



Unobtrusive multi-modal biometric recognition using activity-related signatures

A. Drosou^{1,2} G. Stavropoulos² D. Ioannidis² K. Moustakas² D. Tzovaras²

¹Department of Electrical Engineering, Imperial College London, London SW7 2AZ, UK

²Informatics and Telematics Institute, Hellas 57001, Thessaloniki, Greece

E-mail: drosou@iti.gr

Abstract: The present study proposes a novel multimodal biometrics framework for identity recognition and verification following the concept of the so called ‘on-the-move’ biometry, which sets as the final objective the non-stop authentication in an unobtrusive manner. Gait, that forms the major modality of the scheme, is complemented by new dynamic biometric signatures extracted from several activities performed by the user. Gait recognition is performed through a robust scheme that is based on geometric descriptors of gait energy images and is able to compensate for undesired gait behaviour like walking direction variations and stops. On the other hand, the biometric signatures, based on the user activities, are extracted by tracking of three points of interest and are seen to provide a powerful auxiliary biometric trait. Finally, score level fusion is performed and the experimental results illustrate that the proposed multimodal biometric scheme provides very promising results in realistic application scenarios.

1 Introduction

Biometrics have recently gained significant attention from researchers, while they have been rapidly developed for various commercial applications ranging from surveillance and access control against potential impostors [1] to medical analysis purposes [2]. A number of approaches have been described in the past attempting to fulfil the different requirements of each application, such as reliability, unobtrusiveness, permanence etc. Generally speaking, biometric methods are categorised to physiological and behavioural [3], depending on the type of used features.

On the one hand, physiological biometrics are based on both biological measurements and inherent characteristics of each individual. Fingerprint is a typical example of physiological biometric traits that is widely used in law enforcement for identifying criminals [4], whereas other recent applications are based on iris- [5] or face-identification [6]. Despite their high recognition performance, they all demonstrate a very restricted applicability to highly controlled environments.

On the other hand, behavioural biometrics are related to specific actions and the way that each person executes them. They can potentially allow the non-stop (on-the-move) authentication or even identification in an unobtrusive and transparent manner to the subject and become part of an ambient intelligence (AmI) environment. Behavioural biometrics are the newest technology in the field biometrics and they have yet to be researched in detail. Even if physiological biometrics are considered more robust and reliable, behavioural biometrics have the inherent advantage of being less obtrusive [3, 7].

Recent work and efforts on human recognition have shown that human behaviour (e.g. extraction of facial dynamics features [8]) and motion (e.g. human body shape dynamics during gait [9]) provide the potential of continuous on-the-move authentication, when considering activity-related signals.

1.1 Related work

Regarding gait recognition, significant advances have been lately achieved [9, 10]. Most of the recent gait analysis methods can be divided into two categories of complementary nature [11], the model-based and the feature-based ones.

Model-based approaches study static and dynamic body parameters of the human locomotion [12], like stride length, stride speed and cadence [13]. A noise resistant method has been presented in [14], whereby the model-based gait signature is extracted by applying Fourier series and temporal trait gathering techniques. In general, model-based approaches [12–14] create models of the human body from the input gait sequences. Previous work on these approaches has shown that they can guarantee good degrees of view- and scale-invariance. However, experimental evaluation in larger, publicly available databases is still required, to compare their performance to that of feature-based methods.

On the contrary, feature-based techniques do not rely on the assumption of any specific model of the human body for gait analysis. They usually employ simple methods, such as temporal correlation, linear time normalisation [15], full volumetric correlation on partitioned silhouette frames

[16] and dynamic time warping (DTW) [17]. In [16], the extraction of features was performed on whole silhouettes; in [10], an angular transform was applied on silhouette sequences, whereas in [18, 19] gait recognition based on hidden Markov models (HMM).

Contrary to the fronto-parallel view assumption or other view dependent approaches like [20], some recent approaches deal with non-canonical view gait recognition, or view-invariant recognition of gait sequences, including model-based schemes with self-camera calibration [21]. Similarly, view transformations based on planar silhouette approximation are presented in [22] while tracking of body parts' trajectories are studied in [23] that can be also used to reconstruct the articulated full body motion [24].

Extending the concept of behavioural biometrics, 'reach and interact' biometrics can also been thought as a specialisation of activity related biometrics [25]. In this concept, an interesting biometric characteristic can be the user's response to specific stimuli within the framework of an Aml environment.

Such an approach of using biometric traits from everyday activities for biometric authentication purposes has been presented in [26]. Apparently, the use of motion trajectories towards biometric recognition/authentication exhibited significant potential towards unobtrusive user authentication [27, 28].

It is a common place that the combination of more biometric traits of the same identity not only improves the recognition performance of the biometric systems, but also provides reduced discrimination to people, whose biometrics cannot be recorded well (i.e. owing to certain disabilities etc.). In particular, multimodal biometric systems, that capture a number of unrelated biometrics indicators, are seen to have several advantages over unimodal systems. Specifically, they are much more invulnerable to fraudulent technologies, since multiple biometric characteristics are more likely to resist to spoof attacks than a single one [29].

In this respect, a lot of work has been carried out in the last decade by the scientific community on multimodal biometrics. Recent research activities in multimodal biometrics evaluate the use of gait as a promising stand alone biometric modality or even in combination with other complimentary modalities [30] like face. A gait 'on-the-move' recognition system has been proposed in [31], whereby gait traits have been combined with face recognition in a controlled environment with fixed cameras. Other multimodal approaches have combined face images and speech signals [32, 33], while face and fingerprint have been combined [34]. In a similar manner, soft biometrics is fused with other biometrics such as face recognition [35] and fingerprint [36].

1.2 Contribution

The present paper proposes a novel scheme for the integration of two activity-related traits in a multimodal biometric recognition system, using a score level fusion of the individual modalities. The selected modalities are chosen so as to satisfy the unobtrusiveness constraints of the framework.

The novelty aspects introduced within the current work lie in the following issues. First, it is the first time that a multimodal biometric system is totally based on dynamic biometric traits that provide fully unobtrusive, on-the-move recognition, since nor are any sensors or markers attached to the users, neither are they asked to undergo any special recognition procedure, other than they would have normally

been doing. Last but not least, the proposed framework significantly provides added value in terms of invariance with respect to the environment, since no constraints in the users' movements, such as specific standing position or continuous walking on a fixed line, are set.

Regarding the gait recognition module, a feature-based recognition system is proposed that can handle realistic events, such as user stops and random walking paths. Thus, the gait system can be adopted for environments, where the user can freely move within the working space and perform everyday activities.

Further, the proposed activity related traits are associated with activities including reaching and interacting with objects. This concept is based on the assumption that each user has a characteristic way of reaching and interacting with objects, while performing specific activities [37]. The position of the body, and the relative movement of the head and the palms with respect to the object are analysed towards the extraction of unique signatures of dynamic nature, that form a reliable biometric signal for authentication, since it is more difficult to imitate one's behaviour during time than falsify a static biometric pattern once (e.g. fingerprint). The problem of small variances in the interaction setting, which may be introduced by the arbitrary position of the environmental objects with respect to the user at each trial is handled via a spatial warping method, which compensates for all small displacements of the environmental objects. Finally, fusion that is optimally parameterised by a genetic algorithm (GA) is applied at the score level.

Among others, the current work utilises the concept of the so-called 'on-the-move' biometry [38], which sets as the final objective the non-stop authentication in a very unobtrusive and transparent manner, where the user is not requested to perform any special action.

2 Overview of the proposed framework

In the current framework, a bi-modal system is presented that efficiently manages to fuse complementary information from two uncorrelated activity related traits.

2.1 Scenario

The application scenario expects that the user walks along a corridor in arbitrary walking paths, as shown in Fig. 1. In the middle of the path, there exists a control panel, where the user is supposed to stop, to insert his authorisation card and to type his personal pin. Then, the user continues his way to the door at the other end of the corridor. The whole scene is constantly recorded by two stereoscopic cameras, as shown in Fig. 1.

2.2 Proposed approach

The architecture of the proposed biometric recognition framework is illustrated in Fig. 2. Initially, the moving silhouettes are extracted from the captured gait image sequence, the shadows are removed and the gait cycle is estimated using state-of-the-art (SoA) algorithms [9, 39]. Using a stereoscopic camera, those frames in the sequence are initially detected, whereby the user is not walking, and then removed from those where the user is walking. Then the visual hull of the moving silhouette is extracted using disparity estimation. Once view normalisation is applied by rotating the silhouette, the 3D reconstructed silhouettes are

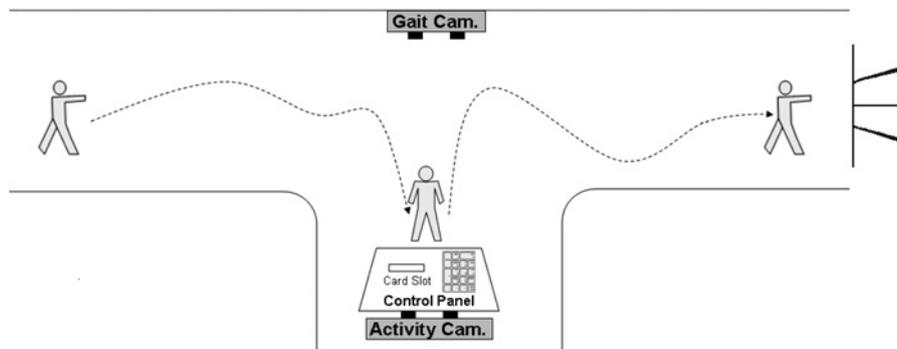


Fig. 1 Scenario setting

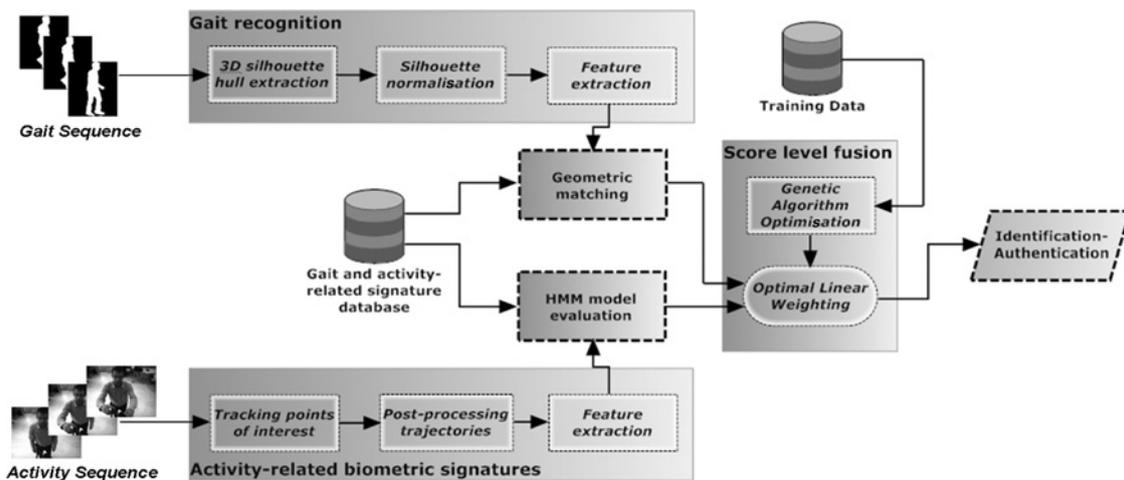


Fig. 2 Architecture of the proposed gait recognition framework

denoised via spatiotemporal filtering, to improve their quality. Finally, two novel geometric descriptors are extracted based on the sequence gait energy image (GEI).

In parallel, the user's activities are captured, and after tracking only specific points of interest, the extracted trajectories are post-processed, so as to filter out tracking errors, to smooth and homogenise them in terms of vector length. Finally, after they have been warped-normalised to a common reference, they are used as input to a HMM algorithm, either for training or for classification.

Moreover, a linear fusion mechanism is proposed, so as to perform score-level fusion of the two biometric traits incorporated. The optimal weights are selected utilising a GA instead of a typical Bayesian classifier owing to the lack of the knowledge of the distribution of the estimated distances. In the following section, the individual modules of the proposed framework are described in detail.

3 Gait recognition

A robust gait recognition framework, that employs novel methods for denoising and post-processing the gait silhouette sequences, is proposed in the context of the proposed framework. As of today, most of the automated gait recognition systems cope with various gait scenarios considering changes either in the subject appearance (e.g. carrying an object, clothing etc.) or in other covariate factors, such as time or view angles between the camera and the moving subject. However, in more complex environments (i.e. security control rooms, workplace

environments with authorised employees, such as nuclear plant personnel etc.), there is obvious need for continuous gait recognition systems that can handle difficult scenarios, such as sudden subject stops and engagement in various everyday activities. For these cases, online estimation of the gait direction with respect to the observing camera local coordinate system is needed. In the following paragraphs, the proposed event-based recognition approach is presented that can also compensate for significant variations in the walking direction between the testing and the stored sequences.

3.1 Detection of stops in a gait silhouette sequence

Initially, the walking human binary silhouette is extracted as described in [9]. Let I_i denote the i th binary human silhouette (second row in Fig. 3). To detect the stops during a gait sequence, motion estimation through the calculation of a motion history image (MHI) [40] is performed in the silhouette image sequence. Motion history template M_t at time instance t is estimated by counting the number of non-zero pixels in the difference image $D(I)$ of two sequential silhouette frames (I_i, I_{i-1}) , as indicated by (1).

$$M_t(x, y) = \begin{cases} b, & \text{if } D(I(x, y)) = 1 \\ \max(0, M_{t-1}(x, y) - 1), & \text{otherwise} \end{cases} \quad (1)$$

where in the context of the proposed framework the value of b is experimentally chosen to be $b = 2$.

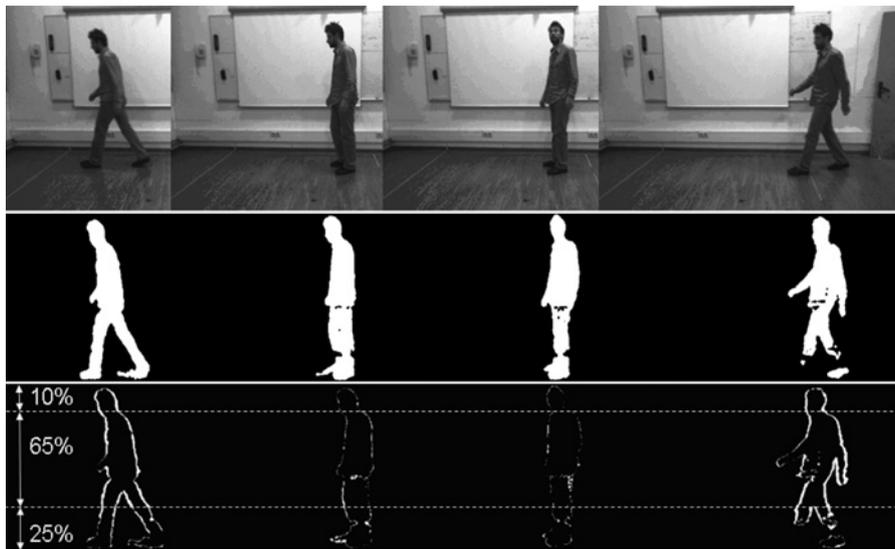


Fig. 3 First row: the user walks along the corridor, makes a short stop (two frames in the middle) and walks on – second row: silhouette extraction for the corresponding frames – third row: MHI of two sequential silhouette images

Area of interest is restricted to the lower 25% of the image height. The upper region which covers 10% of the image, is considered to include the head, which is used for the estimation of the walking angle

The recording phase starts with the detection of silhouette motion in the scene that is when the motion indicator function is over a fixed threshold ε_1

$$f_{\text{motion}} = \sum_{i=0}^{N_x} \sum_{j=0}^{N_y} M_t > \varepsilon_1 \quad (2)$$

whereby N_x and N_y are the resolution dimensions of the image M_t and ε_1 the noise threshold of a non-motion image. Similarly, stops in the user’s walking are detected, when the motion indicator f_{motion} regarding the lower 25% of the silhouette image height – the part of the legs below the knees [41] – falls below ε_2 (third row in Fig. 3). The values of both ε_1 and ε_2 have been experimentally defined.

Once the stop and (re)start frames are detected, the whole gait cycles that include stop frames are removed from the recorded sequence. Thus, a new set of silhouette sequence \tilde{I} is derived. In the following, the gait periods are extracted, as described in [9], and the gait cycles indexes are estimated accordingly.

3.2 Walking angle estimation and compensation

Let the term ‘gallery’ refer to the set of reference sequences, whereas the term ‘probe’ sequence stands for the test sequences to be verified or identified. As reported in the literature, the gait recognition systems achieve high recognition rates when the gallery and the probe sequences demonstrate similar walking angles [16], with respect to the observing camera local coordinate system. On the contrary, in cases whereby people walk with arbitrary view angles or different model-based types of angle transformations are applied [42]. However, the accuracy of angle view transformations at model-based approaches relies on small angle variations that are easily affected by slightly noisy images.

Thus, a novel feature-based method is introduced within the proposed framework that applies, prior to the feature extraction phase, 3D reconstruction on the silhouette itself,

encoding at the same time shape information about the user’s body. Specifically, range data are utilised for the compensation of angular variation in the walking direction.

The first step is to estimate the relative walking angle. The walking direction with respect to the camera (Fig. 4) can be estimated in a straight forward manner under the assumption of straight gait within each gait cycle. Given that the head of the silhouette image can be trivially detected within the highest part of the silhouette (Fig. 3), the gait direction v_1 in the 3D space can be explicitly estimated from the position of the subject’s head in the first h_0 and last frame h_L of the respective gait cycle $v_1 = h_0 - h_L$. It should be clarified that the variance of the walking direction within the same cycle is very rare in practice and thus, it is not taken into consideration in the current context.

Thus, the walking angle, which is considered constant through each gait cycle, is calculated using the equation below

$$\vartheta = \arccos\left(\frac{v_1 \bullet v_2}{|v_1||v_2|}\right) \quad (3)$$

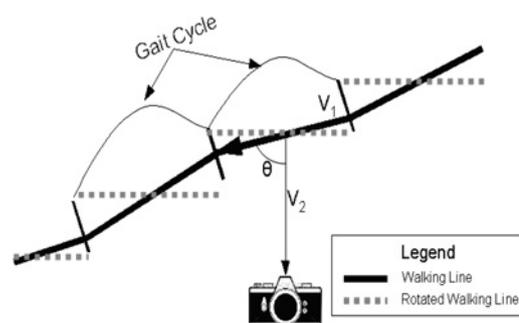


Fig. 4 Walking angle determination is calculated by the across of the inner product of the walking direction vector and the parallel to image plane vector

where \mathbf{v}_1 denotes the walking direction vector and \mathbf{v}_2 the parallel to the image plane vector.

After estimating the walking angle, the silhouettes are rotated, so as to register to the fronto-parallel view. This is achieved by extracting the 3D coordinates of each silhouette pixel, using the disparity data from the stereoscopic camera. This way, a 3D cloud of points \mathbf{p}_i is generated and their rotation is performed as follows

$$\tilde{\mathbf{p}}_i = \mathbf{p}_i \cdot \begin{bmatrix} \cos(\vartheta) & \sin(\vartheta) & 0 \\ -\sin(\vartheta) & \cos(\vartheta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

The points $\tilde{\mathbf{p}}_i$ of the rotated point cloud are now reprojected on the camera to create a new silhouette. The gait features are then extracted from the new set of silhouettes I' .

Despite the notable simplicity of (4), its direct application in the generation of the virtual view includes some inherent problems, that is the reconstructed point clouds could generate non-consistent surfaces, including holes and non-realistic edges, when projected on new virtual views (Fig. 5c). Figs. 5a and b depict the input and the depth silhouette of the user.

Therefore in the proposed framework a 3D surface is initially formed from the 3D point cloud, so as to generate a consistent surface and silhouette image in the synthesised virtual view (Fig. 5d). The surface is created using only a subset of the points of the image, so as to reduce the redundancy and size of the triangulated surface to be generated. Then, the silhouette for a particular view is generated by re-projecting it using the Z-buffering principle so as to rapidly perform depth culling in the new rendered image.

At this point, it should be noted that the acceptable changes $\Delta\theta$ in the user's waking angle are restricted within a range $-20^\circ \leq \Delta\theta \leq 20^\circ$ with respect to the front parallel view. This restriction is imposed by both the relatively coarse precision in the depth information provided by the stereoscopic camera, but also by the fact that for wider angle changes significant part of the user's body is occluded. In the same respect, it has been observed that the average gait cycle direction never exceeded an angle θ of 30° . Thus, whenever $\theta \geq 20^\circ$ the corresponding gait cycle was discarded.

3.3 Signature extraction and matching

The feature extraction process of the gait sequences is based on the radial integration transformation (RIT) and the Krawtchouk moments (KRM) [9]. These two features have

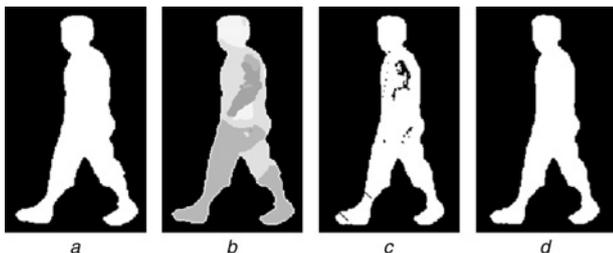


Fig. 5 Several types of extracted silhouette images

- a Original silhouette
- b Depth image of the silhouette
- c Rotated silhouette
- d Rotated silhouette after refinement

been utilised because of a series of advantages they exhibit. In particular, they provide compact feature representation of high discriminative power. Moreover, they are scale invariant, which means that the different resolutions of the recorded images in gallery and probe are allowed. Last but not least, RIT and KRM features provide extensive analysis of the users' biometric characteristics, by merging spatiotemporal information along with shape analysis of their body shape. Instead of applying those transforms on the binary silhouette sequences themselves, the GEI is utilised, which has been proven not only to achieve remarkable recognition performance, but also to speed up the gait recognition [43, 44].

Given the extracted binary gait silhouette images I' and the gait cycles, the GEI is defined over the k th gait cycle according to equation

$$\text{GEI}_k = \frac{1}{C_L} \cdot \sum_{j=n_0}^{n_0+C_L-1} I'_j \quad (5)$$

where n_0 and C_L are the first frame and the length of the current gait cycle, respectively, while k is an index to the gait cycles extracted in the current gait image sequence.

The RIT transform is applied on the GEI, to construct the gait template for each user, as shown below

$$\begin{aligned} \text{RIT}(t\Delta\theta) &= \frac{1}{J} \sum_{j=1}^J \text{GEI}_k(x_0 + j\Delta u \cdot \cos(t\Delta\theta), y_0 + j\Delta u \cdot \sin(t\Delta\theta)) \\ &\text{for } t=1, \dots, T \text{ with } T=360^\circ/\Delta\theta \end{aligned} \quad (6)$$

where Δu and $\Delta\theta$ are the constant step sizes of the distance u and angle θ , while J is the number of the pixels that coincide with the line that has orientation R and are positioned between the centre of gravity (x_0, y_0) of the silhouette and the end of the image in the direction of θ .

In the same respect, the $\text{KRM}_{n,m}$ of order $(n+m)$ transform is applied as follows

$$\begin{aligned} \text{KRM}_{n,m} &= \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} \bar{K}_n(x; p_1, N_x - 1) \\ &\cdot \bar{K}_m(y; p_2, N_y - 1) \cdot \text{GEI}_k(x, y) \end{aligned} \quad (7)$$

whereby $\bar{K}_n(x; p, N) = K_n(x; p, N) \sqrt{\frac{w(x; p, N)}{\rho(n; p, N)}}$.

Correspondingly, $K_n(x; p, N)$ are the Krawtchouk polynomials, while the variables $w(x; p, N)$ and $\rho(n; p, N)$ are given by the two following equations

$$w(x; p, N) = \binom{N}{x} p^x (1-p)^{N-x} \quad (8)$$

$$\rho(x; p, N) = (-1)^n \left(\frac{1-p}{p} \right)^n \frac{n!}{(-N)_n} \quad (9)$$

whereby the symbol $(-N)_n$ in is the Pochhammer symbol given by $(-N)_n = -N(-N+1)(-N+2) \dots (-N+n+1) = (\Gamma(-N+n))/(\Gamma(-N))$, whereby $\Gamma(n) = (n-1)!$ denotes the gamma function.

In the proposed framework the weighted 3D KRMs are estimated using the recurrent relations suggested in [45], since their direct estimation is of heavy computational complexity $O(n^6)$.

The comparison between the number of gallery G_{GEI} and probe P_{GEI} gait cycles for a specific feature $E \in \{\text{RIT}, \text{KRM}\}$ is performed through the dissimilarity score d_E .

$$d_E = \min_{i,j} (\|s_i^G - s_j^P\|) \forall i, j; i \in [1, G_{\text{GEI}}] \text{ and } j \in [1, P_{\text{GEI}}] \quad (10)$$

$\|\cdot\|$ is the L_2 -norm between the s^G and s^P values of the corresponding extracted feature for the gallery and the probe collections, respectively.

4 Activity-related recognition system

The proposed framework extends the applicability of activity-related biometric traits [26], and investigates their feasibility in user authentication applications.

In [25, 26], it is observed that the traits of a user's movements during an activity, that involves reaching and interacting with an environmental object, can be very characteristic for recognition of his/her identity. Indeed, given the major or minor physiological differences between users' bodies in combination with their individual inherent behavioural, habitual or stylish pattern of moving and acting, it has been reported that there is increased authentication potential in common everyday activities such as answering a phone call etc.

In the following, an improved activity-related recognition framework is proposed, that employs a novel method for the normalisation of the trajectories of the user's tracked points of interest. The proposed algorithm also introduces a warping method that compensates for small displacements of the environmental objects without affecting the behavioural information of the movement at all.

Contrary to [26], where a fixed environmental setting was assumed, in real life scenarios, significant performance degradations can be observed because of the small variances in the interaction setting. These variations are mainly introduced by the arbitrary position of the environmental objects with respect to the user at each trial. Thus, a post-processing algorithm towards the improvement of the overall authentication performance that can be employed into biometric systems, which utilise the reaching and interacting concept, is presented in the following.

4.1 Motion trajectory extraction

Following the scenario of Section 2.1, the user's movements are recorded by a stereoscopic camera that is placed on top of

the control panel and the raw captured images are processed, to track the users head and hands. Robust tracking (Fig. 6) is performed via the successive application of image masks on the captured image.

In particular, given the n th frame F^n of the recorded image sequence, a skin-colour mask $S(F^n)$ [46] combined with background extraction $B_{\text{head}}(F^n)$ with respect to the head's position can offer an initial approximation of the possible palms' location. The head can be efficiently tracked via the head detection algorithm based on [47] and mean-shift object tracking [48]. Given the pre-calibrated set of change-coupled device (CCD) sensors mounted on the stereo camera, the real depth information can be easily derived, first by performing disparity estimation from the input stereoscopic image sequence. Thus, it can be written that the derived filtered image $D(F^n)$ is given as $D(F^n) = S(F^n) \cap B_{\text{head}}(F^n)$.

Then, by defining as $M(F^n)$ the pixel-wise subtraction of the two sequential filtered images $D(F^n)$ and $D(F^{n+1})$: $M(F^n) \equiv D(F^n) - D(F^{n+1})$, the remaining blobs on the image I_f^n provide a good estimation of the palms' positions

$$I_f^n(x, y) = \begin{cases} 2, & \text{if } M(F^n(x, y)) = 1 \\ \max(0, I_f^{n-1}(x, y) - 1), & \text{otherwise} \end{cases} \quad (11)$$

After post-processing [26] that is applied on the raw tracked points, based on moving average window and Kalman filtering, smooth 3D motion trajectories (Fig. 7), which are then used as activity related biometric traits for proposed modality. The users' intra-similarity and inter-variance, as expressed by their motion trajectories are illustrated in Figs. 7 and 8, respectively. All drawn trajectories refer to a combined activity, which includes the inserting of a card (left hand) and the typing of a pin (right hand), while standing in front of a control panel.

A motion trajectory for a certain limb l (head or palms) is described in the current work as a 3D N -tuple vector $s_l(t) = (x_l(t), y_l(t), z_l(t))$ that corresponds to the x, y, z -axes location of limbs centre of gravity at each time instance t of an N -frame sequence. The x, y and z data of the trajectories s_l , are concatenated into a single vector and all vectors, produced by the limbs that take part in a specific activity C form the trajectory matrix S_C . Each repetition of the same activity by a user creates a new matrix. The set of matrices for each user for a specific activity are subsequently used to train a stochastic model for each class as explained in Section 4.3.

4.2 Trajectory warping

The environmental invariance of the extracted trajectories in slightly different interaction settings between separate trials



Fig. 6 Snapshots of the tracking during the performance of an activity

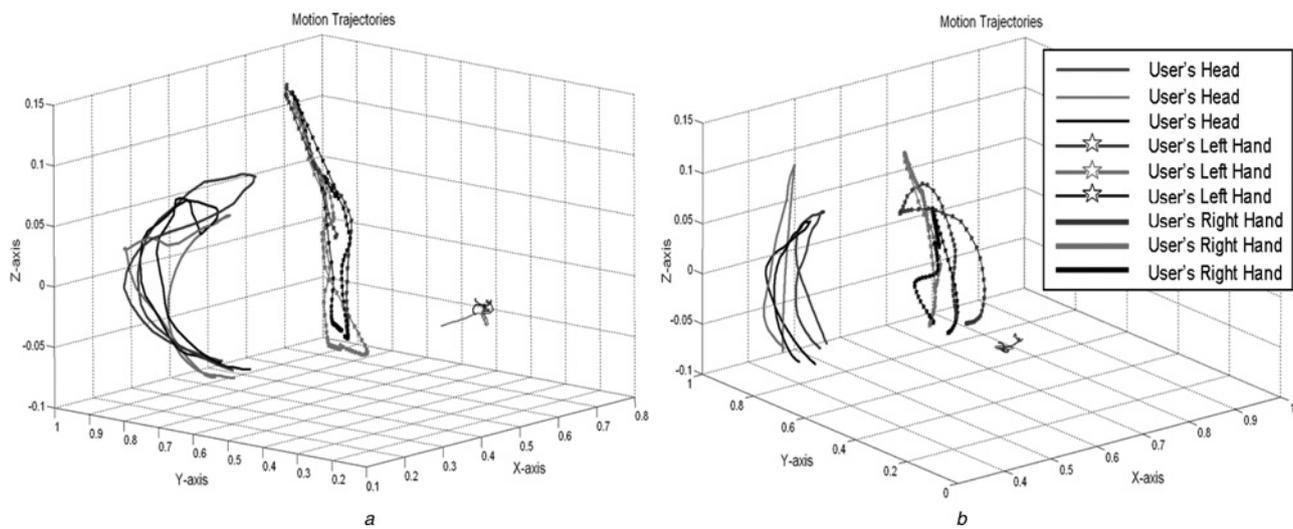


Fig. 7 Extracted motion trajectories from (a) User 1 and (b) User 2 during the combined movement of inserting a card and typing a pinword
 a Extracted motion trajectories of User 1 during the combined movement of inserting a card and typing a pinword
 b Extracted motion trajectories of User 2 during the combined movement of inserting a card and typing a pinword

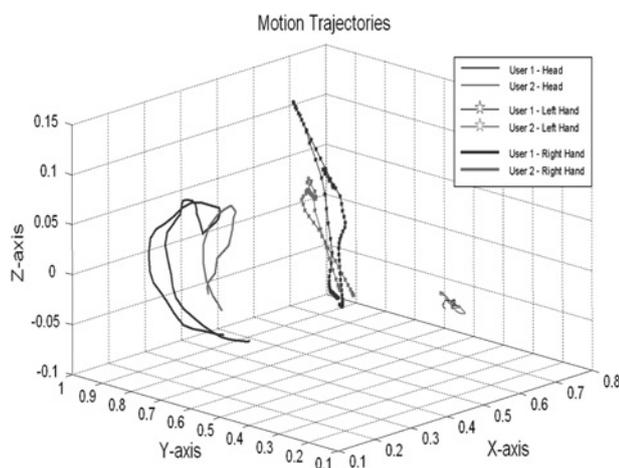


Fig. 8 Differentiations in the trajectories' shape between two different users

(different positions of the interaction objects) is of high importance, since it is very difficult and practically impossible to observe exactly the same relative user-object position between consecutive trials even for the same subject. Lack of environmental invariance could result in an increase in the false rejection rate.

The relative distance between the initial hand position and the panel is not expected to remain fixed, either owing to a shift of the users body or owing to small displacements of the panel. Thus, to provide enhanced invariance of the extracted trajectories, with respect to such environmental variables, the concept of spatial warping is introduced, following the principles of DTW [49].

Without loss of generality, the movement of the user's palms (end-effectors) is considered, while the panel remains fixed. In the current scenario, where the user is asked to type a pin on a panel (left hand) and to insert an authentication card (right hand), a starting and an ending position in the hand's movements is assumed (Fig. 9a). They can both be seen in Fig. 9b at location R_{panel} , when the user has touched the panel just before he starts typing and at location R_{Hand} and P'_{Hand} , when the hand hangs relaxed at the user's side in the enrollment and the classification stage, respectively. The

distance between these two 'extreme' spots may vary even between the same user from trial to trial, since it depends on the slight variations of the environmental setting. This can lead to unexpected deformations of the extracted motion trajectories, resulting to false rejections. However, since the main interest lies in the shape of the motion pattern of the trajectories and not in their length, the warping method is applied on the motion trajectories.

Specifically, the exact location of these two locations in the 3D space is automatically stored in the database for each user during the enrollment procedure. Accordingly, in the authentication process, the trajectory is warped to the environmental characteristics of the enrollment sequence.

In other words, the hand-to-panel distance d is used for the warping (stretching/compression) of an incoming set of trajectories, according to the claimed ID. Specifically, the solid line in Fig. 9 (right) indicates the actual extracted trajectory in the authentication stage. R_{Panel} and R_{Hand} are the stored locations of the user's head and the phone, respectively, obtained in the enrollment phase. The suggested method indicates that R_{Hand} and P'_{Hand} are mapped onto each other, whereas all other locations R_D of the XYZ signature in between are linearly transformed to the new location P'_D as

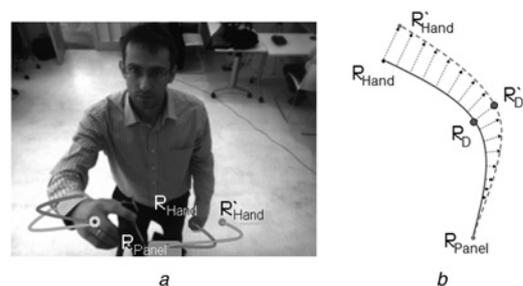


Fig. 9 Probe trajectories are mapped onto the gallery trajectories of the user, whose ID is claimed, towards enhanced invariance in respect to the environmental objects

a Different initial position of the user may result to longer or shorter motion trajectories

b Probe trajectories are mapped onto the gallery trajectories of the user, whose ID is claimed, towards enhanced invariance in respect to the environmental objects

indicated below by the normalisation factor q

$$R_{D'} = qP(D)$$

where

$$q = \frac{\|R_{\text{Hand}} - R_{\text{Panel}}\|}{\|R_{\text{Hand}'} - R_{\text{Panel}}\|}$$

4.3 Model-based recognition – HMMs

Given that the extracted trajectories $s_i(t)$ exhibit a strong dependence on temporal ordering, HMM have been utilised, as a multiple-state Gaussian Mixture Model (GMM) algorithm, for both the training and the authentication/recognition session of the current recognition module. A modelling alternative could have also been based on simple GMMs; however, GMMs have been proven inadequate for capturing the temporal relations and ordering of the successive limbs locations in the course of limb movements [27].

The first parameter specified for an HMM is the number of states. The number of states of the utilised HMM has been set equal to the mean number of changes in the direction of each of the palms and the head during the performance of any activity. Thus, a five-state, left-to-right, fully connected HMM is trained from several enrollment sessions of the same user for the given activity. Once the number of states is fixed, the complete set of model parameters describing the HMM is given by

$$\lambda = \{\pi_j, \alpha_{ij}, b_j\} \quad (12)$$

where π_j is the probability of the j th state being the first state among all the trajectories, α_{ij} denotes the probability of the j th state occurring immediately after the i th state and b_j denotes the probability density function (pdf) of the j th state.

Given that a single GMM $_z$ is completely specified by parameters $\Theta_z = \{w_k, \mu_k, \Sigma_k\}_{k=1}^K$, with mean vector μ_k and covariance matrix Σ_k , each GMM-based representation, used a state pdf that is calculated as

$$P(\mathcal{S}_C | \Theta_z) = \sum_{k=1}^K w_k \mathbb{N}(\mathcal{S}_C; \mu_k, \Sigma_k) \quad (13)$$

whereby \mathcal{S}_C denotes the input trajectory signal, K is the number of mixing Gaussian components, w_k are the mixing weights for which $\sum_{k=1}^K w_k = 1$, and $\mathbb{N}(\mathcal{S}_C; \mu_k, \Sigma_k)$ denotes the multivariate Gaussian function.

Then, the state variable q_t which corresponds to the t th state of the utilised HMM, takes one of T values $q_t \in \{s_1, \dots, s_T\}$. Since a Markovian process is assumed, the probability distribution of q_{t+1} depends only on q_t . This is described by the state transition probability matrix A whose elements a_{ij} represent the probability that q_{t+1} corresponds to state s_j given that q_t corresponds to s_i . The initial state probabilities are denoted by π_j – the probability that q_1 corresponds to state s_1 .

The observational data O_t from each state of the HMM are generated according to a PDF dependent on the instant of t th state, denoted by $b_j(O_t)$. This state-conditional observation

PDF is modelled as a Gaussian mixture as indicated below

$$b_j(O_t) = \sum_{z=1}^Z w_{jz} \frac{1}{(2\pi)^{p/2} |\Sigma_{jz}|^{1/2}} \times \exp\left\{-\frac{1}{2}(O - \mu_{jz})^T \Sigma_{jz}^{-1} (O - \mu_{jz})\right\} \quad (14)$$

whereby w_{jz} , μ_{jz} and Σ_{jz} denote the scalar mixing parameter, p -dimensional mean vector and $p \times p$ covariance matrix of the z th Gaussian component in the j th state. Here, each Gaussian component is a multivariate normal distribution of the same dimensionality, since all trajectories are described with three dimensions. The parameters of the HMM are initialised to random values and the Baum–Welch algorithm is used for estimation using the forward–backward procedure [50].

Once the training phase has been completed, new trajectories are categorised as one of the learned users for the specific activity based on the maximum likelihood criterion principle. Given the HMMs for the L enrolled subjects, $\lambda_1, \lambda_2, \dots, \lambda_L$, and a probe trajectory matrix $\mathcal{S}_C^l(t)$ of the incoming trajectory vectors from the new recording (i.e. the observation sequence) O_1, O_2, \dots, O_m , we assign user's label m as the HMM that maximises the likelihood [50]

$$m = \text{*arg max}_{i \in [1, \dots, L]} \sum_j P(O_{t+1:z} | q_t^i = j, O_{1:t}) P(q_t^i = j, O_{1:z}) \quad (15)$$

The above computation can be efficiently performed using the forward recursion procedure in the Baum–Welch algorithm [50]. The distance metric d_H for the current HMM classifier is defined as the value of the probability P_m that corresponds to the user's label m .

5 Multi-modal biometric fusion scheme

5.1 Datasets

The proposed methods were evaluated on two datasets: the proprietary 29-subject ACTIBIO-dataset and a 14-subject custom dataset. The ACTIBIO-dataset was captured in an AmI indoor environment and is extensively described in [26]. Both datasets include recordings from multiple repetitions of each subject performing the same activity (i.e. gait and activity). Regarding the 'reach and interact' scenario, the average frame rate at 15 fps for high resolution images ($1280 \times 960 \times 24$ BPP), while regarding the gait scenario the camera recorded images of $640 \times 480 \times 24$ BPP resolution with 48 fps.

The workplace recordings, used for gait recognition, include people walking in various paths within the environment, while performing various activities. The main course of walking is around 6 m and the distance from the stereoscopic camera varies from 2–6 m. The maximum detected intercycle angle differentiation with respect to the front-parallel view was found at 26° , while the intracycle walking angle variations ranged from 0° to 52° . Among other experiments recorded for 29 subjects, such as the 'normal', the 'briefcase/bag', the 'coat' experiments, the 'view-stop' condition is mobilised, whereby the subject performs a random path and stops to do specific work activities (e.g. operate the main room panel, press buttons etc.).

In the custom dataset, each of the 14-subjects included is walking again in a random path and stops for performing a dual activity. Specifically, each user should type a pin in a panel and then apply a card on a card reader. Both the gait sequences and the rest activities have been captured by stereo cameras. In this dataset, the size and quality of the gait recording were identical to the ACTIBIO-dataset, whereas the recorded images were of lower resolution ($320 \times 240 \times 24$ BPP) with 15 fps regarding the ‘reach and interact’ scenario.

Unfortunately, it was not possible to benchmark the proposed algorithms on public databases, since they do not include depth data.

5.2 Fusion algorithm

Towards fusion, the scores are normalised to a common basis according to the following equation

$$d^{\text{norm}} = \left(\frac{0.5}{T_L}\right) e^{-(d/d^{\text{max}})} \quad (16)$$

where d^{norm} is the normalised score value, d the non-normalised score, d^{max} the maximum possible score value and T_L an experimentally set threshold for the modality L . Variable d refers to both d_E for the RIT and KRM, as well as for d_H for HMM classification scores.

Given the absence of a priori knowledge regarding the distribution of the estimated similarity scores, a GA has been utilised to do the fusion between the proposed biometric modalities. In general, GAs are very efficient optimisation methods, since they are capable of detecting near global optimum solutions without the need of a priori knowledge of the premise space and of any non-convexities within it. Thus, to optimise the performance of the multimodal gait biometric system and supplementary fuse the activity-related biometric traits, the GA described in [51] is utilised for the estimation of the optimal weights for the three biometric descriptors.

In particular, the optimal weights used for score fusion based on a simple weighted averaging scheme are estimated using the GA described in Appendix. For the training of the fusion algorithm the 14-subjects custom dataset (Section 5.1) has been utilised. Specifically, the used gallery and the corresponding probe sequences stem from different repetitions than the ones that have later been used for the actual recognition purposes.

The experimental tests resulted in the following optimal weighted values

$$w_{\text{RIT}} = 0.34075, \quad w_{\text{KRM}} = 0.21425, \quad w_{\text{HMM}} = 0.445 \quad (17)$$

The final weighted distance between the probe x and the gallery y is estimated as $D_{\text{total}}(x, y) = 1/\text{Sim}(x, y)$, whereby $\text{Sim}(x, y)$ is defined as

$$\begin{aligned} \text{Sim}(x, y) &= \sum_{n \in T} \frac{w_n}{D_n} \\ &= \frac{w_{\text{RIT}}}{D_{\text{RIT}}(x, y)} + \frac{w_{\text{KRM}}}{D_{\text{KRM}}(x, y)} + \frac{w_{\text{HMM}}}{D_{\text{HMM}}(x, y)} \end{aligned} \quad (18)$$

whereby x ranges from 1 to N_p number of probes to identify, y denotes all the subjects in the training database $y = \{1, \dots, N_G\}$ and $D_n(x, y) = 1/\text{Sim}_n(x, y)$ denotes the total dissimilarity, between the probe subject x and the gallery

subject y given the feature set $n \in E^{\text{full}}$, where $E^{\text{full}} = \{\text{RIT}, \text{KRM}, \text{HMM}\}$.

The proposed fusion method is only used to estimate the optimal weights once and then the trained algorithm is applied as is, for the online identification of individuals with no further training or altering of the weights.

6 Experimental results

6.1 Experiments

The experiments that have been carried out on both databases include subjects walking in random paths with a stop for performing a specific activity. All users walk along a corridor in random paths, until they stop for interaction with a device (e.g. to type a password on a panel or to speak to a microphone-bell on the wall), as shown in Fig. 6, and then they walk away to any direction.

As far as the gait recognition in the 29-subject dataset is concerned, the ‘normal’ set that corresponds to fronto-parallel sequences has been used as the gallery set, while the ‘view-stop’ condition scenario is utilised for the authentication measurements. The activity selected from the same database is the ‘talking to a microphone panel’.

6.2 Gait recognition results

The pixel-wise differences in the extracted GEI images between the non-stop-detection approach and the proposed framework are demonstrated in Fig. 10. The reader can notice the significant denaturation of the GEI image in the absence of the stop detection, owing to the contribution of those frames, whereby the user has been standing still. The proposed gait recognition system has been tested on the 29-subject ACTIBIO dataset.

The improvements of the proposed gait recognition modality (i.e. stop detection and silhouette rotation) when the RIT and KRM algorithms are utilised as classifiers can be seen in Fig. 11.

Specifically, the reader can notice the significant contribution of the rotation-algorithm to the method proposed in [9]. In particular, the identification rates (red line in Fig. 11) are increased by a mean ratio of 20% (peek ratio improvement 35%) in the case of the RIT-classifier. Similarly, as far as KRM features are concerned, an improvement of a mean ratio of 10% (peek ratio

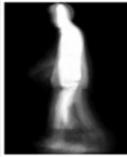
	Gallery GEI	Probe GEI	Absolute Difference	Difference Percentage
Without Stop Detection				6.50%
With Stop Detection				3.59%

Fig. 10 First row: great variations between the gallery and the probe even between a client user, when stop detection is disabled; Second row: low denaturation rate of the probe GEI, when stop detection is enabled at the probe sample

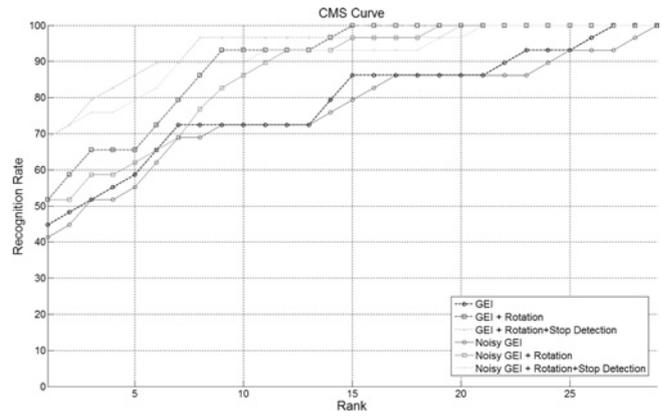
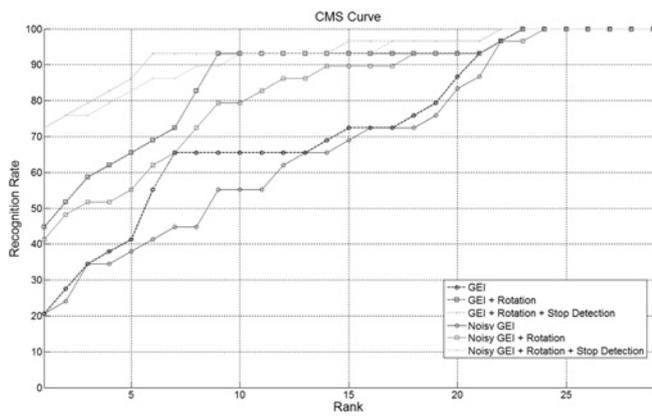


Fig. 11 Improvements in gait because of silhouette rotation and stop detection algorithm (29 subjects) – left: (RIT classifier)/right: (KRM classifier)

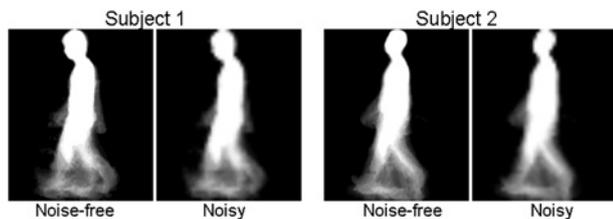


Fig. 12 Noise free versus noisy (PSNR = 24.1237 dB) silhouettes

improvement 23%) can be observed. In addition, when stop detection algorithm was enabled, the identification rates increased even more by a mean ratio value of 25 and 20%, in both the RIT and KRM classifier cases, respectively.

The proposed algorithms have been also tested in terms of their resistance against noise. For this reason, additional white Gaussian noise (AWGN) with a peak signal-to-noise ratio (PSNR) of 24.1237 dB has been added to the extracted gait silhouettes, by the successive down-scaling to 25% of the original size of their resolution and up-scaling them back [52], prior to the generation of each GEI (see Fig. 12). The derived results (Fig. 11) caused only a small degradation in the module’s recognition performance, which proved the robustness of the proposed approach under noisy environments.

In the same respect, the proposed enhancements exhibit significant improvements regarding the authentication performance of the gait module, as indicated by the equal error rate (EER) results in Table 1. Similarly, the degradation caused by noised insertion can be considered rather low.

6.3 Activity-related recognition results

The contribution to the framework of [26], regarding the activity-related biometric trait, is presented here. Given the 29-subject database, the corresponding improvements in the recognition performance, as well as in the EER score can be seen in Fig. 13 and in Table 2, respectively. The reader can notice an improvement of more than 25% in the

Rank-1 identification rate, and a simultaneous fall of 1.5% in the EER authentication score of the trajectory-based, activity-related system.

The improvements exhibited in the current activity trajectory-based module can be justified because of the enhanced invariancy added with respect to the small arbitrary environmental variations in the interaction setting (i.e. slight displacements of the environmental objects, different stop position of the user etc.).

The robustness of the proposed system against noise has been extensively tested in [26], whereby only a slight overall degradation of 1.5% has been noticed.

6.4 Fusion results

The score level fusion between the three classifiers (RIT–KRM–HMM) is performed, as described in Section 5. The recognition and verification performance of the final improved multimodal system, as they have been derived from tests carried out on the 29-subject ACTIBIO database, can be seen in Fig. 14a and in Table 3, respectively.

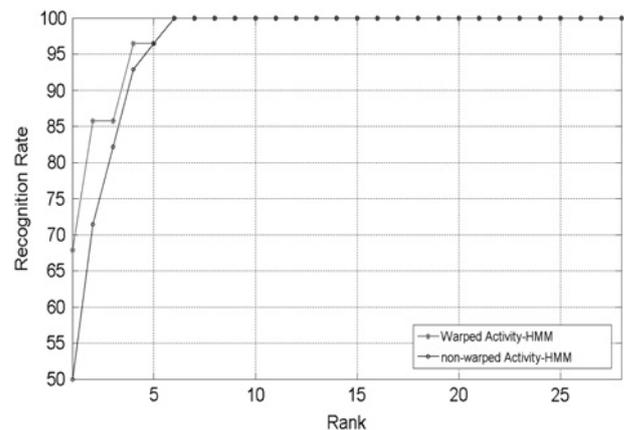


Fig. 13 Improvements in activity-related recognition modality because of the warping algorithm (29 subjects)

Table 1 Activity (ACTIBIO dataset) – EERs

	RIT, %	KRM, %	RIT (AWGN), %	KRM (AWGN), %
EER (29 subjects)	15.9	16.5	16.8	17.7

Table 2 Activity (ACTIBIO dataset) – EERs

	Non-warped trajectory, %	Warped trajectory, %
EER (29 subjects)	14.7	13.3

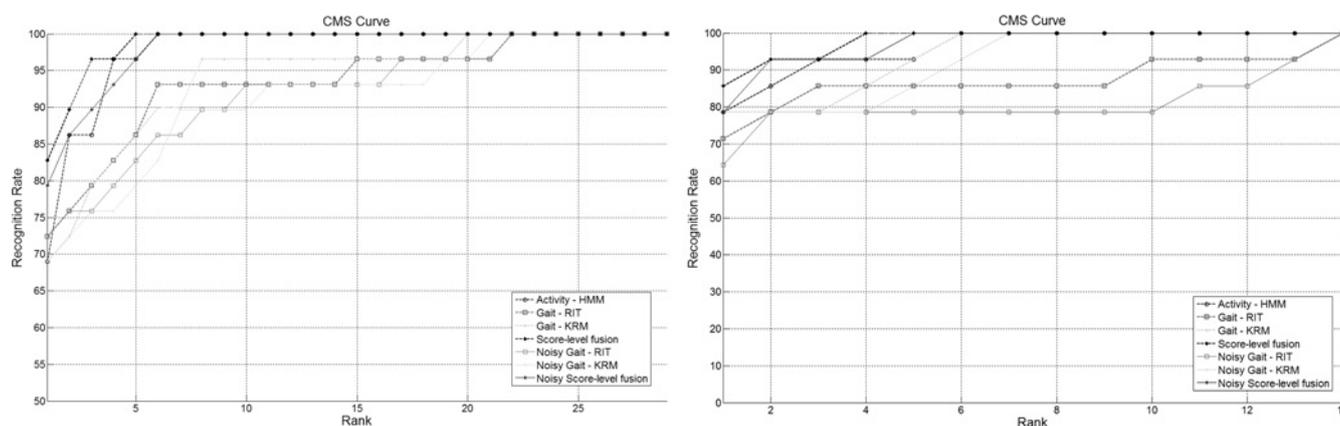


Fig. 14 CMS diagram of the final multimodal system – left: (29-subjects dataset)/right: (14-subjects dataset)

Table 3 Multimodal system (ACTIBIO dataset) – EERs

	RIT, %	KRM, %	HMM, %	Fusion, %	Fusion (noise), %
EER (29 subjects)	15.9	16.5	13.3	9.6	10.8
EER (14 subjects)	14.3	17.2	12.1	8.9	9.7

Specifically, the significantly increased Rank-1 identification rate of the multimodal system has reached a score of 83%, while at Rank-5 the identification rate has correctly recognised all the users. Additionally, the score-level combination of the two activity related traits (i.e. trajectory-based activity recognition and gait recognition) has managed to decrease the overall EER score of the system to 9% as indicated in the last column of the Table 3. In the same respect, the system exhibited strong resilience in both authentication and identification performance, even during the ‘noisy’ experiment, as shown in the Fig. 14 and in Table 3.

Similarly, the corresponding CMS curves and the EER scores for the custom dataset including 14-subject are depicted in Fig. 14b, and along the second row of Table 3, respectively.

The utilisation of the GA algorithm (see Appendix), towards weighted fusion, has driven to an overall performance improvement of the system of 5%, compared with the case, where uniformly distributed weights ($w_{RIT} = 0.33$, $w_{KRM} = 0.33$, $w_{HMM} = 0.33$) have been assigned to each of the derived modality scores.

6.5 Discussion

Given that no human performs exactly the same activity identically twice, it can be claimed that the proposed enhancements provided additional invariance to the extracted features of each of the described recognition modules. Specifically, arbitrary, not deterministic, slight changes in the walking angles along the same path or even different standing distances, in respect to the control panel can now be compensated towards more robust ‘on-the-move’ recognition performance.

In addition, the identification performance and the authentication EER of the proposed multimodal system are, as expected, significantly improved compared with the recognition potential of each single biometric trait. Despite the fact that such an improvement is relatively expected, the tuning of a multimodal system is not always a trivial task. In this context, apart from the utilised GA, great

improvement has also been imposed by the fact that the proposed biometric traits are totally uncorrelated.

Moreover, the performed experiments proved the system’s resilience against noisy environments or low quality recordings (tested with PSNR = 24.1237 dB). Regarding the noise suppression or the compensation for occluded information, that is introduced by rotation, the proposed algorithm has shown significant improvements within a range of 20° with respect to the front-parallel view. For wider walking angles a degradation of the recognition performance has been noticed and thus, the corresponding gait cycles were discarded.

To handle larger changes in the users’ walking direction, a multi-camera system is suggested. Specifically, a set of at least five cameras, whereby each camera would cover an angle of 30°. The presented algorithm would be significantly augmented by the utilisation of a time-of-light (ToF) camera, which would offer high accuracy resolution images.

Regarding the ‘reach and interact’ activity-related scenario, it has been observed that there is an upper bound in recognition potential provided by the three-point tracking. A full body model would significantly improve the recognition performance, since it would take into account not only dynamic information about the users’ movements, but also their static anthropometric profile. However, such a body model would require high accuracy and real-time processing, that is not yet available in SoA.

Despite the fact that no comparison with SoA methods was able to be performed, owing to lack of public databases that include depth data, it has been shown that the proposed preprocessing algorithms significantly improve the current recognition methods. Thus, it is expected to offer advantages if they are applied to any gait or activity-related recognition algorithm.

7 Conclusion

In this paper, a multimodal biometrics scheme that is based on two unobtrusive modalities is presented. Gait, that forms the major modality of the scheme, is complemented by new

dynamic biometric signatures extracted from several activities performed by the user. The proposed framework is seen experimentally to provide very promising recognition and verification rates, even under noisy environments. Moreover, taking also into account that no hard constraints are forced during the capture of the input signals, the proposed approach makes a step forward in the context of the very challenging problem of unobtrusive on-the-move-biometry.

8 Acknowledgment

This work was supported by the EU funded ACTIBIO ICT STREP (FP7-215372).

9 References

- Xiao, Q.: 'Security issues in biometric authentication'. Information Assurance Workshop, IAW 2005, 2005, pp. 8–13
- Rubenstein, L.Z., Solomon, D.H., Roth, C.P., *et al.*: 'Detection and management of falls and instability in vulnerable elders by community physicians', *J. Am. Geriatr. Soc.*, 2004, **52**, pp. 1527–1531
- Jain, A.K., Ross, A., Prabhakar, S.: 'An introduction to biometric recognition', *IEEE Trans. Circuits Syst. Video Technol.*, 2004, **14**, (1), pp. 4–20
- Maltoni, D., Jain, A., Prabhakar, S.: 'Handbook of fingerprint recognition' (Springer Professional Computing, 2009, 2nd edn.)
- Chowhan, S., Shinde, G.N.: 'Iris biometrics recognition application in security management', *Congr. Image Signal Process. (CISP)*, 2008, **1**, pp. 661–665
- Chang, K.I., Bowyer, K.W., Flynn, P.J.: 'An evaluation of multimodal 2D + 3D face biometrics', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, pp. 619–624
- Delac, K., Grgic, M.: 'A survey of biometric recognition methods'. Proc. Elmar 2004, 46th Int. Symp. Electronics in Marine, 2004, pp. 184–193
- Liu, X., Chen, T.: 'Video-based face recognition using adaptive hidden markov models'. IEEE Proc. Computer Society Conf. Computer Vision and Pattern Recognition (CVPR), 2003, vol. 1, pp. 340–345
- Ioannidis, D., Tzovaras, D., Damousis, I.G., Argyropoulos, S., Moustakas, K.: 'Gait recognition using compact feature extraction transforms and depth information', *IEEE Trans. Inf. Forensics Sec.*, 2007, **2**, (3), pp. 623–630
- Boulgouris, N., Chi, Z.: 'Gait recognition using radon transform and linear discriminant analysis', *IEEE Trans. Image Process.*, 2007, **16**, (3), pp. 731–740
- Veeraraghavan, A., Chowdhury, A.R., Chellappa, R.: 'Role of shape and kinematics in human movement analysis'. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'04), vol. 1, pp. 730–737
- Wagg, D.K., Nixon, M.S.: 'On automated model-based extraction and analysis of gait'. Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition, 2004, pp. 11–16
- Wang, L., Ning, H., Tan, T., Hu, W.: 'Fusion of static and dynamic body biometrics for gait recognition', *IEEE Trans. Circuits Syst. Video Technol.*, 2003, **14**, (2), pp. 149–158
- Cunado, D., Nixon, M., Carter, J.N.: 'Automatic extraction and description of human gait models for recognition purposes', *Comput. Vision Image Underst.*, 2003, **90**, pp. 1–41
- Boulgouris, N., Plataniotis, K., Hatzinakos, D.: 'Gait recognition using linear time normalization', *Pattern Recognit.*, 2006, **39**, pp. 969–979
- Sarkar, S., Phillips, P.J., Liu, Z., Robledo-Vega, I., Grother, P., Bowyer, K.W.: 'The human ID gait challenge problem: data sets, performance, and analysis', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2007, (2), pp. 162–177
- Boulgouris, N., Plataniotis, K., Hatzinakos, D.: 'Gait recognition using dynamic time warping'. IEEE Sixth Workshop on Multimedia Signal Processing, Siena, 2004, pp. 263–266
- Changhong, C., Jimin, L., Heng, Z., Haihong, H., Jie, T.: 'Factorial HMM and parallel HMM for gait recognition', *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.*, 2009, (1), pp. 114–123
- Liu, Z., Sarkar, S.: 'Improved gait recognition by gait dynamics normalization', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006, **28**, (6), pp. 863–876
- Goffredo, M., Carter, J.N., Nixon, M.: 'Front-view gait recognition'. IEEE Second Int. Conf. Biometrics: Theory, Applications and Systems (BTAS'08), 2008
- Goffredo, M., Seely, R.D., Carter, J.N., Nixon, M.S.: 'Markerless view independent gait analysis with self-camera calibration'. IEEE Int. Conf. Automatic Face and Gesture Recognition, 2008
- Kale, A., Chowdhury, A., Chellappa, R.: 'Towards a view invariant gait recognition algorithm'. Proc. IEEE Conf. Advanced Video and Signal Based Surveillance, 2003, pp. 143–150
- Jean, F., Albu, A.B., Bergevin, R.: 'Towards view-invariant gait modeling: computing view-normalized body part trajectories', *Pattern Recognit.*, 2009, **42**, (11), pp. 2936–2949
- Spencer, N., Carter, J.: 'Towards pose invariant gait reconstruction'. Proc. IEEE Int. Conf. Image Processing, ICIP 2005, 2005, pp. 261–264
- Kale, A., Cuntoor, N., Chellappa, R.: 'A framework for activity-specific human identification'. IEEE Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), 2002, vol. 4, pp. 3660–3663
- Drosou, A., Moustakas, K., Ioannidis, D., Tzovaras, D.: 'On the potential of activity-related recognition'. Int. Joint Conf. Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP 2010), 2010
- Drosou, A., Moustakas, K., Ioannidis, D., Tzovaras, D.: 'Activity related biometrics based on motion trajectories', in Broemme, A., Busch, C. (Eds.): 'BIOSIG 2010: biometrics and electronic signatures' (GI Edition, 2010), pp. 127–132
- Drosou, A., Moustakas, K., Tzovaras, D.: 'Event-based unobtrusive authentication using multi-view image sequences'. Proc. ACM Multimedia/Artemis Workshop ARTEMIS'10, Florence, 2010, pp. 69–74
- Fairhurst, M., Deravi, F., Mavity, N., George, J., Sirlantzis, K.: 'Intelligent management of multimodal biometric transactions' (Springer, Berlin-Heidelberg, 2003)
- Moustakas, K., Tzovaras, D., Stavropoulos, G.: 'Gait recognition using geometric features and soft biometrics', *IEEE Signal Process. Lett.*, 2010, **17**, (4), pp. 367–370
- Seely, R., Samangoeei, S., Lee, M., Carter, J., Nixon, M.: 'The University of Southampton multi-biometric tunnel and introducing a novel 3D gait dataset'. Second IEEE Int. Conf. Biometrics: Theory, Applications and Systems (BTAS), 2008, pp. 1–6
- Potamianos, G., Neti, C., Luetttin, J., Matthews, I.: 'Audio-visual automatic speech recognition: an overview', 2004
- Landais, R., Bredin, H., Zouari, L., Chollet, G.: 'Vérification audiovisuelle de l'identité'. Proc. Traitement et Analyse de l'Information: Méthodes et Applications (TAIMA), Hammamet, Tunisia, 2007
- Jain, A.K., Nandakumar, K., Lu, X., Park, U.: 'Integrating faces, fingerprints, and soft biometric traits for user recognition'. Proc. Biometric Authentication Workshop, 2004, pp. 259–269
- Marcialis, G., Roli, F., Muntoni, D.: 'Group-specific face verification using soft biometrics', *J. Vis. Lang. Comput.*, 2009, **20**, pp. 101–109
- Jain, A.K., Dass, S.C., Nandakumar, K.: 'Soft biometric traits for personal recognition systems'. Proc. Int. Conf. Biometric Authentication (ICBA), 2004, pp. 731–738
- Rosenbaum, D.A., Meulenbroek, R.J., Vaughan, J., Jansen, C.: 'Posture-based motion planning: applications to grasping', *Psychological Rev.*, 2001, **108**, (4), pp. 709–734
- Matey, J., Naroditsky, O., Hanna, K., *et al.*: 'Iris on the move: acquisition of images for iris recognition in less constrained environments', *Proc. IEEE*, 2006, **94**, pp. 1936–1947
- Cucchiara, R., Grana, C., Piccardi, M., Prati, A., Sirotti, S.: 'Improving shadow suppression in moving object detection with HSV color information', *Intell. Transp. Syst.*, 2001, pp. 334–339
- Bobick, A., Davis, J.: 'The recognition of human movement using temporal templates', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (3), pp. 257–267
- Goffredo, M., Bouchrika, I., Carter, J.N., Nixon, M.S.: 'Self-CALIBRATING VIEW-INVARIANT GAIT Biometrics', *IEEE Trans. Syst. Man Cybern. B*, 2009, pp. 997–1008
- Goffredo, M., Bouchrika, I., Carter, J.N., Nixon, M.S.: 'Performance analysis for automated gait extraction and recognition in multi-camera surveillance', *Multimedia Tools Appl.*, 2009, pp. 75–94
- Han, X., Liu, J., Li, L., Wang, Z.: 'Gait recognition considering directions of walking'. Proc. IEEE Conf. Cybernetics and Intelligent Systems, 2006, pp. 1–5
- Yu, C., Cheng, H., Cheng, C., Fan, K.-C.: 'Efficient human action and gait analysis using multiresolution motion energy histogram', *J. Adv. Signal Process. (EURASIP)*, 2010, **2010**, p. 13
- Koekoek, R., Swarttouw, R.F.: 'The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue', 1998
- Gomez, G., Morales, E.F.: 'Automatic feature construction and a simple rule induction algorithm for skin detection'. Proc. ICML Workshop on Machine Learning in Computer Vision (MLCV), 2002, pp. 31–38
- Viola, P., Jones, M.: 'Rapid object detection using a boosted cascade of simple'. IEEE Proc. Computer Society Conf. Computer Vision and Pattern Recognition (CVPR), 2001, vol. 1, pp. 511–518

- 48 Ramesh, D.C.V., Meer, P.: 'Real-time tracking of non-rigid objects using mean shift'. *IEEE Proc. Computer Vision and Pattern Recognition 2007 (CVPR)*, 2000, vol. 2, pp. 142–149
- 49 Sakoe, H., Chiba, S.: 'Dynamic programming algorithm optimization for spoken word recognition', *Readings in speech recognition*, 1990
- 50 Rabiner, L.R.: 'A tutorial on hidden Markov models and selected applications in speech recognition', *Proc. IEEE*, 1989, **53**, (3), pp. 257–286
- 51 Damousis, I., Bakirtzis, A., Dokopoulos, P.: 'Network-constrained economic dispatch using real-coded genetic algorithm', *IEEE Trans. Power Syst.*, 2003, **18**, (1), pp. 198–205
- 52 Nixon, M.S., Tan, T., Chellappa, R.: *2006 Human Identification Based on Gait*, 2006

10 Appendix

Genetic fusion algorithm, the genotypes or chromosomes for the current GA are provided by the concatenation of w_{RIT} , w_{KRM} and w_{HMM} . An initial population of m chromosomes is generated. Each of them denotes the weight for the gait features scores (RIT, KRM) and for the activity-related recognition scores (HMM), respectively. They all range between 0 and 1, similarly to the training patterns, which stand for the dissimilarity scores of the extracted feature. Then, the total similarity $\text{Sim}(x,y)$ of each person (gallery) in the database to the client (probe) is given by (18).

The user's ID that achieves the greatest matching score is notated as C

$$C = \underset{y \in \mathbb{R}}{\text{arg max}} \text{Sim}(x, y) \quad (19)$$

Following, the quality of a specific chromosome for the subject C is measured with respect to its 'fitness' function

f_{fitness} , as follows

$$f_{\text{fitness}} = \sum_{x=1}^{N_p} \text{correct_id}_x \quad (20)$$

where x denotes the probe i_d . In this context, the correct_id_i is given by the following

$$\text{correct_id}_i = \begin{cases} 1, & \text{if } \text{Sim}(x, C) = \max(\text{Sim}(x, y)), y = \{1, \dots, N_G\} \\ 0, & \text{if } \text{Sim}(x, C) < \max(\text{Sim}(x, y)), y = \{1, \dots, N_G\} \end{cases} \quad (21)$$

The final weight scores have been taken after the generation of 300 new generations of chromosomes, since thereafter the algorithm converged sufficiently. Seemingly, the fitness maximises through the evolution of the population and so does the number of correctly identified individuals in the database, as well.

To avoid overfitting and database-dependent weights, the proposed fusion method was only used to estimate the optimal weights for each modality. After their calculation, the weights have directly applied for the online identification of individuals and no further training or altering of the weights occurred for the database. Hence, here we only introduce a fusion at the score level whereas leaving our feature extraction algorithms to execute without any additional training procedures.