, Italy.

He has been a Co-Founder/Owner of a videogame development team focused on 3-D sports simulations. He is currently an Assistant Professor with the Department of Biosciences, University of Molise, Campobasso, Italy. He has co-authored about 70 research papers, including international journals, book chapters, and conference proceedings. His current research interests include biometry, virtual and augmented/mixed reality, haptics systems, and human–computer interaction.

Dr. Ricciardi serves as an external Expert for the Research Executive Agency of the European Commission. He is a member of the the International Association for Pattern Recognition (GIRPR/IAPR).