Pattern Recognition 48 (2015) 1743–1759

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Activity related authentication using prehension biometrics

Anastasios Drosou [a,*], Dimosthenis Ioannidis [a], Dimitrios Tzovaras [a], Konstantinos Moustakas [b], Maria Petrou [a]

[a] Information Technologies Institute, Centre for Research and Technology Hellas, P.O. Box 361, 57001 Thermi-Thessaloniki, Greece
[b] Electrical and Computer Engineering Department, University of Patras, Patra, Greece

ARTICLE INFO

Article history:
Received 10 April 2013
Received in revised form
19 January 2014
Accepted 9 December 2014
Available online 17 December 2014

Keywords:
Biometrics
Behavioural biometrics
Biometric recognition
Biometric authentication
Prehension biometrics
Activity Curves
Activity surfaces
Activity related biometrics
Spherical harmonics
DTW

ABSTRACT

This paper presents an extensive study on prehension-based dynamic features and their use for biometric purposes. The term prehension describes the combined movement of reaching, grasping and manipulating objects. The motivation behind the proposed study derives from both previous works related to the human physiology and human motion, as well as from the intuitive assumption that different body types and different characters would produce distinguishable, and thus valuable for biometric verification, activity-related traits. A novel approach for analyzing such movements is presented herein, based on the generation of an activity related manifold, the *Activity hyper-Surface*. The authentication capacity of the extracted features on the activity hyper-surface is evaluated in terms of their relative entropy and their mutual information within a complete framework targeting user verification. Experimental results on two datasets of 29 real subjects each and a third one of 100 virtual subjects show that the introduced concept constitutes a promising approach in the field of biometric recognition.

1. Introduction

Biometrics have recently gained significant attention from researchers while they have been rapidly developed for various commercial applications, ranging from access control against potential impostors, to the management of voters to ensure no one votes twice [1]. These systems require reliable personal recognition schemes to either confirm or determine the identity of an individual requesting their services. A number of approaches have been proposed in the past to satisfy the different requirements of each application, such as unobtrusiveness, reliability, and permanence.

Biometric methods are categorized into physiological and behavioural [2], depending on the type of used features. Physiological biometrics are usually based on static biological measurements and inherent characteristics of each human. The most typical example in this area is the fingerprint [3], which is widely used in law enforcement for identifying criminals [4]. Further, static biometrics include DNA, facial characteristics [5], iris [6] and/or retina [7], and hand geometry [8] or palm print [9] recognition. Despite their high accuracy, a general shortcoming of these biometric traits is the obtrusive process of obtaining the biometric signature. The subject has to stop, go through a specific measurement protocol, which can be very uncomfortable, wait for a period of time and get clearance after authentication is positive. Besides being obtrusive and uncomfortable for the user, static physical characteristics can be digitally duplicated, e.g. the face could be copied using a photograph, a voice print using a voice recording and the fingerprint using various forging methods. In addition, static biometrics could be intolerant of changes in physiology, such as daily voice changes or appearance changes.

On the other hand, recent technologies in biometrics resemble more natural ways of recognizing people. Similar to the methods or techniques humans utilize in order to recognize each other, modern trends in biometrics focus on the recognition of dynamic face grimaces, gait, movements, etc. In other words, they tend to recognize liveness rather than static features as the aforementioned traits do (fingerprint, iris, etc.). In this respect, behavioural biometrics are related to specific actions and the way that each person executes them. On the whole, behavioural biometrics are less obtrusive and simpler to implement [2,10], although they are less reliable than physiological biometrics. This way, integral drawbacks of regular biometrics can be lifted; for instance, inborn physiological

* Corresponding author.
E-mail addresses: a.drosou09@alumni.imperial.ac.uk (A. Drosou), djoannid@iti.gr (D. Ioannidis), tzovaras@iti.gr (D. Tzovaras), moustakas@ece.upatras.gr (K. Moustakas).

http://dx.doi.org/10.1016/j.patcog.2014.12.008
0031-3203/

characteristics may be mixed with stylish and behavioural ones, so that even twins can be separated.

The imposed obtrusiveness by non-behavioural biometrics lies in both the utilized sensors (e.g. fingerprint or iris reader) and the authentication procedure to which the users are subjected to. Contrary to sensor-based recognition [11] using behavioural signals, recent research trends have been moving towards the vision-based methods [12]. Additionally, recent work and efforts on human recognition have shown that there is quite a number of behavioural traits on which recognition can be based (e.g. extraction of facial dynamics features [13]). However, the most well-known example of behavioural biometrics is the human body shape dynamics [14] or joints tracking analysis [15] for gait recognition. In the same respect, the analysis of dynamic activity-related trajectories [19] provides the potential of continuous authentication for discriminating people, when considering behavioural signals.

Not rarely, action recognition approaches are utilised in biometric systems either complementarily [19] or inspirationally. Such a view-, style- and appearance-invariant enhancement of the well know Motion History Images concept for action recognition has been proposed by Vitaladevuni et al. in [46], whereby ballistic dynamics are utilised in a Bayesian network. Another view-invariant action recognition algorithm has been proposed by Cuntoor et al. in [44], whereby the actions are recognised and labelled as abnormal per case, via the detection of significant spatiotemporal changes by properly trained HMMs. Similarly, dramatic changes in the speed and the direction of the trajectory are detected in [45] by estimating the spatiotemporal curvature of the latter, via a compact and view invariant representation based on dynamic descriptors. Finally, taking advantage of the hierarchical decomposition approach of specific actions based on the detection of notable characteristic changes in the motion sequence, Tanveer Syeda-Mahmood [43] tried to resemble human perception in the video segmentation process.

In the present paper, the term *prehension-based biometric authentication* is introduced as the combination of a reaching movement and a grasping activity. This concept derives from the simple observation that a person uses/manipulates the objects in the surrounding environment (e.g. answering a phone call) with his/her own style. Prehension biometrics belong to the general category of behavioural biometrics and can also be thought of as a specialization of activity related biometrics [16,17].

Given that most of the activities performed in everyday life include the human's physical interaction either with other people or with objects, the current work attempts to detect and to evaluate a series of stable, invariant, time lasting and unique activity related biometric characteristics for each human. In this paper, the focus is on the arm's movement [19] and on the movement of the fingers. Thus, both movements are thoroughly studied during specific actions that include people manipulating objects. Although palm dynamic features have not been employed in the field of biometrics yet, significant amount of research has been performed on various aspects of dynamic palm gestures [28].

The core topic with which this study deals is introduced in Section 3 and the proposed feature extraction methodology is described in detail in Sections 4 and 5. The tools used for feature evaluation are presented in Section 6.1, while the case-study scenarios, along with the obtained results can be found in Section 7. Finally, the conclusions follow in Section 8.

2. Motivation

As mentioned earlier, recent trends in biometrics deal with analyzing the dynamic nature of various biometric traits, targeting user convenience and optimal performance in various realistic environments. Activity-related biometrics have been recently studied in [16,17], where signals from various modalities are measured, while the subject is performing specific activities. These signals are then used to create unimodal or multimodal activity-related biometric signatures of each subject. Moreover, activity-related biometrics have been proven to have the potential to discriminate accurately between subjects, while remaining stable over time for the same subject.

However, not any movement can be seen as a potential identifier. The requirements that a biometric trait should satisfy are defined below [2]:

- *Universality*: Each user should posses it.
- *Distinctiveness*: The extracted features are characterized by great inter-individual differences.
- *Reproducibility*: The extracted features are characterized by small intra-individual differences.
- *Permanence*: No significant changes occur over time, age, environmental conditions or other variables.
- *Collectability and automatic processing*: It is possible to recognize or verify a human characteristic, which can be measured quantitatively, in a reasonable time and without a high level of human involvement.
- *Circumvention*: It should be difficult to be altered or reproduced by an impostor who wants to fool the system.

Additionally, there are a number of other issues that should be considered when designing a practical biometric system, like its recognition performance (i.e. achievable recognition accuracy and speed) and the resources required to achieve the desired recognition accuracy and speed. Further operational and environmental factors, such as the frequency with which a given activity is performed on a daily basis and the degree of approval of a certain technology by the society, are also significant issues to be taken into account.

In the concept of the current study, the *Universality* requirement is satisfied by definition. Moreover, there are plenty of models which depict that the user seeks the “most convenient” and the less effort demanding way of performing each movement. Specifically, there is the *Flash and Hogan's Minimum Jerk Model* [29] which indicates that the hand paths in extrinsic space should be straight. Curved hand paths can be generated, of course, but according to this model, they must be produced by concatenating straight-line segments. Similarly, the *Uno, Kawato and Suzuki Minimum Torque Change Model* [30] assumes a hand movement according to the minimization of the torque during the movement. Based on these observations, but also on Turvey et al.'s [31] and Goodman et al.'s [32] findings, it can be claimed that not only the *Distinctiveness*, *Reproducibility*, but also the *Permanence* requirements are also fulfilled, since all these parameters are related to the user's anthropometric variables, that exhibit significant variance within the population. Of course, like all biometrics, there is the issue of aging, which can only be overcome via the update of the biometric signature over time. However, expressions of behaviour are less vulnerable to sudden changes (i.e. a fingercut is much more frequent and has a direct and quick effect on the authentication than a change in the movement due to arthritis or other diseases).

Similarly, the *Permanence* requirement is guaranteed, given that the human body remains unchanged over the years, in terms of anthropometric proportions, like the distances between the joints. Moreover, the proposed approach utilizes a combination of physiological with stylish and behavioural characteristics. Thus, the proposed biometric traits are very *hard to circumvent*, if not impossible, by an impostor. Furthermore, provided the fact that recent technological achievements, especially regarding miniaturized sensors and accurate vision-based tracking algorithms, allow

the unobtrusive application of such biometric technologies. Additionally, given that the recognition process is incorporated in the daily activities of the user, it can be stated that the acceptability and frequency criteria are covered, as well. Finally, the *Automatic processing* requirements, including the recognition accuracy and speed, are highly dependent on the features and algorithms deployed.

Similar to gait, a prehension movement is a very frequent activity performed in everyday life that describes the sequential occurrence of two independent and complementary activities. Namely, it includes the activities of reaching and grasping an object in the vicinity of a user. Such activities may further involve the handling of the doorknob in order to enter or to leave a room, the answering of a phone call by picking up the phone, the grasping of the wheel when driving, the interaction with the mouse when working with the computer, etc.

The intuitive assumption is again that all users have their own characteristic way of reaching, grasping or in general manipulating specific objects, while performing specific activities. The motivation behind this assumption stems from the expectation that different articulated structures (e.g. human body, palm and fingers) and different human behaviours would produce distinguishable activity-related traits. In this context, the movement of the arm towards the object, the positions of the hand, the palm and the fingertips with respect to the object are analyzed, in order to extract unique signatures of dynamic nature that would form a reliable biometric signal for authentication.

The scope of the current paper is the demonstration of the recognition potential of prehension biometrics. In other words, this paper forms a study of activity related features that aims to suggest efficient and robust algorithms for their processing during certain prehension movements, which resemble the presented ones. The major improvement over previous works is the framing of the reaching biometrics in the more generic concept of prehension biometrics. In this extent, the concepts of *Activity Curves* and *Activity hyper-Surfaces* are both proposed as novel descriptors for prehension activities and evaluated accordingly. The term Activity Curves (ACs) describe the angular or spatial transition of a certain joint over time. Especially, for the case of a joint's spatial transition, ACs can be extended to a more descriptive activity manifold, the Activity hyper-Surface (AhS), which encompasses the relative distance between the joint's location from a moving reference point (e.g. the location of the head) as a function of time.

3. Related work–theoretical aspects on prehension

Biological systems exhibit complex behaviours of functioning, which sometimes can not be explicitly explained. Thus, observed behaviours may be attributed to certain "black boxes", that optimize either some activity related criteria or the teleological behaviour of the whole organism. On the contrary, complex behaviours could result from observable physical properties of the systems and their environment, and/or from explicitly expressed common control principles [31].

In this respect, complex multijoint movements, such as reaching or grasping an object, are planned and executed not only according to one's exclusive personal behaviour, but also due to various physical properties and phenomena (e.g. the physiology of the human body) [31]. According to Goodman et al. [32], some features, like the ones discussed hereafter (Section 3.1), have been proven to be independent of movement distance, direction, starting position and external load. Thus, it is reasonable to claim that by relying on such invariant features, user-specific activity-related properties can be modeled as biometric signatures for authentication purposes.

Moreover, according to Hoff et al. [33] a prehension activity can be divided into two parts: (i) a fast initial movement, whereby the user moves the arm to transport the hand towards the object and preshape their fingers (Fig. 1(a)), and (ii) a slow approach movement, whereby the final stage of the grasping scheme takes place (Fig. 1(b)). In this sense, these two sub-activities are approached in the current work both as complementary movements, but also as distinct biometric identifiers. Thus, the current work uses a dual approach, whereby each part of the prehension activity is studied separately. At the end, the results are fused in order to provide a single authentication framework.

In the context of the current study, the features selected for both phases of a prehension activity are mainly of dynamic nature. However, it can be claimed that static physiological information is also encoded (e.g. the relative mean or maximum distance values between the head and the hands during the reaching movement). This assumption can be easily extended to the fingers' movement, whereby the dynamic, pre-grasping movement (opening of the palm and closing to the object's dimensions) forms the dynamic part, while the final hand posture is seen as the static, user oriented one. Thus, the features described next are related to both the users' anatomy and their habitual behaviour.

3.1. Theoretical aspects on Reaching

Regarding the reaching task, the most important so-called invariances have been analyzed in [32] and in [34]. Specifically, any reaching task of a human arm is characterized by the following common properties:

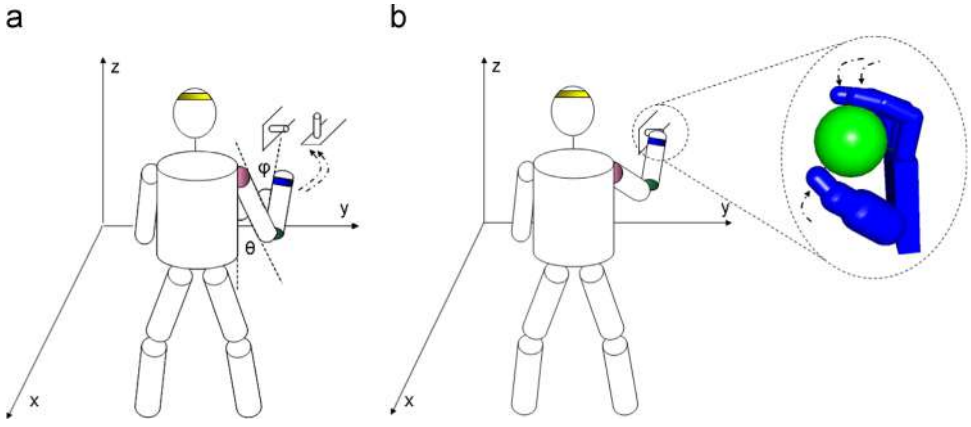

Fig. 1. (a) During a *Reaching Movement* shoulder and elbow angles change in a predefined way. (b) During a *Grasping Activity* the fingers' and the palm's angles are moving towards the hand's final posture. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

1. It is equifinal (i.e., the limb end-point reaches the vicinity of a target under a wide range of external conditions).
2. Most of its path usually lies along a straight line, although it can be slightly curved and hooked at the end.
3. The time profile of the limb end-point tangential velocity is approximately bell-shaped, with some distortions at its end.
4. The trajectory reflects speed-sensitive (uniform rates of joint torque development) or speed-insensitive (variable rates of joint torque development) movement strategies, depending upon the specifications of the movement task.
5. In case of a double-step target, the path is curved and the velocity–time profile is bi-modal.

Several models have been proposed in the past attempting to describe a reaching movement in a deterministic way. In this respect, the dependency of the arm's angles is explicitly stated [35] and the fact that the users seek the "most convenient" and the least effort demanding way to perform each movement [30,29] has also been proven.

One of the most important studies has been conducted by Rosenbaum et al. [34], who came up with the finding that the final body postures are not simply considered as the results of movements, but as goals that movements serve to satisfy. These notions were justified as follows.

1. Optimal movements can be generated once initial and final postures are known. As assumed in several models, knowing the final as well as initial postures allows the creation of optimal movements.
2. Memory for final positions is better than memory of movements [36].
3. Variability of end positions is generally smaller than variability of movements towards those end positions [37].
4. The end-state comfort effect, defined as willingness to adopt initially uncomfortable postures for the sake of comfortable final postures, is better predicted by ratings of final-posture than by ratings of movement ease.

3.2. Theoretical aspects on Grasping

As an extension to the reaching task, the finishing of a prehension activity involves the grasping of the object. Generally, the movement of the fingers follows the same basic rules as for any articulated human model (e.g. *Memory for Final Posture*).

The authentication capacity of such an action has been initially presented by Vogiannou et al. in [38], where the whole concept of grasping-based biometric features, as behavioural, dynamic biometric features related to the dynamic manipulation of objects, has exhibited promising potential for biometric person identification. In this respect, *Grasping Biometrics* can be seen as a special case of activity related biometrics, which deal with the characteristic features of human grasping, including both hand posture and activity related dynamic traits that contribute to the discrimination between different subjects.

Additionally, the work that has been performed in [39] showed significant variance in the movement of the finger joints during grasping of several objects among a variety of subjects. Inspired from that, but also based on the physiological differences between the palms and fingers of different users, it can be claimed that increased authentication potential is encoded in the way one grasps an object. This claim can be also supported by Rosenbaum's model [34], which states, among others, that angular trajectories demonstrate high variability within a population, although segments of paths may be relatively straight.

At this point, it is important to point out that grasping-related features of the hand are not the same as hand biometrics which have already been employed for human recognition [40]. Although certain hand characteristics, such as the size of the palm or the length of the fingers, have an effect on the way humans manipulate objects, grasping biometrics are primarily concerned with the behavioural features and the dynamics of the specific action. Thus, descriptors invariant to palm sizes are going to be exploited herein (e.g. angular acceleration and total angular distance covered by the fingers). Similarly, given the fact that the final hand gesture is dependant on the object involved, the measurements performed in the current study are grouped with reference to the same activity-experiment (e.g. the picking of a phone).

3.3. Context dependencies in activity-related biometrics

As suggested in several works of Rosenbaum et al., the context (i.e. both the environmental setting and the temporal order of ongoing events), in which a movement based human identification takes place, may significantly affect its outcome, either by influencing the perception of the identifier [24] or by affecting the planning and execution of the prehension movement of the person to be identified [25,26].

Following this, it becomes evident that behaviour analysis and context are in close relation with each other. Thus, in order to proceed with the analysis of someone's behaviour, the context has to be known, in which the observed behavioural signal has been displayed. In this respect, the definition of the context should be provided either via the *W*4 (where, what, when and who) or even better via the *W*5+ (where, what, when, who, why and how) methodology.

However, provided that the problem of context-sensing is extremely difficult to solve, especially for a general case, answering the "why" and the "how" questions in a *W*4-context-sensitive manner is virtually unexplored area of research [27]. Thus, without loss of generality, in the current work only the apparent perceptual aspect of the context (*W*4), in which HCI takes place, will be dealt with.

Based on the *W*4 approach, in order to design a prehension-based recognition system that will work circumventing the context dependency of prehension, the following requirements should be fulfilled:

1. The relative position of the actor performing the prehension movement should be fixed with respect to the interaction object (i.e. object to be reached and grasped).
2. The space between the actor and the interaction object should remain similar as during the registration procedure (e.g. no obstacles should interfere).
3. The actor should be at a similar affective state, as the time he/she has been registered to the system.
4. The interaction object should remain the same, in terms of shape and size.
5. The order of the movements in a specific scenario should stay unchanged, so as the initial and the final position of the body parts that participate in the prehension movement are the same as during the registration.

To this extent, possible applications where such application specific scenarios can be successfully designed include restricted areas in military bases, sensitive infrastructures where classified data are stored (e.g. personal data, medical data, and war plans), highly secure areas in nuclear power plants (e.g. control rooms), control rooms of surveillance/security companies and administrator rooms in the central servers of companies managing big amounts of data.

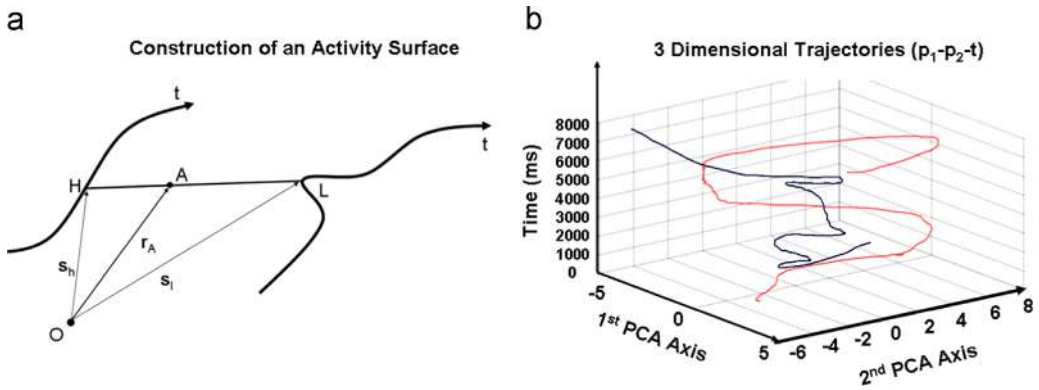

Fig. 2. (a) Construction of an activity surface. O is the origin of the axes. The two black curves represent the trajectories of vectors $\mathbf{s}_h(t)$ and $\mathbf{s}_l(t)$ parameterized by time t. Point A is any point of segment HL defined by the position of the head H and the limb L at a specific time, with position vector $\mathbf{r}_A$. (b) Trajectories of the head and the palm in space and time after dimensionality reduction via PCA of the spatial coordinates. Joining the corresponding points of the two curves forms the characteristic surface.

4. Reaching–feature extraction using Activity hyper-Surfaces

A novel descriptor for the reaching part of a prehension movement is presented herein. Specifically, the arm's movement will be described via the novel concept of *Activity hyper-Surfaces* (AhS). However, in order to introduce the AhS, the concept of *Activity Curve* (AC) has to be defined. In particular, an AC describes the spatial displacement of the head or an arm's joint during the prehension movement. Thus, an AC is defined for a certain limb l as the curve made up by the point with position vector $\mathbf{s}_l(t)$ of the limbs in the 3D space:

$$\mathbf{s}_l(t) = (x_l(t), y_l(t), z_l(t)) \tag{1}$$

As time t is sampled with N points, this curve is made up from N consecutive points.

Similarly, the series of points $\mathbf{s}_h(t) = (x_h(t), y_h(t), z_h(t))$ of the position of the head h makes up the AC of the head.

An *Activity hyper-Surface* (AhS) is defined as the surface made up from the points with position vectors $\mathbf{r}_A = (1-\mu)\mathbf{s}_h + \mu\mathbf{s}_l$, where $0 \le \mu \le 1$ (see Fig. 2(a)).

The area A of such a surface is given by

$$A = \int_{AhS} \left[(1-\mu)\mathbf{s}_h(t) + \mu\mathbf{s}_l(t)\right] dA \tag{2}$$

where $dA = dt\, d\mu$ stands for the infinitesimal surface element.

In this respect, three AhSs can be extracted from an arm's movement: (a) the head-to-shoulder AhS, (b) the head-to-elbow AhS and (c) the head-to-palm AhS. From now on, the term AhS will refer only to the head-to-palm for the rest of the paper. This simplification is justified by the facts that the movement of the elbow's joint is significantly correlated with the palm's movement ([35] and Section 7.4) and that the shoulder's and head's movements exhibit high dependency on each other.

The proposed AhS is quite a complicated manifold to be used directly. However, it has been reported that during a prehension activity, the movement is mainly concentrated on a 2D plane [29]. Thus, dimensionality reduction principles can be applied in order to simplify the calculations. In this respect, axis rotation is performed in the (x, y, z) subspace of the AhS via Principal Components Analysis (PCA) and the eigenvector with the smallest eigenvalue is removed. The remaining two dimensions plus the time axis form a 3D space in which the original AhS manifold is represented by a surface, characteristic of the ongoing activity, as illustrated in Fig. 2(b).

Definition. The simplified Activity Surface (AS) comprises the union of all points that lie on the lines connecting the corresponding representative points of the head and hand trajectories. Geometrically, the surface is defined as the area within the spatial bounds of the head's and hand's Activity Curves and the temporal limits of the activity's duration (Fig. 3).

As time increases monotonically, this definition means that the Activity Surface is made up by a series of line segments, parallel to the plane defined by the spatial coordinates. So this representation explicitly encodes the relative distance between the head and the hand during the ongoing activity, additionally to the information provided by the original shape of the motion trajectories.

The head is considered as a direct extension of the human torso. Provided the restricted range in the angles within which the neck can bend/rotate and especially under the assumption that the full face can be simultaneously recognized, the head can be utilized as a very robust reference point for the rest of the body [21]. In the same respect, head pose has been extensively studied in [22] and in end-effector of the whole arm. In this context, a strong dependency and thus a high correlation between the movement of the hand to the rest arm joints during a prehension movement are not only supported in [38] but also experimentally proven in the mutual information confusion matrices.[1] In this respect, it is assumed herein that combination of these two salient points (i.e. head and hand) does not only represent the prehension movement to a significant extent, but also reveal a great amount of biometric information for the user. Finally, the use of the hand (i.e. position of the palm/wrist) is justified since it is considered as the end-effector of the whole arm.

Moreover, the movements' velocity distribution is also encoded implicitly by this representation, given that the movement's duration is mapped on a separate axis. In Fig. 3, the intra-similarities and the inter-variances of the Activity Surfaces between several users are visually illustrated.

In the following, a series of activity related features will be extracted based on several subspaces of the introduced AhSs. In particular, each consisting Activity Curve will form both a set of features itself and the basis for further feature extraction, while the Activity Surfaces will be processed, so as to produce novel features in terms of *Spherical Harmonic Analysis* (SHA) [41]. In particular, SHA have been chosen, so as to preserve the view invariance of the AS, which is a combination of two trajectories.

4.1. Spherical harmonic coefficients as biometric descriptors

As any surface in a 3D space, the generated Activity Surfaces can be uniquely described in terms of Spherical Harmonic

[1] http://www.iti.gr/~drosou/PrehensionBiometrics/MIConfusionMatrices.pdf

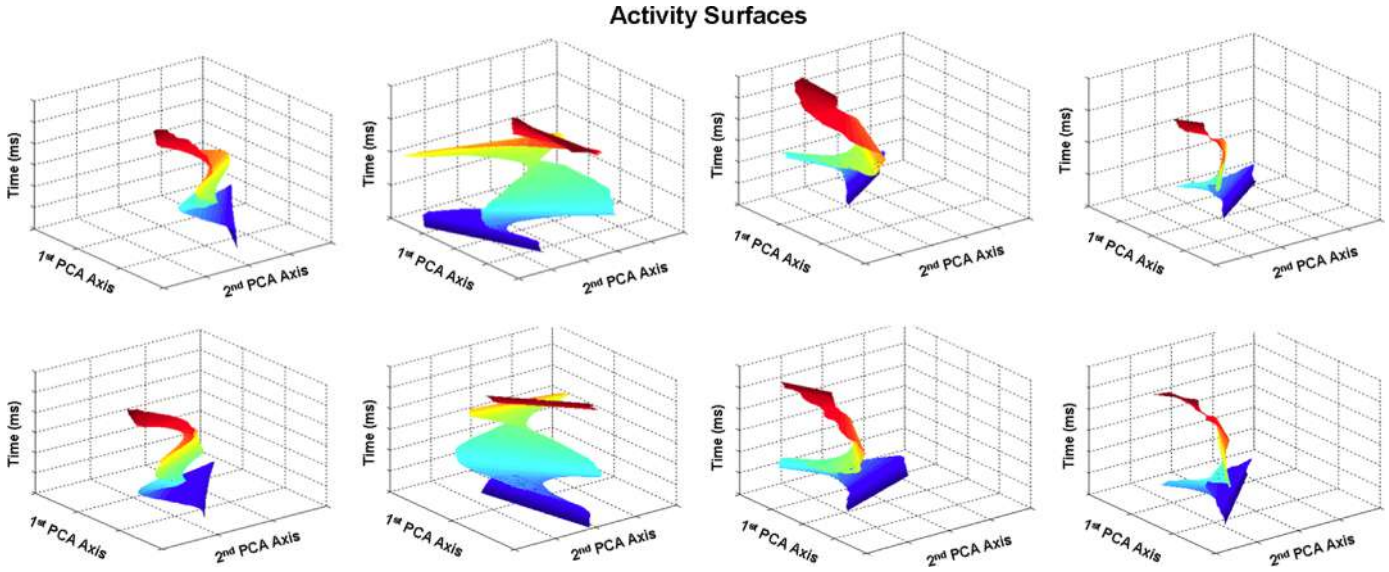

Fig. 3. Eight activity surfaces exhibiting visually intra-similarity and inter-variances. The surfaces in different columns correspond to different people, all executing the same action. The PCA based dimensionality reduction is only used to produce the simplified activity surface (AS), that is then used for the extraction of the spherical harmonics related features.

Coefficients (SHCs) via Spherical Harmonic Analysis (SHA). For their calculation, surface AS has to be expressed in terms of spherical coordinates (ρ,θ,ϕ), triangulated and resampled with a constant sampling density $\mathbf{d}_s=(d_s^{\phi},d_s^{\theta})$ in the two angular coordinates. The reference point R, i.e. the origin of the spherical coordinates, for which $\rho=0$, has to be carefully selected. Given the limitations of the SH algorithm, that only one value of r can be assigned to a (θ,ϕ) pair, it is critical to select an R point in our coordinate system, that minimizes the number of multiple intersections of the various radii with the surface. Since no special method is available for defining such an optimal origin point, multiple reference points ($\mathbf{R}=\{R_i|i\in[1,7]\}$) are proposed for the optimal description of the whole surface, as shown in Table 1. This way, the uncertainty introduced from a single-view SH coefficient extraction will be minimized.

Let $f_{R_i}:\mathbb{R}^3\rightarrow\mathbb{S}^2$ denote the function that performs the mapping of the surface to the corresponding R_i. Specifically, $f_{R_i}(\overline{\omega}):\{\overline{\omega}\in\mathbb{R}^3:f_{R_i}(\overline{\omega})\in\mathbb{S}^2)\}$, whereby $\overline{\omega}$ is a point $\{P_1,P_2,t\}$ of the AS in spherical coordinates:

$$f_{R_i}(\theta,\phi)=\min_{k=1,\ldots,K}\{d(R_i,v_k)\} \tag{3}$$

whereby $v_k=v_k(R_i,\theta,\phi)$ is the kth intersection of the AS with the ray that starts from the R_i point for a given pair of values of θ and ϕ. K is the total number of intersections in the particular direction, while $d(R_i,v_k)$ stands for the Euclidean distance between R_i and v_k.

The Spherical Harmonic Coefficients c_l^m for each $f_{R_i}(\theta,\phi)$ can then be easily calculated by utilizing the orthonormalized spherical harmonics, multiplying them with the aforementioned function, and integrating the product over the solid angle $d\theta\,d\phi$:

$$c_l^m=\sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}\int_0^{2\pi}\int_0^{\pi}f_{R_i}(\theta,\phi)P_l^m(\cos\theta)e^{jm\phi}\,d\theta\,d\phi \tag{4}$$

whereby P_l^m is the associated Legendre polynomial. As it can be seen, the Legendre polynomial takes two integer arguments l and m. In particular, l is used as the Spherical Harmonic Band (SHB) index to divide the class into bands of functions, resulting in a total of $(l+1)l$ polynomials for a lth band series, while $m\in[-l,l]$. It should be noted that between any P_l^m and a $P_{l'}^{m'}$ for different m values on the same band, the polynomials are orthogonal with each other, unless neither $m=m'$ nor $l=l'$ holds.

Next, in order to transform the extracted harmonics to comparable, rotation- and thus, view-invariant indicators, the following normalizations should be successively applied:

$$c_l^*=\sum_{m=-K_l}^{K_l}|c_l^m| \tag{5}$$

whereby c_l^* denotes the lth SHB coefficient, that is rotation invariant by definition, and K_l stands for the total number of SHs for the given SHB index l, while $m=2l-1$. For more details regarding the spherical harmonic analysis, the reader is referred to the report of Schonefeld [41].

4.2. Orientation as biometric descriptor

By further studying the previously constructed Activity Surface, the state vectors $\boldsymbol{\theta}_{arm,l}$ and $\boldsymbol{\phi}_{arm,l}$, for each limb l of the arm that performs the movement, can be extracted, as well. These two state vectors describe the orientation of the limb l with respect to the head of the user, for the whole duration of the movement, as shown below:

$$\theta(t)=\arctan\left(\frac{y_{head}(t)-y_{limb}(t)}{x_{head}(t)-x_{limb}(t)}\right) \tag{6}$$

$$\varphi(t)=\arccos\left(\frac{z_{head}(t)-z_{limb}(t)}{r(t)}\right) \tag{7}$$

where

$$r(t)=\sqrt{(x_{head}(t)-x_{limb}(t))^2+(y_{head}(t)-y_{limb}(t))^2+(z_{head}(t)-z_{limb}(t))^2}$$

and t a certain time instance within the duration of the movement.

The meaning of this transformation is graphically explained in Fig. 4(b). Thereby, the straight line that connects the points H (i.e. the position of the head at time t) and L (i.e. the position of limb l at time t) has a certain orientation in the 3D space. In particular, it forms the relative angle with which the limb l is positioned in the space that has its origin in the current position of the head. The orientation of the axes θ_0 and ϕ_0 is selected so that they are aligned to the direction, where the limb l is found at the beginning of the movement.

4.3. Speed, acceleration and jerk

In [35] it was reported that fluctuations of hand's speed during a prehension activity are generally described by a bell shaped distribution and that the movement is independent from the speed. Yet, speed, acceleration and jerk are not only affected by behavioural habits of the user, but also contribute in describing the temporal dimension of the activity. Thus, they will be included as indicative, behavioural traits in the evaluation performed within the current study.

In particular, by utilizing each bounding AC of the aforementioned AhS, a notion of the movement's speed and the instant speed variances at which an activity has been performed can be obtained as the distance between two successive sampling points. Given the spatial transitions and the temporal information, the speed, acceleration and jerk vectors of the head and palm during a prehension movement can be trivially calculated using the central differences of the well-known formulae:

$$\mathbf{v}_{x,y,z}(t)=\frac{\mathbf{s}_{x,y,z}(t+1)-\mathbf{s}_{x,y,z}(t-1)}{2\Delta t} \tag{8}$$

$$\boldsymbol{\alpha}_{x,y,z}=\frac{\mathbf{v}_{x,y,z}(t+1)-\mathbf{v}_{x,y,z}(t-1)}{2\Delta t} \tag{9}$$

$$\mathbf{j}_{x,y,z}=\frac{\boldsymbol{\alpha}_{x,y,z}(t+1)-\boldsymbol{\alpha}_{x,y,z}(t-1)}{2\Delta t} \tag{10}$$

4.4. Curvature & torsion trajectories

Working with the bounding ACs of the Activity hyper-Surface, a further set of four characteristic view invariant traits can be extracted. Namely, they are the curvature κ, torsion τ and their first order derivatives (κ_s and τ_s) with respect to the Euclidean arc-length parameter which is expressed by the position vector $\mathbf{s}(t)$ of the points along the curve, as follows:

$$\kappa(t)=\frac{\mathbf{s}(t)\times\ddot{\mathbf{s}}(t)}{\|\dot{\mathbf{s}}(t)\|^3};\quad \tau(t)=\frac{(\mathbf{s}(t)\times\ddot{\mathbf{s}}(t))\cdot\dddot{\mathbf{s}}(t)}{\|\dot{\mathbf{s}}(t)\times\ddot{\mathbf{s}}(t)\|^2} \tag{11}$$

for $t\in[1,T]$, where T is the total number of samples of the curve. Similarly, the corresponding derivatives are employed to construct the signature

$$\kappa_s(t)=\frac{d\kappa(t)}{ds}=\frac{d\kappa(t)}{dt}\frac{dt}{ds}=\frac{d\kappa(t)}{dt}\frac{1}{\|\dot{\mathbf{s}}(t)\|} \tag{12}$$

$$\tau_s(t)=\frac{d\tau(t)}{ds}=\frac{d\tau(t)}{dt}\frac{dt}{ds}=\frac{d\tau(t)}{dt}\frac{1}{\|\dot{\mathbf{s}}(t)\|} \tag{13}$$

The intra-similarities and the inter-dissimilarity of some extracted curvature and torsion traits from three arbitrary experiments (Section 7.4) are illustrated in Fig. 5(a) and (b) respectively.

The analytical methodology, that has been followed herein for the estimation of the aforementioned quaternion of activity-related traits (i.e. curvature, torsion and the corresponding derivatives), is based on differential invariants. In particular, the signature components that depend on high order derivatives are sensitive to noise and round-off errors. In order to reduce this effect and to make the extracted traits robust, the straight-forward calculation of the high order derivatives (i.e. single point calculation) is avoided by involving multiple neighboring points. In other words, the aforementioned traits are numerically approximated from $\mathbf{s}(t)$ by using the joint Euclidean invariants (inter-point Euclidean distances), as described in [42].

Table 1
(P_1P_2t) coordinates of each utilized reference point. CoG_{P_2t}, CoG_{P_1t} and $CoG_{P_1P_2}$ stand for the Center of Gravities for each of the following planes P_2-t, P_1-t and P_1-P_2, respectively. $\min(AS_{P_1})$, $\min(AS_{P_2})$, $\min(AS_t)$ and $\max(AS_{P_1})$, $\max(AS_{P_2})$, $\max(AS_t)$ denote the minimum and maximum value of the AS in the corresponding dimension, respectively.

Point no.	P_1-location	P_2-location	t-location
$\mathbf{R}_1$	CoG_{P_2t}	CoG_{P_1t}	$\min(AS_t)$
$\mathbf{R}_2$	CoG_{P_2t}	$\min(AS_{P_2})$	$CoG_{P_1P_2}$
$\mathbf{R}_3$	$\min(AS_{P_1})$	CoG_{P_1t}	$CoG_{P_1P_2}$
$\mathbf{R}_4$	CoG_{P_2t}	CoG_{P_1t}	$CoG_{P_1P_2}$
$\mathbf{R}_5$	CoG_{P_2t}	CoG_{P_1t}	$\max(AS_t)$
$\mathbf{R}_6$	CoG_{P_2t}	$\max(AS_{P_2})$	$CoG_{P_1P_2}$
$\mathbf{R}_7$	$\max(AS_{P_1})$	CoG_{P_1t}	$CoG_{P_1P_2}$

4.5. Dynamic spatial cost

According to Rosenbaum et al. [34], motion planning and especially human movements are governed by two task-relevant costs; the spatial error cost and the travel cost. The travel cost, which depends on the changes in the angles at joints, cannot be applied as a descriptor for the arm's movement, since the joint angles are not provided by the tracker. However, the spatial cost dsc_p can be extended so as to form a useful descriptor that describes the total distance that is covered by the limb l during the activity. Originally, the travel cost was defined as

$$\mathrm{dsc}_p=\sqrt{(X_o-X_c)^2+(Y_o-Y_c)^2+(Z_o-Z_c)^2} \tag{14}$$

where (X_0,Y_0,Z_0) and (X_c,Y_c,Z_c) are the Cartesian coordinates of the target object o and the contact point c, respectively. In other words, this feature describes the absolute distance between the initial and the final position of a joint during a movement. Although this distance metric is of limited discrimination capacity, it can be easily extended to a valuable and meaningful dynamic trait, the *Dynamic Spatial Cost* (DSC), which indicates the covered

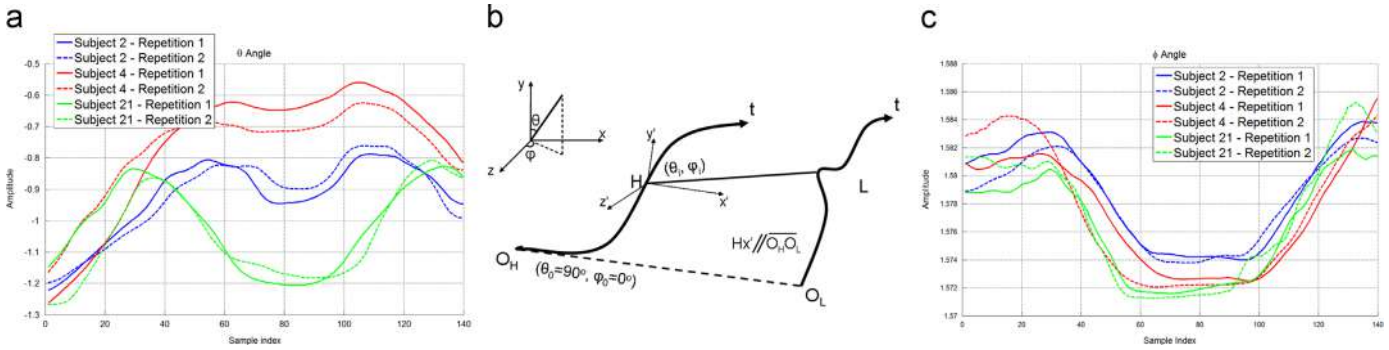

Fig. 4. The extraction of the orientation vectors of each joint of the arm is calculated with reference to the user's head. The origin of the axis is aligned with the initial pose of the user (i.e. Hx' // $\overline{O_HO_L}$). As such, two orientation vectors $(\boldsymbol{\theta},\boldsymbol{\phi})$ are generated for each joint of the arm, describing the changes of the movement's orientation in the 3D space. The distinctiveness between different users is clearly exhibited via (a) the θ-angle vectors than (c) the vertical ϕ-angle vectors.

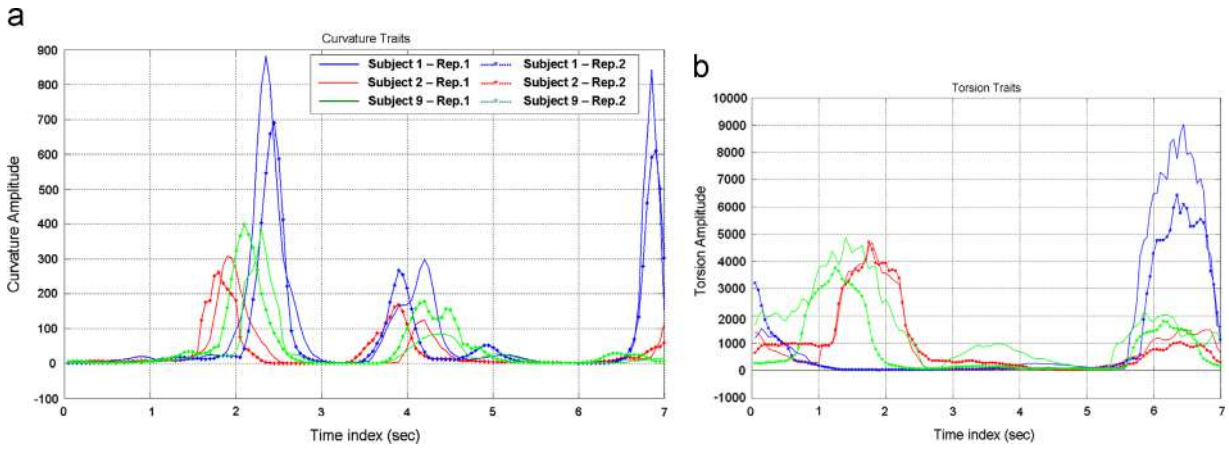

Fig. 5. (a) Curvature (*Phone Conversation* experiment), (b) Torsion (*Reach* & *Grasp* experiment) traits from different users.

distance at a given time-instant, is described by the following recursive equation:

$$\mathrm{dsc}_p(t) = \mathrm{dsc}_p(t-1) + \sqrt{(x_l(t)-x_l(t-1))^2 + (y_l(t)-y_l(t-1))^2 + (z_l(t)-z_l(t-1))^2}$$

In this respect the $\mathrm{dsc}_p(T)$ stands for the total distance covered by the joint during the activity.

Under these observations, it can be concluded that the motion towards an object is context specific, when described by the spatial error cost of Eq. (15) and thus, depends on the surrounding environment. Thus, the repeatability can be considered valid only for a fixed environment, which will be the case in the performed experiment (Section 7). On the contrary, the end position and posture of the user's hand is reported to be governed by the target and exhibits very low variation over time for the same user, therefore satisfying the Permanence requirement.

4.6. Activity Curves

Last but not least, all extracted Activity Curves $\mathbf{s}(t)$ are concatenated, so as to form a single state vector:

$$\mathbf{S}(t) = (\mathbf{s}_{head}(t), \mathbf{s}_{shoulder}(t), \mathbf{s}_{elbow}(t), \mathbf{s}_{hand}(t)) \tag{15}$$

The full set of extracted features, regarding the reaching movement, consists of the activity curves, the 7 sets of Spherical Harmonic coefficients, the speed, acceleration and jerk vector, the curvature, torsion and derivatives vector, as well as the dynamic Spatial Cost vector as functions of time ($\mathbf{V} = \{\mathbf{S}(t), \mathbf{c}_l^*, \mathbf{v}(t), \boldsymbol{\alpha}(t), \boldsymbol{\theta}_{arm,l}, \boldsymbol{\phi}_{arm,l}, \mathbf{c}(t), \mathbf{dsc}(t)\}$).

5. Grasping—feature extraction using Activity Curves

The most relevant scientific field to grasping analysis can be regarded as the Sign Language Recognition (SLR) and gesture recognition (GR) approaches, which can be divided into two types: the probabilistic [49–51] and geometric ones [48]. The latter, to which the proposed scheme belongs, usually deal with static gestures. However, one of the novelties of the proposed approach is that it manages to transform a dynamic gesture into a static hyper-surface and can thus apply geometric techniques.

Furthermore, SLR approaches have been found as inappropriate for utilization in the current framework due to the fact that their main target is to diminish and discard the intra-class (class = action X) variability of an activity, with respect to the users and therefore many recent approaches deal with 'subject independent recognition' problems. On a totally different perspective the proposed approach aims to pronounce and detect intra-class variability for biometric recognition purposes. Therefore the approaches used for SLR are in general case not well suited and cannot be used directly for prehension biometrics.

In the respect, the most typical geometric feature regarding grasping is the posture of the user's hand after the grasping equilibrium has been reached. The posture $\mathbf{P}$ is then defined as the set of angles θ_j of each joint j with respect to the predefined reference angles in the equilibrium position. Thus, the posture feature space can be defined as $\mathbf{P} = \{\theta_1, \theta_2, \ldots, \theta_N\}$, where N stands for the total number of the utilized fingers' joints.

However, the grasping biometrics introduced herein aim to encode the habitual behaviour of humans performing grasping actions (i.e. how they are used to grasp objects), along with the corresponding anatomical characteristics (i.e. the grasping posture depends on the shape, size and kinematics of each individual's palm). Following this, the aforementioned static features $\mathbf{P}$ can be extended to biometric features of dynamic nature by defining a sequence of successive postures over time $\mathbf{P}(t)$. In this respect, the feature space of the dynamic hand posture $\mathbf{P}(t)$ can be described by a set of N *Activity Curves* and can now be written as

$$\mathbf{P}(t) = \{\boldsymbol{\theta}_1(t), \boldsymbol{\theta}_2(t), \ldots, \boldsymbol{\theta}_N(t) | t \in [t_e - t_0, t_e]\} \tag{16}$$

where t_e refers to the grasping equilibrium time and t_0 is a timing offset appropriately defined and sufficiently large, so as to include the transitional motion of the hand just before grasping, but also relatively small, so that the corresponding motion in the interval $[t_e - t_0, t_e]$ should not be prone to variance due to environmental parameters, such as interfering objects. Because we have 4 Degrees of Freedom (DoF) for each finger and another 3 for the palm's base (Section 7.3), $N = 23$.

5.1. Angular speed, angular acceleration and angular jerk

The 1st, 2nd and 3rd time derivatives of each angle θ_i, $i \in \{1, \ldots, 22\}$ between the finger phalanxes describe the angular velocity ω_θ, the angular acceleration α_θ and the angular jerk, respectively:

$$\boldsymbol{\omega}_\theta = \frac{d\boldsymbol{\theta}}{dt}, \quad \alpha_\theta = \frac{d\boldsymbol{\omega}}{dt} = \frac{d^2\boldsymbol{\theta}}{dt^2}, \quad \boldsymbol{\beta}_\theta = \frac{d\boldsymbol{\alpha}_\theta}{dt} = \frac{d^2\boldsymbol{\omega}}{dt^2} = \frac{d^2\boldsymbol{\theta}}{dt^2} \tag{17}$$

5.2. Dynamic travel cost

Rosenbaum et al. [34] introduced in their study the total travel cost $\mathbf{dtc}_\mathbf{p}(t) = \sum_{j=1}^{N=23} \mathrm{dtc}_j(\boldsymbol{\theta}_j(t), T_j)$, where θ_j is the angular displacement of the jth joint from its starting to its end angle posture p and T_j denotes the time needed for the absolute angular

displacement. In addition, the Dynamic Travel Cost (DTC) $\mathrm{dtc}_j(\theta_j(t), T_j)$ is going to be utilized as a descriptor of the hand movements.

Specifically, the cost $\mathrm{dtc}_j(\theta_j, T_j)$ of moving joint j through an angle of size θ_j in a time T_j that may or may not equal the joint's optimal time, $T_j^*(\theta_j)$, for that same angular displacement is defined as

$$\mathrm{dtc}_j(\theta_j(t), T_j) = k_j\theta_j(1+[T_j - T_j^*(\theta_j)]^2) \tag{18}$$

where θ_j is measured in degrees, T_j in ms^2, while the optimal time is defined as

$$T_j^*(\theta_j(t)) = k_j \ln(\theta_j + 1), \quad k_j \geq 0 \tag{19}$$

whereby k_j is the joint expense factor that is assigned a value from 0 to 1, according to the angles' relative entropy value (see Section 6.1).

It is expected that distinctive variations will be extracted by Eq. (18) due to the unique finger size and articulation characteristics of each user. Thus, the *Hard to circumvent requirement* will be also satisfied (see Section 3). More evidence on this issue will be also presented in the recognition capacity analysis that follows in Section 6.1.

The total set of extracted features, regarding the grasping movement, consists of the angle vectors for each finger, the angular speed, angular acceleration and angular jerk vector and the Dynamic Travel Cost vector as functions of time $\mathbf{W} = \{\boldsymbol{\theta}(t), \boldsymbol{\omega}(t), \boldsymbol{\alpha}_\theta(t), \boldsymbol{\beta}_\theta(t), \mathbf{dtc}(t)\}$.

6. Feature selection and classification

This section deals with the evaluation of the authentication potential of the extracted features (Section 6.1) that will then be used as input in the utilized classifier, so as to proceed with the user authentication step (Section 6.2).

6.1. Evaluation of prehension related features authentication potential

Herein, two measures are presented that will quantify to a certain extent the features' authentication capacity, in terms of both their distinctiveness and their mutual dependency. These are the *Relative Entropy*, a metric for distinctiveness, and the *Mutual Information*, a metric for independency between distinct distributions. The outcomes of this analysis will lead to the final selection of the most independent features that exhibit high authentication capacity for the proposed biometric system.

6.1.1. Relative entropy and mutual information

Initially, it is assumed that for each of the aforementioned dynamic biometric features i there are two different probability density functions $f_i^{intra}(r)$ and $f_i^{inter}(r)$ for the intra- and inter-variances of the discrete random variables F_i^{intra} and F_i^{inter}, respectively. In this context, the relative entropy [52] between the inter-individual (f_i^{inter}) and intra-individual (f_i^{intra}) probability distributions of an entire population S is defined as follows:

$$D(f_i^{intra} \| f_i^{inter}) = \int f_i^{intra} \log \frac{f_i^{intra}}{f_i^{inter}} dr \tag{20}$$

For the relative entropy, also known as Kullback and Leibler divergence [52], $D(f_i^{intra} \| f_i^{inter})$ is describing the "distance" of f_i^{intra} from f_i^{inter}. However, the term "distance" is not intended to be taken in its most literal sense, since is not a metric. From the information theory viewpoint, $D(f_i \| f_j)$ can be interpreted as a measure for the expected discrimination information for f_i over f_j.

Still, since relative entropy is asymmetric i.e. $D(f_i^{intra} \| f_i^{inter}) \neq D(f_i^{inter} \| f_i^{intra})$, a notion of symmetry is usually inserted by the mean relative entropy:

$$D_{sym}(f_i^{intra} \| f_i^{inter}) = D_{sym}(f_i^{inter} \| f_i^{intra}) = \frac{D(f_i^{intra} \| f_i^{inter}) + (f_i^{inter} \| f_i^{intra})}{2}$$

Mutual information measures the information that is shared between two distributions. It is expected that the mutual information of independent distributions is zero. On the contrary, the mutual information between two identical distributions is as high, as the actual entropy $H(F_i) \equiv H(F_j)$ of each.

The mutual information value for each possible pair of features is calculated and normalized over the sum of both features' entropies, in order to obtain a standardized measure for the features' intra-dependency:

$$I^{norm}(F_i^{inter}, F_j^{inter}) = \frac{I(F_i^{inter}, F_j^{inter})}{H(F_i^{inter}) + H(F_j^{inter})} \tag{21}$$

whereby $I(f_i^{inter}, f_j^{inter})$ is calculated as

$$I(F_i^{inter}, F_j^{inter}) = \sum_{f_i^{inter} \in \mathbb{R}} \sum_{f_j^{inter} \in \mathbb{R}} f_{i,j}^{inter} \log \frac{f_{i,j}^{inter}}{f_i^{inter} f_j^{inter}} \tag{22}$$

6.2. Classification and user authentication

The format of the extracted features is a set of M state vectors obtained via frequent measurements of the interaction, whereby M is the number of simultaneously observed features. Although these state vectors provide quantitative snapshots of the interaction, only a subset M' of the total number of features will contribute towards the users' final verification.

Given that all features, apart from the Spherical Harmonics Coefficients (SHCs), exhibit a strong dependence on temporal relations and ordering, it is essential that classification is performed via some appropriate spatiotemporal means. The Dynamic Time Warping (DTW) algorithm [53] has been utilized as the classifier in the present scheme, since it sufficiently manages to capture the spatiotemporal information of the biometric traits. The SHC-related features are compared with each other via the *L1-norm*.

The DTW algorithm has been widely used in a series of matching problems, varying from speech processing [53] to biometric recognition applications [54]. Its main advantages are its simple implementation and its satisfactory performance given the required processing time. A possible implementation basis on estimating the closed area formed by the path around the diagonal of the rectangular DTW-grid (Fig. 6(a)). The total dissimilarity d_{DTW} between the vectors under comparison is defined as the product of the area A_c and the minimum difference cost $D_{min}(T, T)$, that are calculated via dynamic programming [53]:

$$d_{DTW} = A_c \cdot D_{min}(T, T) \tag{23}$$

The general process that is followed is that each "probe" feature vector or feature vector set is compared with the "gallery" template of the claimed ID that is stored in the database. In order to combine authentication scores from different modalities so as to derive an authentication metric for the full prehension movement, the scores from each tracking device have to be fused. It should be noted that the camera-based and sensor-based tracking devices are used in turns, in combination with the glove-based tracking device. The fusion of scores from different tracking devices is performed via score-level fusion:

$$D_{tot} = \sum_{j \in \{C,M,G\}} w_j d_{j,DTW} \tag{24}$$

whereby $d_{j,DTW}$ stands for the score provided by each tracking device j (C:Camera; M:Magnetic; G:Glove), while w_j is the

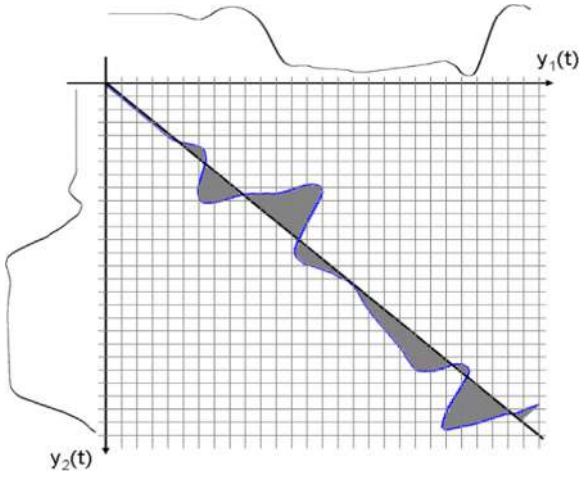

Fig. 6. Estimation of the matching score between the "gallery" and the "probe" vectors using a DTW-Grid. The plotted diagonal presents the (optimal) path with the least difference cost, i.e. "gallery" and "probe" vectors are identical. The value for A_c is calculated as the area enveloped between the optimal and the actual path on the DTW-Grid, as described in [53].

corresponding weight coefficient and is proportional to the total number of bits of information of the utilized features:

$$w_j = \frac{bits\ of\ information\ for\ all\ features\ of\ device\ j}{total\ number\ of\ bits\ for\ all\ utilized\ features} \quad (25)$$

7. Case study—prehension movements experimental evaluation

Following the analysis presented above regarding the reaching and grasping activities, two separate experiments have been conducted, in order to study the movement of the arm and the finger in practice. Thereby, each user was instructed to perform both a simple *Reaching and Grasping* activity and a more complex one, a short *Phone Conversation*. The complete framework that is proposed herein includes the tracking of the user's head, arm's and fingers' joints via special equipment (Section 7.2).

7.1. Integrated system architecture

The system that has been set up for the experiments' execution is a three-layered system (i.e. Tracking, Feature Extraction and Decision Taking), which consists of two trackers for the arm movement and one tracking device for the movement of the fingers, as shown in Fig. 7.

As it can be easily noted, the proposed framework aims at exploiting a series of motion-related features for user authentication. Following both Hoff's assumption [33] about the two distinct phases of a prehension movement and the theoretical background provided in Sections 3.1 and 3.2, a two-fold approach is followed here.

7.2. Tracking module

In this section, the methods utilized for tracking the human motion during a prehension movement are described. Specifically, two methods are outlined regarding the arm movement, a camera based and a (magnetic) sensor based, which also forms the ground truth. Further, the finger movements are tracked by the CyberGlove®, a dataglove,[2] which converts the deformations on its surface into real angle values (Fig. 8).

[2] The utilization of the glove significantly reduces (if not eliminates) the unobtrusiveness offered by activity-related (and vision-based) biometric approaches. However, the scope of the current paper is the demonstration of the recognition potential of prehension biometrics. In other words, this paper forms a study of activity related features that aims to suggest efficient and robust algorithms for their processing during certain prehension movements, which resemble the presented ones. The major improvement over previous works is the framing of the reaching biometrics in the more generic concept of prehension biometrics.

7.2.1. Tracking of reaching movement

The core of the utilized vision-based upperbody tracking[3] is presented in [19] and is briefly described in the following, so as to make the paper self-contained. The user's movements are recorded by a stereo camera and the raw captured images are processed, in order to track his/her head and hands via the successive application of filtering masks on the captured image. Specifically, a skin-colour mask combined with a motion-mask can provide the location of the palms, while the head can be accurately tracked via a combination of a Viola–Jones face detection algorithm and a mean-shift object tracking algorithm. The 2.5*D* information can be easily derived, performing disparity estimation from the input stereoscopic image sequence.

Real 3*D* data are acquired from the disparity values. The origin of the axes at each repetition is the head's initial location. In order to reduce the effect of noise in the calculation of sensitive high order derivatives and to make the signature robust, the differential invariants in the signature undergo a preprocessing, which consists of smoothing via Kalman filtering.

The accuracy of the proposed vision-based tracker was evaluated via the Magnetic Motion Tracker of *Ascension Technology Corp*. Specifically, two small magnetic sensors were mounted on the user's head and hand as indicated by the coloured spots in Figs. 1 and 8, during the execution of the experiment. Simultaneously, the user's head and hand were tracked by the stereoscopic camera. A series of post-processing algorithms [19] applied to the raw tracked points of both trackers, extracted smooth motion trajectories which were then used as biometric signatures. The comparison between the derived motion trajectories (Fig. 9(a)) demonstrates the capabilities of the proposed tracker.

The small offset that can be seen in the trajectories of the camera tracker was mainly caused by the fact that the magnetic tracker was mounted at the user's wrist, while the camera tracker detects the gravity center of the palm. It turns out that, although not being able to capture the motion in a detail as the magnetic tracker, the performance of the proposed visual tracker is satisfactory enough for the needs of the current experiment, as it will be proven by the authentication results (Section 7.4).

7.3. Tracking the grasping movement

In order to cover the second part of a prehension movement, the tracking of the fingers during the grasping activity is required. The device that has been utilized for this scope is the CyberGlove®. It provides the angles between the phalanxes of the hand by translating into current changes caused by the deformations of integrated thin metallic layers. In this respect, the 3*D* reconstruction of the hand is possible for visual verification of the tracking. Specifically, each finger has been assigned 4 Degrees of Freedom (DoF), while it consists of 3 phalanxes (Fig. 8(a)). Similarly, another 3 DoFs have been assigned to the palm's base.

At this point, it should be noted that according to the authors knowledge, there is no available vision-based tracker for detecting

[3] Contrary to the finger tracking device, the tracking of the upperbody is achieved via the incorporation of a vision based head/hand tracker has been studied, which is able to produce acceptable recognition accuracy, although slightly lower that its magnetic (more obtrusive) alternative.

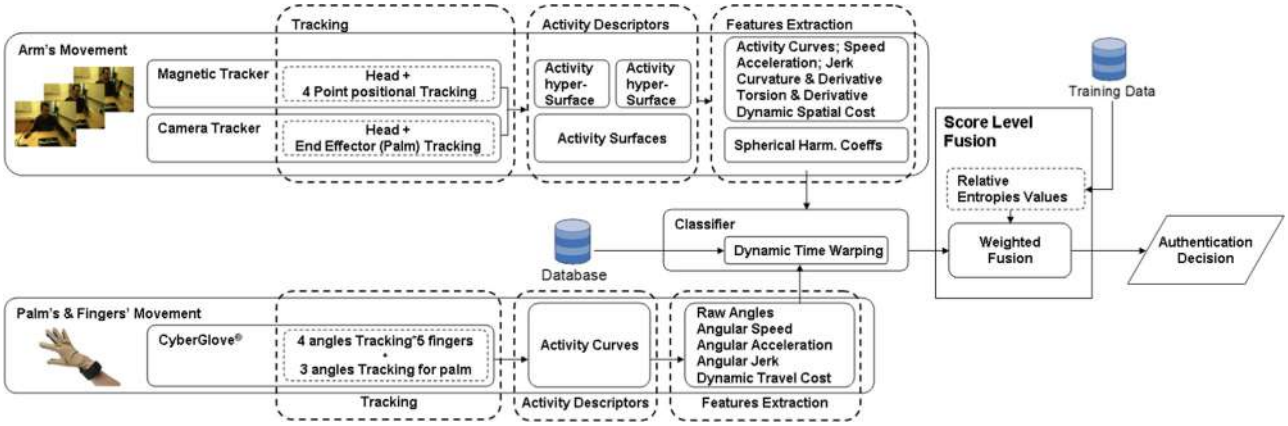

Fig. 7. Overview of the proposed system: Regarding the upperbody movements, the Magnetic and the Camera trackers track the 4 (i.e. head, shoulder, elbow and wrist) and the 2 (i.e. head and wrist) points of interest of the human body , respectively, and thus, both the Activity hyper-Surface and the simplified Activity Surface are generated. Two complementary sets of activity related features are extracted then by each of the aforementioned descriptors, as shown in the 'Feature Extraction' building block in the top right corner of the image. Regarding the movement of the fingers, 23 angles in total are tracked by the CyberGlove device and the corresponding Activity Curves for the angles are produced. Then, the indicated features are extracted. Finally, the DTW classifier forwards a matching probability for each feature separately in the fusion module, which produces the final recognition result, by taking into account the distinctiveness and the uniqueness/redundancy of each feature. Some more detailed views of the proposed architecture can be found in http://www.iti.gr/~drosou/PrehensionBiometrics/SystemArchitecture.pdf.

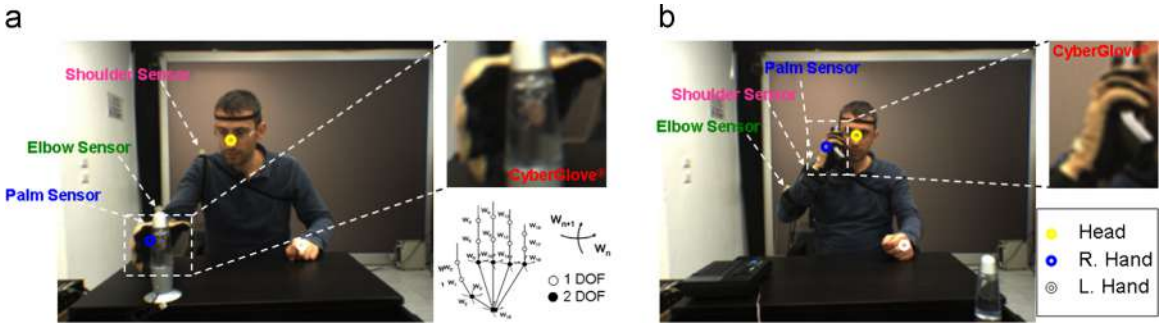

Fig. 8. Simultaneous data recording from three tracking sources during (a) a *Reach and Grasp* activity and (b) a *Phone Conversation*. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

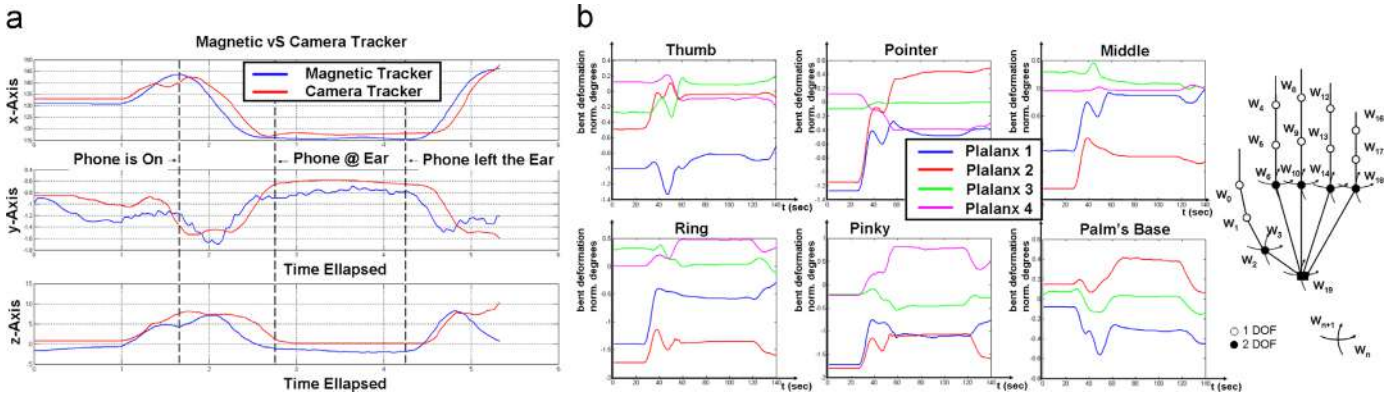

Fig. 9. (a) Comparison between vision-based tracker and ground truth (magnetic tracker). (b) Finger phalanxes' angles. (c) Notation of finger names.

and recognizing the palm gestures accurately over time. Moreover, the development of such a tracker is a complicated task that is out of the scope of the current work. Thus, the dynamics of the palm and finger movements will be studied only via the utilization of the CyberGlove® sensor. However, this fact does not reduce the level of unobtrusiveness in future prehension based biometric systems, when appropriate trackers will become available. Last but not least, it should be noted that finger based biometrics are studied hereby, for reasons of completeness of the prehension based movements and act supportively to the already proven authentication potential of the arm based movements [19].

Herein, a set of postprocessing actions is performed on the data derived by the CyberGlove® device. In particular, the raw data underwent some filtering and processing on the timeseries information, such as resampling, smoothing via low-pass filtering for the removal of artificially generated peaks.

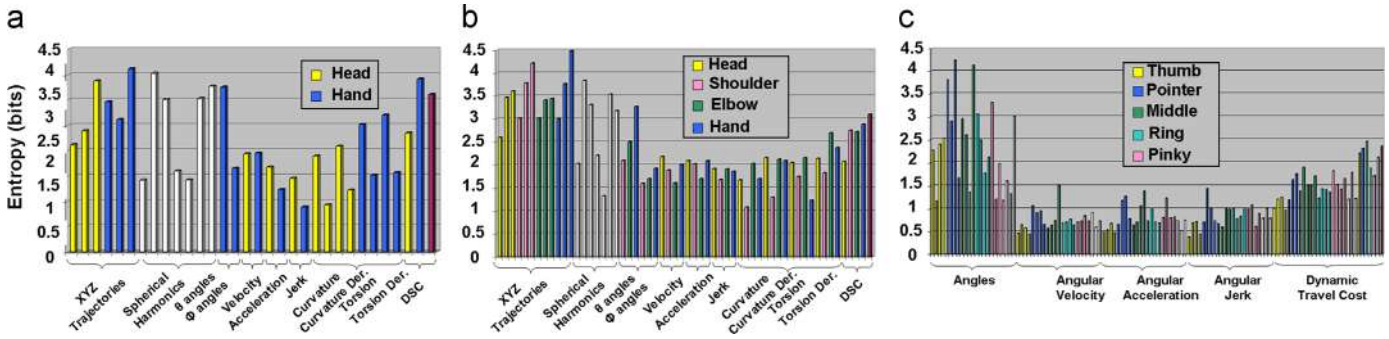

Fig. 10. Relative entropy values from features extracted by (a) Camera Tracker, (b) Magnetic Tracker, (c) CyberGlove® for the "Reaching and Grasping" experiment. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

7.4. Experimental evaluation of the extracted features in a lab-environment database

The experimental setup includes one testing action that is performed by 29 subjects, as many as the population of the dataset in [19]. In particular, each user has been asked to perform an activity denoted as a raw "Reaching and Grasping" task (Experiment 1): the user had to lean forward and grasp a lamp standing on the desk. The second activity was a short "Phone Conversation" (Experiment 2). In particular, the user had to pick up the ringing phone with his/her right hand, bring it next to the ear, hold a short conversation and put it back on its base.

Each experiment was repeated by each user 6 times. The first 3 were used for the enrollment of the user (gallery), while the 4th and the 5th repetitions were used for testing. The 6th repetition of each user was kept as backup for corrupted recordings. The users had been advised to act free and no special constraints were imposed. Repetition 4 was recorded immediately after the enrollment sessions. As such, there is a high resemblance in both the attitude and the movement of the user. On the contrary, the 5th repetition was recorded in a different time session (i.e. after a day), in order to test the permanence of the proposed traits over time.

Regarding Experiment 2, a post-processing algorithm was applied on the extracted trajectories in order to compensate for the following issue. Given that the duration of a phone conversation may be of arbitrary length, the meantime between the moment the phone reaches the ear and the moment the phone leaves the ear is rejected from the trajectory.

7.4.1. Feature evaluation

Provided the features extracted as described in Sections 4 and 5 and the evaluating tools presented in Section 6.1, an analysis was conducted in order to investigate the content of a biometric feature representation of each single individual, with respect to the whole population. More specifically, the general distribution f_i^{intra} was constructed, by using all measurements of feature i from all users. Similarly, f_i^{inter} was constructed by using all measurements of feature i from the training sessions of a single user.

The extracted biometric traits were grouped with reference to the tracker used for each experiment. The relative entropy values are exhibited in Figs. 10 and 11.

One can notice the high discriminative capacity of most of the novel Spherical Harmonic Coefficients as activity-related features. Given the low relative entropy values (in bits) of specific reference points, one can conclude that these views of the Activity Surfaces are characterized by a large number of intersections (see Section 4.1). Equally interesting is the fact that the spatial cost of the hand is of high discriminative capacity. Intuitively, it can be claimed that the larger the total spatial cost, the larger the user and vice versa.

Regarding the Cyberglove® features, one can see that the most indicative features are the angles and the Dynamic Travel Costs (DTC) of each finger, while angular velocity and acceleration of some phananxes may provide enhanced distinctiveness among users. The red line stands for the total DTC, summed up over all fingers.

Still, among the most indicative features, there may still be redundancy, given that it is very likely that some features are not independent. In order to detect them, the extracted features are evaluated with respect to their inter-dependency, via their mutual information $I(F_i^{inter}, F_j^{inter})$.

Given the vast number of utilized features, the confusion matrices[4] are difficult to be read, the most important findings are discussed hereby. First, a high dependency value is exhibited between the features associated with elbow and hand movement. Similar quite high dependence has been detected between the shoulder's and the head's movement, as it is expressed via the extracted features (i.e. activity curves, orientation vectors, curvature, etc.). These findings verify Lacquaniti et al.'s assumption [35] about the strong correlation of all the joints of the arm during a prehension movement. Finally, the full spatial cost is highly related with the hand's spatial cost, especially in the *Phone Conversation* experiment.

Regarding the fingers' movement, let us first assign the following identification letters to each finger: *a*-thumb, *b*-pointer, *c*-middle, *d*-ring and *e*-pinky. During both experiments, it was noticed that there was high dependency in the angles' movement of all joints of fingers *d* and *e*. An equally high dependency was detected in the movement of the base's angles of fingers *b*, *c*, *d* and *e* during the *Reaching and Grasping* movement, while the movement of the pointer's base was differentiated significantly during the *Phone Conversation* experiment. Similarly, to the above, the full travel cost of all fingers was roughly the same.

Finally, in order to estimate the optimal number of most indicative features that should be used for authentication, the following process was attempted. In particular, an alternative approach to a classification problem following the basic principles of typical classification techniques (e.g. minimum redundancy maximum relevance (mRMR)) is utilized herein, taking into account both the Kullback–Leibler divergence (i.e. relative entropy) for evaluating each feature individually and the mutual entropy for co-evaluating the correlations between all features.

For each experiment (i.e. *Reach and Grasp* and *Phone Conversation*) and for each tracking device (i.e. Camera Tracker, Magnetic Tracker and CyberGlove®) the Equal Error Rate value was calculated, as a function of utilized features (Fig. 12), starting from 1 to the total number of extracted features $N_{movement,tracker}$, with respect to the tracker and the movement studied. Based on the confusion matrix with the mutual entropies, the $n^i_{movement,\ tracker}$ features are preserved

[4] www.iti.gr/~drosou/PrehensionBiometrics/MIConfusionMatrices.pdf

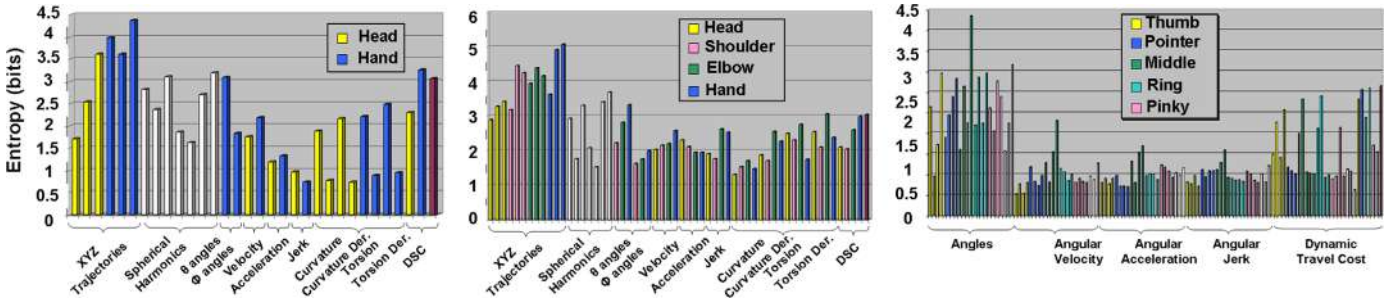

Fig. 11. Relative entropy values from features extracted by (a) Camera Tracker, (b) Magnetic Tracker, (c) CyberGlove® for the "Phone Conversation" experiment. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

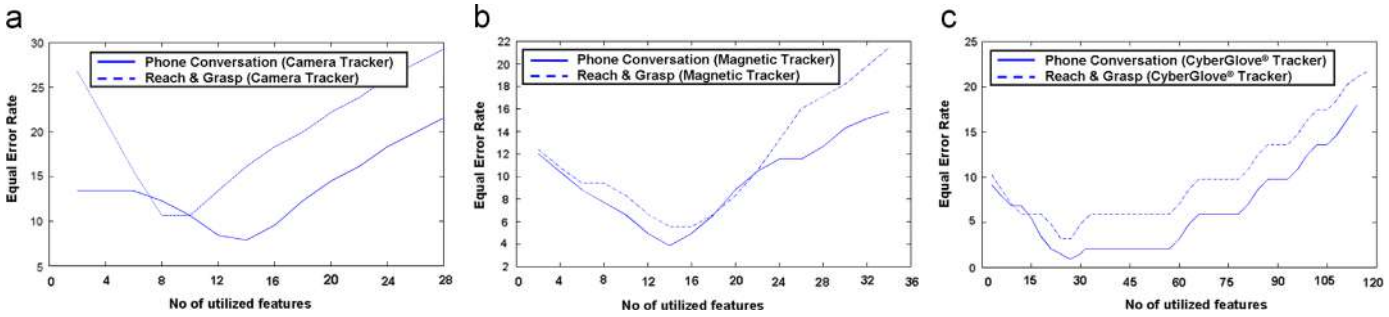

Fig. 12. The equal error rate value as a function of the number of the utilized most indicative features—applied in the decreasing order of relative entropy—for (a) Camera Tracker, (b) Magnetic Tracker, (c) CyberGlove®.

that have the highest relative entropy value and are not strongly correlated with others. The index i denotes the number of the current iteration of the algorithm (i.e. $n^{i+1}_{movement,tracker} = n^{i}_{movement,\ tracker} + 1$). Each utilized feature had undergone a min–max normalization, while the classification at this stage was performed with the Dynamic Time Warping Algorithm (see Section 6.2). This way an EER score is estimated in the testing dataset and noted down, while the algorithm proceeds to the next iteration.

As it has become evident up to now, during the carried out experiments, reaching and grasping movements have been recorded separately by two autonomous monitoring devices, and thus, these distinct processing procedures can be regarded as non-correlated. Following this, in the absence of one monitoring device, the other is not affected, while the recognition potential of the latter remains the same. An improvement in the biometric recognition occurs due to the combination of two distinct traits (see Table 2), as it is the general case in multi-biometric approaches that are properly fused, in the absence of redundancy between the features.

It has been noted that after a certain number of utilized features $n_{movement,\ tracker}$, the authentication performance decreases (i.e. EER score increases accordingly in Fig. 12), since the use of less distinctive or redundant features has a negative effect to the authentication performance of the system. Thus, the indicated unimportant features will be discarded from the authentication procedure, since their utilization has a negative contribution. In this respect, features with high mutual information values with others can be discarded, without serious loss in the overall discrimination capacity of the system. Thus, the table in Fig. 13 includes the features per activity that are maintained, as the most valuable ones (a description of the notations used in this figure can be found in Fig. 10(a) and (b) for the joints of the arm and in Fig. 13(c) for the joints of the fingers).

Additionally in Fig. 12, the reader can notice that the minimum authentication error of the camera tracker is significantly larger than the one derived from the magnetic tracker. This is to be explained by the fact that sometimes the camera tracker fails to capture accurately the velocity and acceleration information of the movement, by being more sensitive to noise from variable illumination and shadows. Although this does not affect the general form of the trajectory, it causes some unavoidable flickering around the tracked point (head and hand) along the frame sequence, which is crucial for capturing the velocity, acceleration and jerk information.

Table 2
Overall authentication errors after final fusion.

Experiment Name	*Session*1 (%)	*Session*2 (%)	*Virtual subjects* (%)
Phone Conversation			
Camera Tracker & CyberGlove	1.1	3.9	4.7
Magnetic Tracker & CyberGlove	0.8	3.3	3.6
Reaching and Grasping			
Camera Tracker & CyberGlove	6.8	7.7	8.2
Magnetic Tracker & CyberGlove	2.6	6.2	6.4

In order to verify the *Permanence in Time* requirement of our biometric approach, the same users were asked to perform the same activities in a different time session (i.e. after a day). The aforementioned findings regarding the optimal number of preserved, most discriminative features, were utilized herein and the authentication and identification results are shown with the help of ROC and CMS curves, respectively (Fig. 14). Similar ROC curves generated from non-optimal amounts of indicative features are suggestively illustrated, as well, in order to exhibit the system's non-optimal performance.

The reader can notice a degradation of $< 5\%$, for the optimal number of features per tracking device. However, this performance is improved when the authentication scores of both phases of the prehension movement are used simultaneously (see Table 2), as expected.

7.5. Experimental evaluation in a realistic environment with the ACTIBIO database

The findings of the previous section were applied to the real environment of the ACTIBIO database [19], where similar recordings regarding the reaching movement were included. More specifically, the ACTIBIO database consists of another 29 subjects performing everyday office activities, such as a phone conversation and an interaction with a microphone. The second activity includes the leaning of the user towards a microphone, the talking and the getting back in his/her original position. This activity was selected for evaluation because it resembles the reaching phase of a prehension movement.

In order to verify the robustness of the proposed method in the real environment of the ACTIBIO database, the most significant features (see Fig. 13 in Section 7.4.1) were extracted. Fig. 15 illustrates the variations of the system's performance in terms of EER scores for different numbers of the most indicative features (table in Fig. 13). Unfortunately, the ACTIBIO database does not include measurements with CyberGlove. Thus, the evaluation of our framework was based only on the proposed camera tracker (Section 7.2.1).

There was an improvement in the system's performance compared to the results reported in [18], whereby the simple utilization of the Activity Curves exhibited an EER of $>15\%$ and 10% for the two activities.

7.6. Experimental evaluation in a large synthetic database

Finally, in order to study the performance of the proposed method in larger population, it was evaluated in another database of 100 virtual subjects that was created as follows. Let us define the "mean" trajectory and the "mean" velocity for each limb as the average trajectory-velocity of all available enrollments (black line in Fig. 16). Given the estimated first and second order statistics among all users' activity curves $(m^{inter}(t),\ \sigma^{inter}(t))$ from the proprietary dataset (Section 7.4), 100 new "base-features" were created. We used

$$\mathbf{s}_{\mathbf{l}}^{*}(\mathbf{t}) = \mathbf{m}_{\mathbf{s,l}}^{\mathbf{inter}}(t) + \mathbf{n}_{\mathbf{s,l}}(\mathbf{t}), \tag{26}$$

where $\mathbf{n}_{\mathbf{s,l}}(\mathbf{t})$ was a random number drawn from a normal distribution with 0 mean and standard deviation $\sigma_{s,l}^{inter}(t)$, regarding the lth joint.

In order to minimize the effect of flickering along a feature signal, generated this way, we used a low-pass filtering method via a moving average window. Finally, new "Repetitions" (see Fig. 16) of each virtual subject were generated by using the detected intra-variance $(m^{intra}(t),\ \sigma^{intra}(t))$ of each subject. Similarly, virtual velocity vectors can be generated, i.e. $\mathbf{v}_{\mathbf{v,l}}^{*}(\mathbf{t}) = \mathbf{m}_{\mathbf{v}}^{\mathbf{inter}}(\mathbf{t}) + \mathbf{n}_{\mathbf{v,l}}(\mathbf{t})$ and the corresponding intra-parameters.

	Camera Tracker		Magnetic Tracker		CyberGlove®	
Initially extracted features	**6** for the raw trajectories, **7** for Spherical Harmonics Coefficients, **2** for orientation, **8** for the curvature/ torsion and derivatives, **6** for the velocity, acceleration and jerk of each joint and **2+1** for the Dynamic Spatial Cost		**12** for the raw trajectories, **7** for Spherical Harmonics Coefficients, **2** for orientation, **16** for the curvature/ torsion and derivatives, **12** for the velocity, acceleration and jerk of each joint and **4+1** for the Dynamic Spatial Cost		**23** for the phalanxes' angles, **69** the angular velocity, angular acceleration and angular jerk and **23+6+1** for the dynamic travel	
Finally preserved features	**Phone Conversation**	**Reach & Grasp**	**Phone Conversation**	**Reach & Grasp**	**Phone Conversation**	**Reach & Grasp**
	Y,Z Trajectories for Head; X,Y,Z Trajectories for Hand; **5** Spherical Harmonics; angle for orientation; DSC for Hand; Hand velocity; Hand torsion.	Y,Z Trajectories for Head; X,Y,Z Trajectories for Hand; **4** Spherical Harmonics; θ angle for orientation; DSC for Hand; Hand velocity; Hand torsion	X,Y,Z Trajectories for Head; X,Y,Z Trajectories for Hand; **4** Spherical Harmonics; θ angle for orientation for Palm; Palm trajectory for dynamic Spatial Cost; Palm velocity; Palm jerk; curvature of Hand Trajectory.	X,Y,Z Trajectories for Head; X,Y,Z Trajectories for Hand; **4** Spherical Harmonics; Palm orientation; Palm Dynamic Spatial Cost; Palm curvature; Palm torsion; 1st derivative of Palm torsion.	*Finger a:* W_0, W_2 and W_3 angles; W_2 total travel cost. *Finger b:* W_4, W_5 and W6 for angles, W_5 jerk, total travel cost; *Finger c:* W_8, W_9 and W10 for angles, W_{10} for velocity, total travel cost; *Finger d:* W_{12}, W_{13}, W_{14} for angles, total travel cost; *Finger e:* -; Palm: W_{19} for angles (i.e. transverse palm movement), W_{19} travel cost.	*Finger a:* W_2 and W_3 angles, W_3 and total travel cost; *Finger b:* W_5, W_6 and W_6 angles, total for total travel cost; *Finger c:* W_9, W_{10} and W_{11} for angles, W_{11} for velocity, total travel cost; *Finger d:* W_{12} and W_{13} for angles, W_{13} and total for travel cost; *Finger e:* -; Palm: W_{19} and W_{20} for angles (i.e. transverse and sagittal palm movement).

Fig. 13. Summary of the initially extracted and the remaining (i.e. most valuable, in terms of authentication capacity, features per tracking device) features. The notations for the finger joints are explained in Fig. 9(c).

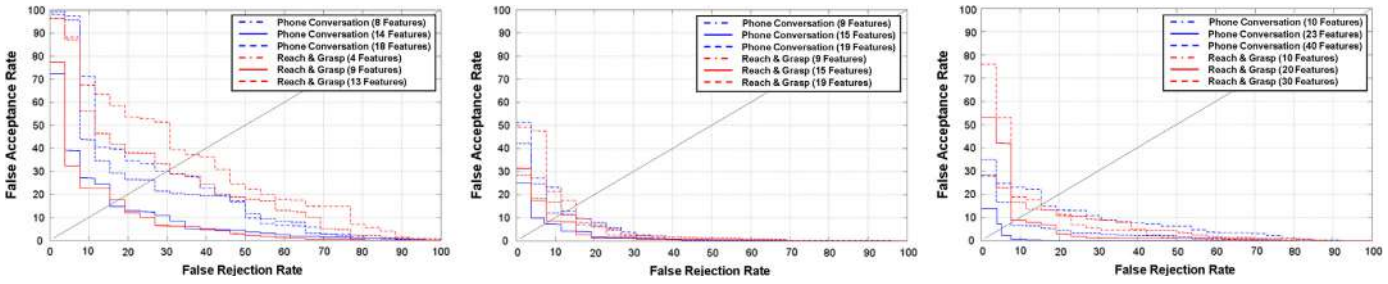

Fig. 14. ROC curves for the performed activities in Session 5 sorted by tracking device: (a) Camera Tracker, (b) Magnetic Tracker, (c) CyberGlove. The corresponding CMS curves can be found in http://www.iti.gr/~drosou/PrehensionBiometrics/CMSGraphs.pdf.

This way, given a set of virtual activity curves and the corresponding velocity vector, one can easily estimated the virtual activity's duration t^*. The rest of the features were then extracted for each subject as described in Sections 4 and 5. In Fig. 16, one can see some samples of virtual features from different users. It can be noticed that the physical notion of the activity curves is preserved.

Similar to the experiments performed in the previous two databases, the changes of the EER scores for different number of utilized features are presented in Fig. 17 for the two proposed activities.

It should be noted that the results from the synthetic database should be compared with the ones of Session 2 (Fig. 14), given that the generated trajectories for the virtual subjects were based on statistics from all sessions.

Finally, it is expected that by fusing the outcomes of the two phases for each activity, the overall authentication performance of the system will be improved. The fused results of the proposed framework are presented in Table 2. It should be noted that the fusion was performed as described at the end of Section 6.2.

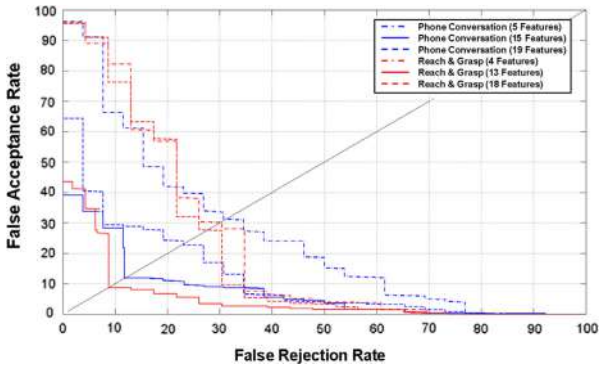

Fig. 15. ROC curves for the performed activities in the ACTIBIO Database, as they were recorded by the Camera Tracker. The corresponding CMS curves can be found in http://www.iti.gr/~drosou/PrehensionBiometrics/CMSGraphs.pdf.

8. Conclusion

In this paper, a novel descriptor for prehension movements was presented. Based on this, a series of activity related features were extracted which capture the dynamic characteristics of reaching and interacting with objects to be used for biometric authentication. The authentication potential of these features was estimated according to their relative entropies, with inter-dependencies detected via the mutual information of the systems.

Although the presented study has shown promising results regarding the authentication potential, the application of such biometrics in real case scenarios, as well as the level of unobtrusiveness it offers is highly dependent on the quality of tracking, as

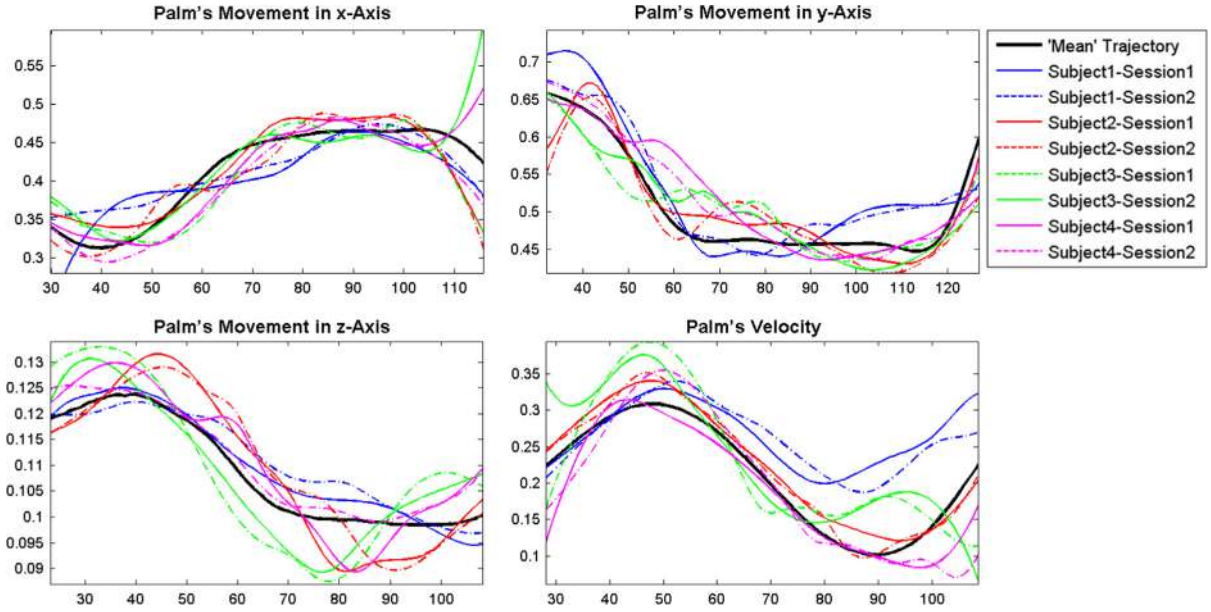

Fig. 16. Multiple repetitions of virtual subject's features.

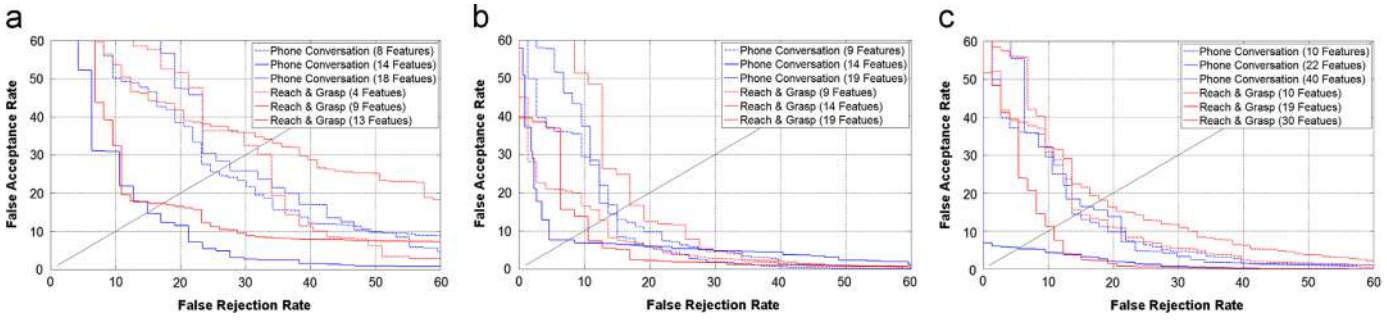

Fig. 17. ROC curves for the performed activities in the synthetic database sorted by tracking device: (a) Camera Tracker, (b) Magnetic Tracker, (c) CyberGlove. The corresponding CMS curves can be found in http://www.iti.gr/~drosou/PrehensionBiometrics/CMSGraphs.pdf.

it has been observed from the comparison between the vision-based and the sensor-based arm tracker. Thus, future trackers are expected to be significantly valuable for the actual incorporation of the proposed modality in actual biometric systems.

The proposed features were tested in two scenarios of three databases (i.e. two medium sized and a large one) and the experimental results showed that the proposed method can achieve high rates of authentication performance. The proposed method can be integrated along with other types of features in a user authentication system, so as to improve its overall efficiency. For instance, prehension biometrics can offer a robust modality for both those who are unwilling to be exposed to inconvenient processes (e.g. iris scan and fingerprint scan), as well as an integral part of an "on-the-go" authentication system [20]. Last but not least, prehension biometrics are recommended for continuous, transparent authentication, so as to renew the validity of the claimed ID of the user transparently.

Conflict of interest

None declared.

Acknowledgements

This work was partially supported by the EU funded ACTIBIO ICT STREP (FP7-215372) and by the Greek national research project BIOTAFTOTITA.

References

[1] Q.t. Xiao, Security issues in biometric authentication, Information Assurance Workshop, IAW 2005, pp. 8–13, 2005.
[2] A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, IEEE Trans. Circuits Syst. Video Technol. 14 (1) (2004) 4–20.
[3] M. Garris, E. Tabassi, C. Wilson, NIST fingerprint evaluations and developments, Proc. IEEE 94 (11) (2006) 1915–1926.
[4] Anil K. Jain, David Maltoni., Handbook of Fingerprint Recognition, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2003.
[5] N. Fox, R. Gross, J. Cohn, R. Reilly, Robust biometric person identification using automatic classifier fusion of speech, mouth, and face experts, IEEE Trans. Multimed. 9 (4) (2007) 701–714.
[6] Z. Sun, T. Tan, Ordinal measures for Iris recognition, IEEE Trans. Pattern Anal. Mach. Intell. 31 (12) (2009) 2211.
[7] N. Schmid, M. Ketkar, H. Singh, B. Cukic, Performance analysis of Iris-based identification system at the matching score level, IEEE Trans. Inf. Forensics Secur. 2 (1) (2006) 154–168.
[8] G. Zheng, C.-J. Wang, T. Boult, Application of projective invariants in hand geometry biometrics, IEEE Trans. Inf. Forensics Secur. 2 (4) (2007) 758–768.
[9] A. Jain, J. Feng, Latent palmprint matching, IEEE Pattern Anal. Mach. Intell. 31 (2009) 1032–1047.
[10] K. Delac, M. Grgic, A survey of biometric recognition methods, in: Proceedings of 46th International Symposium Electronics in Marine, Elmar 2004, 2004, pp. 184–193.
[11] H. Junker, J. Ward, P. Lukowicz, G. Tröster, User activity related data sets for context recognition, in: Proceedings of Workshop on 'Benchmarks and a Database for Context Recognition', 2004.
[12] M. Goffredo, I. Bouchrika, J.N. Carter, M.S. Nixon, "Self-Calibrating View-Invariant Gait Biometrics," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on , vol.40, no.4, pp.997,1008, Aug. 2010 doi: http://dx.doi.org/10.1109/TSMCB.2009.2031091 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5299188&isnumber=5510034.
[13] A. Hadid, M. Pietikäinen, S.Z. Li2, Learning personal specific facial dynamics for face recognition from videos, in: Analysis and Modeling of Faces and Gestures, Springer, Berlin, Heidelberg, 2007, pp. 1–15.
[14] D. Ioannidis, D. Tzovaras, I.G. Damousis, S. Argyropoulos, K. Moustakas, Gait recognition using compact feature extraction transforms and depth information, IEEE Trans. Inf. Forensics Secur. 2 (3) (2007) 623–630.
[15] M. Goffredo, I. Bouchrika, J.N. Carter, M.S. Nixon, Performance analysis for automated gait extraction and recognition in multi-camera surveillance, Multimed. Tools Appl. 50 (2009) 75–94.
[16] A. Kale, N. Cuntoor, R. Chellappa, A framework for activity-specific human identification, in: IEEE Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 4, 2002, pp. 3660–3663.
[17] A. Kale, A. Sundaresan, A. Rajagopalan, N.P. Cuntoor, A.K. Roy-Chowdhury, V. Kruger, R. Chellappa, Identification of humans using gait, IEEE Trans. Image Process. 13 (2004) 1163–1173.
[18] A. Drosou, K. Moustakas, D. Ioannidis, D. Tzovaras, On the potential of activity-related recognition, in: Proceedings of the International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP), 2010, pp. 340–348.
[19] A. Drosou, D. Ioannidis, K. Moustakas, D. Tzovaras, Spatiotemporal analysis of human activities for biometric authentication, J. Comput. Vis. Image Underst. 116(3) (2012) 411–421 (Special Issue on Semantic Understanding of Human Behaviors in Image Sequences).
[20] A. Drosou, G. Stavropoulos, D. Ioannidis, K. Moustakas, D. Tzovaras, Unobtrusive multi-modal biometric recognition approach using activity-related signatures, IET Comput. Vis. 5 (6) (2011) 367–379.
[21] H.P. Jain, A. Subramanian, Real-time upper-body human pose estimation using a depth camera, in: Computer Vision/Computer Graphics Collaboration Techniques, vol. 6930, 2011, pp. 227–238.
[22] M. Chutorian, M. Trivedi, Head pose estimation in computer vision: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 31 (4) (2009) 607–626.
[24] L. Brown, C. Moore, D. Rosenbaum, Feature-specific perceptual processing dissociates action from recognition, J. Exp. Psychol.: Hum. Percept. Perform. 28 (6) (2002) 1330–1344.
[25] A. Churchill, B. Hopkins, L. Roennqvist, S. Vogt, Vision of the hand and environmental context in human prehension, Exp. Brain Res. 134 (2000) 81–89.
[26] J. Vaughan, D. Rosenbaum, R. Meulenbroek, Planning reaching and grasping movements: the problem of obstacle avoidance, Motor Control 2 (2001) 116–135.
[27] M. Pantic, A. Nijholt, A. Pentland, T. Huanag, Human-centred intelligent human-computer interaction (HCI): how far are we from attaining it? Hum. Centred Intell. Hum. Comput. Interact. 1 (2) (2008) 168–187.
[28] K. Moustakas, M. Strintzis, D. Tzovaras, S. Carbini, O. Bernier, J. Viallet, S. Raidt, M. Mancas, M. Dimiccoli, E. Yagci, S. Balci, E. Leon, Masterpiece: physical interaction and 3D content-based search in VR applications, IEEE Multimed. 13 (3) (2006) 92–100.
[29] T. Flash, N. Hogan, The coordination of arm movements: an experimentally confirmed mathematical model, J. Neurosci. 5 (7) (1985) 1688–1703.
[30] Y.K. Uno, R.M. Suzuki, Formation and control of optimal trajectory in human multijoint arm movement—minimum torque-change model, Biol. Cybern. 61 (2) (1989) 89–101.
[31] M. Turvey, Perceiving, Acting, and Knowing: Toward an Ecological Psychology, Lawrence Erlbaum, New Jersey, 1977.
[32] S.R. Goodman, G. Gottlieb, Analysis of kinematic invariances of multijoint reaching movement, Biol. Cybern. 73 (1995) 311–322.
[33] B. Hoff, M.A. Arbib, Models of trajectory formation and temporal interaction of reach and grasp, Motor Behav. 25 (3) (1993) 175–192.
[34] D.A. Rosenbaum, R.J. Meulenbroek, J. Vaughan, C. Jansen, Posture-based motion planning: applications to grasping, Psychol. Rev. 108 (4) (2001) 709–734.
[35] F. Lacquaniti, J.F. Soechting, Coordination of arm and wrist motion during a reaching task, J. Neurosci.: The Official Journal of the Society for Neuroscience 2 (April) (1982) 399–408.
[36] S. Jaric, D. Corcos, G. Gottlieb, D. Ilic, M. Latash, The effects of practice on movement distance and final position reproduction: implications for the equilibrium-point control of movements, Exp. Brain Res. 100 (2) (1994) 353–359.
[37] M. Wiesendanger, O. Kazennikov, S. Perrig, P. Kaluzny, Two handsone action: the problem of bimanual coordination, in: A.H. Wing, P. Haggard, J.R. Flanagan (Eds.), Hand and Brain: Neurophysiology and Psychology of Hand Movement, 1996, pp. 283–300.
[38] A. Vogiannou, K. Moustakas, D. Tzovaras, M. Strintzis, A first approach to contact-based biometrics for user authentication, in Proceedings of the 3rd International Conference on Advances in Biometrics, 2009, pp. 838–846.
[39] J. Aleotti, S. Caselli, Grasp recognition in virtual reality for robot pregrasp planning by demonstration, in: IEEE International Conference on Robotics and Automation (ICRA), 2006, pp. 2801–2806.
[40] E. Kukula, S. Elliott, Implementation of hand geometry: an analysis of user perspectives and system performance, in: IEEE Aerospace and Electronic Systems Magazine, 2006, pp. 3–9.
[41] V. Schönefeld, Spherical Harmonics, Technical Report, 2005.
[42] S. Wu, Y.F. Li, Flexible signature descriptions for adaptive motion trajectory representation, perception and recognition, Pattern Recognit. 42 (2009) 194–214.
[43] T. Syeda-Mahmood, Segmenting actions in velocity curve space, in: Proceedings of the 16th International Conference on Pattern Recognition, vol. 4, 2002, pp. 170–175.
[44] N.P. Cuntoor, B. Yegnanarayana, R. Chellappa, Activity modeling using event probability sequences, IEEE Trans. Image Process. 17 (4) (2008) 594–607.
[45] R. Cen, Y. Alper, S. Mubarak, View-invariant representation and recognition of actions, Proc. Int. J. Comput. Vis. 50(2) (2002) 203–226.
[46] S.N. Vitaladevuni, V. Kellokumpu, L.S. Davis, Action recognition using ballistic dynamics, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.
[48] S.P. Priyal, P.K. Bora, A robust static hand gesture recognition system using geometry based normalizations and Krawtchouk moments, Pattern Recognit. 46 (2013) 2202–2219.

[49] W.W. Kong, S. Ranganath, Towards subject independent continuous sign language recognition: a segment and merge approach, Pattern Recognit. 47 (2014) 1294–1308.
[50] O. Aran, L. Akarun, A multi-class classification strategy for Fisher scores: application to signer independent sign language recognition, Pattern Recognit. 43 (2010) 1776–1788.
[51] H.D. Yang, S.W. Lee, Robust sign language recognition by combining manual and non-manual features based on conditional random field and support vector machine, Pattern Recognit. Lett. 34 (2013) 2051–2056.
[52] M. Lexa, Useful Facts about the Kullback–Leibler Discrimination Distance, Technical Report, Department of Electrical and Computer Engineering, Rice University, Houston, 2006.
[53] H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, in: Readings in Speech Recognition, vol. 26, 1990, pp. 159–165.
[54] N. Boulgouris, K. Plataniotis, D. Hatzinakos, Gait recognition using dynamic time warping, in: IEEE 6th Workshop on Multimedia Signal Processing, Siena, 2004, pp. 263–266.

Anastasios Drosou is a post-doctoral research fellow at the Information Technologies Institute at the Centre for Research and Technology Hellas. He received his Diploma in Electrical and Computer Engineering, with an expertise in Telecommunications, from Aristotle University of Thessaloniki, and an MSc. in Communication Electronics from the Technische Universitt Mnchen in 2004 in 2007, respectively. He also holds a PhD in Signal and Image processing from the Imperial College London since 2013. His research interests include biometric security, visualization and network security, regarding the identification. He serves as a regular reviewer for several technical journals and he is the coauthor of more than 25 scientific papers (i.e. 9 journals, 12 conferences and 4 book chapters).

Dimosthenis Ioannidis was born in Larisa, Greece, in 1977. He received the Diploma degree in Electrical Engineering from the Electrical and Computer Engineering Department of the Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece, in 2000, and the M.A. degree in Advanced Communication Systems and Engineering from AUTH, in 2005. He is currently working as a Research Associate with the Information Technologies Institute of the Centre for Research and Technology Hellas, Thessaloniki, Greece. His main research interests are in the areas of biometrics, 3D data processing, and web semantics. Mr. Ioannidis is a member of the Technical Chamber of Greece.

Dimitrios Tzovaras holds a degree in Electrical and Computer Engineering Thessaloniki (1992) and a Ph.D. in the Electrical and Computer Engineering, from the corresponding department of the Polytechnic School of the Aristotle University of Thessaloniki (AUTH). He is a Senior Researcher (Researcher A') in the Information Technologies Institute (ITI) at the Centre for Research and Technology Hellas (CERTH). His main research interests include visual analytics, 3D object recognition, search and retrieval, behavioral biometrics, assistive technologies, information and knowledge management, multimodal interfaces, computer graphics and virtual reality. Dimitrios Tzovaras has been working as a Researcher since September 1999 and he has been involved in more than 60 projects, funded by the EC and the Greek Ministry of Research and Technology. His involvement with those research areas has led to the co-authoring of over 80 articles in refereed journals and more than 150 papers in international conferences. He has served as a regular reviewer for a number of international journals and conferences. He is a member of IEEE and EURASIP.

Konstantinos Moustakas (M'07) received the Diploma degree and the Ph.D. degree in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2003 and 2007, respectively. He has been a Teaching Assistant in the same department (2004–2007) and a Visiting Lecturer during 2008–2009. He served as a Research Associate in the Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki (2003–2007), where he is currently a Postdoctoral Research Fellow. Since 2012, he is an Associate Professor at the Electrical and Computer Engineering Department of the University of Patras. His main research interests include virtual reality, 3-D content-based search, collision detection, haptics, deformable object modelling and simulation, and stereoscopic image processing. During the latest years, he has been the (co)author of more than 60 papers in refereed journals, edited books, and international conferences. He serves as a regular reviewer for several technical journals. He has been also involved in nine research projects funded by the EC and the Greek secretariat of Research and Technology. Dr. Moustakas is a member of the IEEE Computer Society and the Technical Chamber of Greece.

Maria Petrou (A'90-M'91-SM'05) received a degree in physics from the Aristotle University of Thessaloniki, Thessaloniki, Greece, a degree in Applied Mathematics in Cambridge, U.K., the Ph.D. degree from the Institute of Astronomy, Cambridge, and the D.Sc. degree from Cambridge in 2009. She currently holds the Chair of Signal Processing at Imperial College London, London, U.K., and she is the Director of the Information Technologies Institute of the Centre of Research and Technology Hellas, Greece. She has published more than 350 scientific papers on astronomy, remote sensing, computer vision, machine learning, colour analysis, industrial inspection, and medical signal and image processing. She has coauthored two books, Image Processing: The Fundamentals (Wiley, first edition 1999 and second edition 2010) and Image Processing: Dealing With Texture (Wiley, 2006). She has also coedited the book Next Generation Artificial Vision Systems: Reverse Engineering the Human Visual System. She has supervised to successful completion 43 Ph.D. theses. Dr. Petrou is a Fellow of the Royal Academy of Engineering, a Fellow of the City and Guilds Institute, a Fellow of the Institution of Engineering and Technology (IET), a Fellow of the International Association for Pattern Recognition (IAPR), a Fellow of the Institute of Physics, and a Distinguished Fellow of the British Machine Vision Association. She has served as a Trustee of the IET (2006–2009), as the IAPR Newsletter Editor (1994–1998), and as the IAPR Treasurer (2002–2006).