

Imperial College London
Department of Electrical Engineering and Electronic Engineering

Activity related Biometrics for Person Authentication

Anastasios Drosou

04-10-2013

Supervised by Prof. Maria Petrou

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Electronic Engineering
of Imperial College London
and the Diploma of Imperial College London

Declaration of Originality

I hereby declare that:

- this thesis is the record of original work, that has been carried out by me,
- all other referenced work is appropriately referenced,
- this thesis has not been submitted in any previous application for a higher degree,
- I truthfully documented all methods, data and operational procedures,
- I have not manipulated any data,
- I have identified all persons who have substantially supported me in my work in the acknowledgements.

I understand that the above written work may be tested electronically for plagiarism.

Statement of Copyright

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

One of the major challenges in human-machine interaction has always been the development of such techniques, that are able to provide accurate human recognition, so as to offer either personalized services or to protect critical infrastructures from unauthorized access. To this direction, a series of well stated and efficient methods have been proposed mainly based on biometric characteristics of the user. Despite the significant progress that has been achieved recently, there are still many open issues in the area, concerning not only the performance of the systems but also the intrusiveness of the collecting methods.

The current thesis deals with the investigation of novel, activity-related biometric traits and their potential for multiple and unobtrusive authentication based on the spatiotemporal analysis of human activities. In particular, it starts with an extensive bibliography review regarding the most important works in the area of biometrics, exhibiting and justifying in parallel the transition that is performed from the classic biometrics to the new concept of behavioural biometrics.

Based on previous works related to the human physiology and human motion and motivated by the intuitive assumption that different body types and different characters would produce distinguishable, and thus, valuable for biometric verification, activity-related traits, a new type of biometrics, the so-called prehension biometrics (i.e. the combined movement of reaching, grasping activities), is introduced and thoroughly studied herein. The analysis is performed via the so-called Activity hyper-Surfaces that form a dynamic movement-related manifold for the extraction of a series of behavioural features.

Thereafter, the focus is laid on the extraction of continuous soft biometric features and their efficient combination with state-of-the-art biometric approaches towards increased authentication performance and enhanced security in template storage via Soft biometric Keys. In this context, a novel

and generic probabilistic framework is proposed that produces an enhanced matching probability based on the modelling of the systematic error induced during the estimation of the aforementioned soft biometrics and the efficient clustering of the soft biometric feature space.

Next, an extensive experimental evaluation of the proposed methodologies follows that effectively illustrates the increased authentication potential of the prehension-related biometrics and the significant advances in the recognition performance by the probabilistic framework. In particular, the prehension biometrics related biometrics is applied on several databases of ~ 100 different subjects in total performing a great variety of movements. The carried out experiments simulate both episodic and multiple authentication scenarios, while contextual parameters, (i.e. the ergonomic-based quality factors of the human body) are also taken into account. Furthermore, the probabilistic framework for augmenting biometric recognition via soft biometrics is applied on top of two state-of-art biometric systems, i.e. a gait recognition (> 100 subjects)- and a $3D$ face recognition-based one (~ 55 subjects), exhibiting significant advances to their performance.

The thesis is concluded with an in-depth discussion summarizing the major achievements of the current work, as well as some possible drawbacks and other open issues of the proposed approaches that could be addressed in future works.

In the memory of my grandfather...

Acknowledgements

I wish to honourably dedicate the current thesis to my principal supervisor Professor Maria Petrou for having been an example of scientific integrity and intellectual fortitude. Apart from having mentoring me to push my work to the highest standards, she has always been an example of a caring and supportive supervisor.

Moreover, the writing of the current thesis would have never been possible without the support of many loving persons. First of all, special thanks are granted to my collaborative supervisor at the Information Technologies Institute (CeRTH/ITI) Dr. Dimitrios Tzovaras, who has always inspired me and guided me throughout this period.

Furthermore, I wish to thank Mr. Dimosthenis Ioannidis (CeRTH/ITI) and the Assoc. Professor Konstantinos Moustakas (University of Patras) that have patiently supported me with their experience, capabilities and positive thinking. Also, I am kindly thankful to Dr. Christos Bouganis (Imperial College London), who accepted to guide me through the submission process of my thesis.

Last but not least, very special reference deserve my beloved parents (Drosos & Stella), my sister (Amalia), my grandmother (Amalia) and all my friends for having stood by me in all difficult moments of the last $3\frac{1}{2}$ years.

Contents

1. Introduction	22
1.1. Motivation	23
1.1.1. Motivation for Prehension based biometric recognition	24
1.1.2. Motivation for multi-biometric recognition	25
1.1.3. Contribution in activity related biometric recognition	27
1.2. Problem Formulation	28
1.3. Introduction to Prehension Biometrics	30
1.3.1. Reaching (Arm movement)	33
1.3.2. Grasping (palm/fingers movement)	34
1.4. Validity of prehension related features as biometric traits . .	35
1.4.1. Ergonomic factors in Prehension	38
1.5. Originality Achievements of the Thesis	39
1.6. Thesis Outline	42
2. Literature Survey	45
2.1. Anthropometric characteristics & Soft Biometrics	46
2.2. Hard Biometrics	47
2.2.1. Physical Biometrics	47
2.2.2. Behavioural Biometrics	48
2.3. Multi-biometric Approaches	63
2.3.1. Combination of Hard Biometrics	65
2.3.2. Combination of Soft with Hard Biometrics	67
2.4. Summary	69
3. Activity related application scenarios	72
3.1. Context in activity-related biometrics	72
3.2. Activity related datasets	74
3.2.1. Discussion on the recorded datasets	74
3.2.2. Datasets description	76

3.3. Conclusion	87
4. User Recognition using Prehension Biometrics	88
4.1. Reaching - Feature Extraction using Activity hyper-Surfaces .	88
4.1.1. Tracking of Reaching Movement	89
4.1.2. Activity hyper-Surfaces & Feature Extraction	94
4.2. Grasping - Feature Extraction using Activity Curves	106
4.2.1. Tracking the Grasping Movement	106
4.2.2. Angular Speed, Angular Acceleration and Angular Jerk	109
4.2.3. Dynamic Travel Cost	109
4.3. Evaluation of Prehension related Features Authentication Po- tential	110
4.3.1. Relative Entropy and Mutual Information	110
4.4. Classification and User Authentication	111
4.4.1. Gaussian Mixture Model	112
4.4.2. Hidden Markov Models	114
4.4.3. Dynamic Time Warping	118
4.5. Summary	121
5. Enhancement of Biometric Systems using Anthropometric and Soft Characteristics	124
5.1. Static Anthropometric Profile	125
5.1.1. Attributed Graph Matching	126
5.2. Systematic Error Analysis of Soft Biometrics	128
5.2.1. Modelling Soft Biometrics	129
5.3. Summary	135
6. Experimental Evaluation of Prehension Biometrics	137
6.1. Experimental Results of the Prehension based User Recognition	137
6.1.1. Experimental Setup	137
6.1.2. Feature Classification	138
6.1.3. Authentication Capacity of Activity Curves during the Reaching movement	142
6.1.4. Authentication Capacity of Activity Curves for mul- tiple Authentication	147
6.1.5. Authentication Capacity of Spherical Harmonics dur- ing the Reaching Movement	150

6.1.6.	Authentication Capacity of the full Prehension movement	153
6.1.7.	Experimental results in a realistic environment with the ACTIBIO Database	158
6.1.8.	Experimental results in a large synthetic Database	158
6.2.	Experimental results of Prehension based biometrics in a multimodal approach	162
6.2.1.	Gait Recognition results	164
6.2.2.	Activity-Related recognition results	166
6.2.3.	Fusion results	167
7.	Evaluation Static Anthropometric Trait related Enhancements	170
7.1.	Experimental results of the Soft Biometrics based enhancements	170
7.1.1.	Enhancing Gait Recognition with Soft Biometrics	170
7.1.2.	Enhancing Face Recognition with Soft Biometrics	179
8.	Conclusions	186
8.1.	Summary of the Thesis	186
8.2.	Critical Discussion	189
8.3.	Future Work	191
	Bibliography	193
	Appendix A. Vision-based UpperBody Tracking Algorithm	228
A.1.	Face Detection - Tracking	228
A.1.1.	AdaBoost learning and cascade structure of classifiers	230
A.1.2.	Head Tracking	232
A.2.	Skin Colour	234
A.3.	Background Removal	236
A.4.	Motion Detection	237
	Appendix B. Geometric Gait Features Extraction	239
B.1.	Geometric Gait Recognition	239
B.2.	Height and Stride-length Estimation	242
B.3.	Detection of stops in a gait silhouette sequence	243

B.4. Walking angle estimation and compensation	244
B.5. Genetic fusion algorithm	247

List of Figures

1.1.	(a) During a <i>Reaching Movement</i> shoulder and elbow angles change in a predefined way - (b) During a <i>Grasping Activity</i> the fingers' and the palm's angles are moving towards the hand's final posture	33
1.2.	Human convenience zones. The dark shaded area marks the so-called " <i>Convenient Zone</i> ", while the light shaded area marks the so-called " <i>Kinetosphere</i> ".	39
3.1.	Screenshots of several subjects from the ACTIBIO Reaching Dataset (DB.P.1) performing the "Phone Conversation" activity.	77
3.2.	Multiple Repetitions of several features of the Virtual Subjects.	80
3.3.	A user playing the Stroop test. The real world coordinates of the his/her joints are automatically provided by the Microsoft SDK [5] (skeleton), while the Galvanic Skin Response (GSR) and the ElectroCardioGram (ECG) signal are concurrently recorded via the attached sensors.	81
3.4.	Illustration of the 5 different possible manipulation gestures of the user during the game. The screenshot of the Game Layout refers to the (1 st Trial): The buttons are fixed and the reference font colour matches the displayed colour name (waiting time=4 seconds). In the current snapshot, the correct answer is selected by Movement #1. It should be noted that the presented curves do not necessarily represent real movements, in terms of curvature , but they are drawn so for illustrative reasons.	82

3.5.	Game instances - (a) Trial 2: The buttons are fixed but the reference font colour does not match the displayed colour name (waiting time=3 seconds) - (b) Trial 3: The buttons are not fixed and the reference font colour does not match the displayed colour name (waiting time=2 seconds).	83
3.6.	Screenshots of several subjects from the BIOTAFTOTITA 3D Face Dataset (DB.F.1) during the probe recordings of their faces under several poses/angles/conditions.	85
4.1.	The block diagram architecture of the tracker, that is realized by the sequential application of filtering masks and an enhanced Viola-Jones based face detection algorithm. Initially, the input frame is filtered by a skin mask and the face of the user is localized within the captured image. The input of the face localization is fed to the background removal building block, while a motion detection is applied, so as to exclude all remaining non-moving (i.e. background) objects. Finally, the application of specific biometric restrictions (i.e. the right/left hand are expected to be found on the right/left side of the head) results to the localization of the hands as the last remaining blobs on the incoming frame.	90
4.2.	Processing of the tracked locations of the user's palm (i.e. blue colour) and head(i.e. red colour) towards the extraction of continuous and smooth motion trajectories. The radius of each coloured circle on the images is proportional to the distance of the current location from the camera.	92
4.3.	A practical example of a raw and uniformly re-sampled trajectory is shown herein. The blue points refer to the initially tracked ones, while the green circles refer to the re-sampled trajectory. It is becomes obvious that the velocity/acceleration/jerk information at any time remains intact.	93
4.4.	Simultaneous data recording from three tracking sources during a <i>Reach and Grasp</i> activity.	94
4.5.	Comparison between vision-based tracker and ground truth (magnetic tracker)	95

4.6. (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.	97
4.7. 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 labels of the axes are similar to the ones of Figure 4.6(b) and are omitted here for reasons of visual simplicity.	98
4.8. Each Activity Surface (AS) is analyzed with respect to the seven highlighted different reference points. Hereby, the sampling of the surface with respect to $R_5 \equiv RP_5$ is illustrated.	101
4.9. 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. $Hx'//O_H\bar{O}_L$). As such, two orientation vectors (θ, ϕ) are generated for each joint of the arm, describing the changes of the movement's orientation in the $3D$ space.	102
4.10. The distinctiveness between the orientation vectors of different users is clearly exhibited by illustrating several recordings of (a) the θ -angle vectors and (c) the vertical ϕ -angle vectors.	103
4.11. The distinctiveness between the (a) Curvature (<i>Phone Conversation</i> Experiment) and (b) Torsion (<i>Reach & Grasp</i> Experiment) vectors of different users is clearly exhibited by the illustrated traits.	104
4.12. Simultaneous data recording from three tracking sources during a <i>Phone Conversation</i> activity.	107
4.13. (a) Raw Angles of the Finger Phalanxes, (b) Notation of Finger Names.	108

4.14. a) Graphical representation of the utilized 5-state, left-to-right and fully connected Hidden Markov Model. b) 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 [4].	118
4.15. Flow Chart diagram of the procedure followed for the extraction of dynamic features from the movement of the arm.	121
4.16. Flow Chart diagram of the procedure followed for the extraction of dynamic features from the movement of the fingers.	122
4.17. Flow Chart diagram of the procedure followed for the selection of the most indicative activity related features.	122
4.18. High Level flow chart diagram of the procedure followed for the enrollment/training and recognition phase of the novel biometric system.	123
5.1. Adjusted skeleton model based on: a) hierarchical filtering, b) OpenNI algorithms	125
5.2. Anthropometric Graphs’ Comparison.	127
5.3. a) Orthogonal Cluster - b) Hexagonal Cluster - c) Gaussian Cluster	131
6.1. Overview of the proposed system	138
6.2. Relative Entropy Values from features extracted by the Camera Tracker for the “Reaching and Grasping” experiment	139
6.3. Relative Entropy Values from features extracted by the Magnetic Tracker for the “Reaching and Grasping” experiment	140
6.4. Relative Entropy Values from features extracted by the CyberGlove for the “Reaching and Grasping” experiment	140
6.5. Relative Entropy Values from features extracted by the Camera Tracker for the “Phone Conversation” experiment	141
6.6. Relative Entropy Values from features extracted by the Magnetic Tracker for the “Phone Conversation” experiment	141

6.7. Relative Entropy Values from features extracted by CyberGlove for the “Phone Conversation” experiment	142
6.8. Phone Conversation - ROC Curves for the fused scores.	144
6.9. Interaction with the Microphone Panel - ROC Curves for the fused scores.	144
6.10. Object detection: a)Top camera view, b)Contour extraction c)Objects’ area detection, d)Tagging of objects.	145
6.11. Extracted motion trajectories from the user’s head (black), shoulder (green), elbow (red) and palm (blue) during a spe- cific movement: (a) The intra-similarities in the motion tra- jectories between different repetitions of the same user are obvious. - (b) The inter-variances in the motion trajectories between the same movement performed by different users are obvious.	148
6.12. The authentication results for each movement when each user is at the normal stress level.	149
6.13. ROC Curves comparing the authentication performance of the proposed SH method and SoA.	152
6.14. The EER value as a function of the utilized features applied in decreasing order of relative entropy for the Camera Tracker.155	
6.15. The EER value as a function of the utilized features, applied in decreasing order of relative entropy for the Magnetic Tracker.155	
6.16. The EER value as a function of the utilized features applied in decreasing order of relative entropy for the CyberGlove.	155
6.17. ROC Curves for the performed activities in Session 5 as recorded by the Camera Tracker.	156
6.18. ROC Curves for the performed activities in Session 5 as recorded by the Magnetic Tracker.	157
6.19. ROC Curves for the performed activities in Session 5 as recorded by the CyberGlove Tracker.	157
6.20. ROC Curves for the performed activities in the ACTIBIO Database, as they were recorded by the Camera Tracker	158
6.21. ROC Curves for the performed activities in the synthetic Database as recorded by the Camera tracking device.	159
6.22. ROC Curves for the performed activities in the synthetic Database as recorded by the Magnetic tracking device.	160

6.23. ROC Curves for the performed activities in the synthetic Database as recorded by the CyberGlove tracking device.	160
6.24. Extracted motion trajectories from (a) User 1 and (b) User 2 during the combined movement of <i>Inserting a Card</i> and <i>Typing a pinword</i>	162
6.25. Architecture of the proposed gait recognition framework.	163
6.26. 1 st row: Great variations between the gallery and the probe even between a client user, when stop detection is disabled. - 2 nd row: Low denaturation rate of the probe GEI, when stop detection is enabled at the probe sample.	164
6.27. Improvements in Gait due to silhouette rotation & stop detection algorithm (29 Subjects) - Left: (<i>RIT</i> classifier) / Right: (<i>KRM</i> classifier).	165
6.28. a) Noise free vS. Noisy ($PSNR = 24.1237dB$) Silhouettes.	166
6.29. CMS Diagram of the final multimodal system - Left: (29-Subjects dataset) / Right: (14-Subjects dataset).	167
7.1. The estimation of the stride length is very sensitive to the shadows that are created on the floor walking level of the user's walking path.	172
7.2. The correlation factor between the error measurements of the utilized soft biometrics is found 0.0755. Thus, these error measurements can be characterized as i.i.d. variables, without loss of generality. The x-axes of all presented diagrams are aligned, so as to correspond to the soft biometric values of the same user (6 soft biometric values per user).	173
7.3. Distribution of the Systematic Error in (a) Height and (b) Stride Measurements and the corresponding fitting curve.	174
7.4. Three alternatives for partitioning the feature space are studied: (a) UOG, (b) UHG and (c) non-linear 2D GG.	175
7.5. Cumulative Matching Scores (<i>CMS</i>) for the <i>GEI-RIT</i> algorithm described in Section B.1 as evaluated on both databases.176	
7.6. Cumulative Matching Scores (<i>CMS</i>) for the Baseline <i>RIT</i> algorithm described in Section B.1 as evaluated on both databases.176	
7.7. Cumulative Matching Scores (<i>CMS</i>) for the Baseline <i>CIT</i> algorithm described in Section B.1 as evaluated on both databases.177	

7.8.	Receiver Operating Characteristics (ROC) for the GEI-RIT algorithm described in Section B.1 as evaluated on both databases.	177
7.9.	Receiver Operating Characteristics (ROC) for the GEI-RIT algorithm described in Section B.1 as evaluated on both databases.	178
7.10.	Receiver Operating Characteristics (ROC) for the GEI-RIT algorithm described in Section B.1 as evaluated on both databases.	178
7.11.	Scores Distribution before and after the application of the proposed framework (Gaussian Clusters) in the HUMABIO RIT-Time experiment.	180
7.12.	Scores Distribution before and after the application of the proposed framework (Gaussian Clusters) in the ACTIBIO RIT-Time experiment.	180
7.13.	The red spots indicate the location of the nose tip as detected by processing the colour related information of the face, while the blue ones point the location estimated by processing the depth information.	181
7.14.	First Row: Face-specific ellipses are drawn on the facial images so as to exclude any noisy, non-facial. points from processing; Second Row: 3D reconstruction of the extracted facial image; Third Row: The extracted isogeodesic stripes are drawn within the boundaries defined by each face-specific ellipse.	182
7.15.	The 3D soft biometric feature space has been partitioned in 13 Gaussian Clusters, according to the spatial proximity of the features.	183
7.16.	Distribution of the induced noise during in (a) Eye-to-Eye distance (b) Nose-to-Eye distance (c) Nose-to-Mouth distance and the corresponding fitting curves.	184
7.17.	(a) Cumulative Matching Scores (CMS) and (b) Receiver Operating Characteristics (ROC) for the 3D Face Recognition algorithm with and without counting in the contribution of the Soft Biometric Traits.	185
7.18.	Scores Distribution with and without counting in the contribution of the Soft Biometric Traits.	185
A.1.	Types of rectangle features.	231

A.2. (a)Face detection, (b)Disparity Information, (c)Skin Colour Filtering, (d)Motion Detection, (e)Detected Hand Positions and (f)Tracking Result	237
B.1. a-c) Gait Energy Images from several users; d) Estimation of Height and Stride length from Silhouette Images	240
B.2. Applying (a) the <i>RIT</i> and (b) <i>CIT</i> transforms on a Gait Energy Image using the Center of Gravity as its origin.	241
B.3. 1 st row: The user walks along the corridor, makes a short stop (2 frames in the middle) and walks on - 2 nd row: Silhouette extraction for the corresponding frames - 3 rd row: Motion History Image (MHI) of two sequential silhouette images. The area of interest is restricted to the lower 25% of the image height. The upper region which covers 10% of the image, is considered to include the head, which is used for the estimation of the walking angle.	244
B.4. The walking angle determination is calculated by the across of the inner product of the walking direction vector and the parallel to image plane vector.	245
B.5. (a)Silhouette Image - (b)Depth Image of the Silhouette - (c)Rotated Silhouette - (d)Rotated Silhouette after Refinement.	246

List of Tables

2.1.	Classification and properties of behavioural biometrics [112]	52
2.2.	Biometric Approaches and Relevant Surveys	53
2.3.	Recognition Performance and Dataset Size of significant works on Facial Dynamics related Biometric Recognition	62
3.1.	Summary of the important characteristics of the utilized datasets	86
4.1.	The Mean Squared Error values when comparing the tracking accuracy of the vision based tracker, having as ground truth the response of the magnetic tracker.	96
4.2.	(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 , respec- tively. $min(AS_{P_1})$, $min(AS_{P_2})$, $min(AS_t)$ 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.	100
4.3.	Authentication Performance, measured as EER scores for dif- ferent amount of utilized states in the training of the HMM based signature.	115
6.1.	Authentication Performance (EER) of Activity Curves	147
6.2.	EER Scores for each single movement (see Figure 3.4).	150
6.3.	Authentication results based on Spherical Harmonics.	152
6.4.	Overall Authentication Errors after final Fusion	160
6.5.	Remaining - Most valuable Features per Tracking Device . . .	161
6.6.	Activity (ACTIBIO Dataset) - Equal Error Rates	166
6.7.	Activity (ACTIBIO Dataset) - Recognition Performance . . .	166
6.8.	Multimodal System (ACTIBIO Dataset) - Equal Error Rates	168
7.1.	Gaussian Mixture fitting the Height Distribution	174

7.2. EER Scores Comparison between the proposed method and [6]	179
7.3. Identification Performance Comparison between the proposed method and [6]	179

1. Introduction

Human recognition has always been a field of primary concern in a wide range of applications, such as access control (e.g. secure infrastructures, computer systems, door security, portable media, safes with biometric locks, etc.), time and attendance management and surveillance. Its primary aim is to achieve personalized human-machine interaction by utilizing these tools that will automatically reveal the identity of the user.

Until recently, the aforementioned scope was indirectly fulfilled via the validation of portable identities or access cards. However, existing commercial methods based on passwords or tokens are nowadays attempted to become independent from the latter, since they can be easily lost, stolen, forgotten or shared. Moreover, modern human recognition systems tend to resemble more natural ways for discriminating among people. They tend to utilize more straightforward approaches to identify someone, not by what (s)he has (e.g. passport) or what (s)he knows (e.g. password), but by what (s)he is. To this extend, **biometrics** seem to be offering a reliable solution to the problem of identity management.

Etymologically, the term “biometrics” stems from the Greek words *bio* (i.e. life/livingness) and *metrics* (i.e. to measure) and refers to the measurements of unique physical or behavioral characteristics of individuals that are of high discrimination capacity and are able to reveal the identity of individuals.

Although there have been reported a few examples of biometric applications since the 14th century [1], the first automated biometric systems became available only over the last few decades, due to three major reasons: a) the development of low-cost and high accuracy sensors, b) the significant advances in the performance and efficiency of modern computer processors and c) to the adaptation of legislation to the arising legal and ethical issues [2]. Many of these automated techniques, however, are based on ideas that were originally conceived hundreds, even thousands years ago.

Initially used by the Chinese merchants, who were stamping palmprints and footprints of children on paper with ink, so as to distinguish them, the anthropologist Alphonse Bertillon became the first person to work in biometrics was in the 1890s, by developing a method of bodily measurement. His system was used by police authorities throughout the world, until it was found out that some people shared the same measurements, and based on these measurements alone, two people could be mistaken for one another. After this, fingerprint [143] and palmprint [172] became not only the first biometric traits to be thoroughly studied by many researchers, but also widely accepted by authorities as a robust recognition tool. However, similarly to the majority of human-machine interaction techniques that are inspired by natural processes, there is a tendency in utilizing more obvious and easily collectable biometric traits. For instance, provided that since the beginning of civilization, humans have used faces to categorize individuals to known (familiar) and unknown (unfamiliar) ones, facial characteristics have always been among the most popular for recognizing people.

Up to date, there can be found a significantly large collection of biometric recognition related literature. A common approach that is frequently met in most biometric related studies include the collection, the analysis and processing of the most discriminative characteristics of the human, so as to deliver robust and accurate recognition in the most efficient manner. Although significant progress has been achieved in the modality and feature selection domain, as well as in their corresponding analysis and processing, there are still many open issues that need to be addressed.

1.1. Motivation

Biometrics are meant to simplify the process of recognition in human-machine interaction systems and thus, to offer advanced security and/or personalized services. However, contrary to most human-machine interaction applications, biometrics pose a twofold challenge. Similarly to any computerized system, the smooth functioning and the high performance are undisputable requirements of biometric systems. More importantly, however, modern biometric systems have to overcome both the hesitation of people to be exposed in unfamiliar and possibly unpleasant recognition procedures, as well as their fear of having their personal data misused, so as to

increase public acceptance.

Although public acceptance is mainly affected by the public awareness with respect to biometric technologies and the corresponding legislation, a major shortcoming of all widely used biometric methods is the obtrusive process for obtaining the biometric features. In particular, the subject has to stop, go through a specific measurement procedure, which depending on the biometric modality¹ can be very obtrusive (i.e. iris scan via laser beam), wait for a period of time and get clearance after authentication is positive.

Targeting at the convenience of the users and the optimal performance in various realistic environments, recent trends in biometrics research deal with the analysis of the dynamic nature of various modalities. Emerging biometric technologies, such as gait recognition, dynamic body motion recognition and technologies, such as automated face/gestures dynamics detection, as well as biometrics measured by sensors either worn by the user or transparently integrated in the infrastructure can potentially allow the non-stop (on-the-move) authentication or even identification, which is unobtrusive and transparent to the subject and become part of an Ambient Intelligence Environment (AmI).

1.1.1. Motivation for Prehension based biometric recognition

The motivation behind using **activity-related biometrics**, and thus, **behavioural biometrics**, for recognition purposes is based on two distinct observations. In particular, as with any other biometric trait, two have conditions to be fulfilled; namely, the biometric characteristic has to exhibit high inter-variance among different people, while these distinctiveness has to be visible by an external observer ².

To this extent, there are many works in the literature that prove a direct coupling between the different outcomes of one's of the motor behaviour and his/her inherent individual differences [317]. In particular, Rosenbaum et al. states in several of his publications the observation that complex

¹Although the term modality is used in the general case to distinguish between biometric traits that stem from the different sensors, in the current thesis, it will refer to different biometric trait without any restriction.

²A more detailed motivation for the use of prehension biometrics, based on psychological findings indications can be found in Section 2.2.2

multijoint movements, such as walking or reaching an object, are planned and executed according to one’s personal behaviour and style.

Furthermore, a number of natural “restrictions”, such as the physiology of the human body, possible impairments or the perceived environment [187] are bound to influence constantly the way that specific movements are executed. Thus, it can be claimed that biometric recognition would be potentially feasible when it may be based on all these dynamic environmentally invariant properties (i.e movement’s distance, direction, starting/ending position, external load, etc.) [188].

On the other hand, there are also many evidences in the psychology related literature providing strong indications for the advanced visual and cognitive perception of humans regarding biological motion. In particular, the behavioral ability of humans to recognize subtle changes in another’s movements [294] has been reported to be the main reason for the recognition of emotional state [302], deceptive intent [303], motor effort [301], gender [304], sexual orientation [305] and even style [323], apart from the identification of the human himself [308].

In this direction, the current thesis proposes a novel method for activity-related biometric authentication in the context of an AmI environment. In particular, the users are authenticated by analyzing the invariant features of their movements, as they are performed by their upper-body, during several everyday activities. The analysis of the movements is based on the processing of the extracted motion trajectories, in order to retrieve unique signatures of dynamic nature that would form reliable biometric traits for authentication.

1.1.2. Motivation for multi-biometric recognition

As it is mentioned in [14], humans have the ability to recognize one another, based on the evidence presented by multiple biometric characteristics (behavioral or physical) in addition to several contextual details associated with the environment. In particular, the recognition process itself may be viewed as the reconciliation of evidence pertaining to these multiple traits and/or multiple modalities.

The assumption that multi-biometric approaches can alleviate several practical problems in biometric recognition (i.e. the performance of a bio-

metrics system be improved by integrating multiple biometrics) has been stated in a series of studies in the literature (e.g. [94] [95]), while in 1999, Hong et al. modeled this assumption mathematically and proved its validity [16].

As such, the same authors claimed that multi-biometric systems can be expected to be more accurate due to the presence of multiple pieces of evidence, while they indicated a list of advantages offered by multi-biometrics compared to uni-biometrics.

- (i) They can offer substantial improvement in the matching accuracy of a biometric system depending upon the information being combined and the fusion methodology adopted.
- (ii) They address the issue of non-universality or insufficient population coverage.
- (iii) It is increasingly difficult for an impostor to spoof multiple biometric traits of a legitimately enrolled individual.
- (iv) They can support continuous/multiple monitoring or tracking of an individual in situations when a single trait is not sufficient.
- (v) They may also be characterized as fault tolerant and noise resilient, that continues to operate even when certain biometric sources become unreliable due to sensor or software malfunction, or deliberate user manipulation.

Similarly, a single behavioural biometric trait (i.e uni-biometric) exhibits several drawbacks in respect with modern requirements in biometric recognition, such as acceptable matching performance. Although psychologically justified and despite the fact that all relevant studies have shown great recognition accuracy when dealing with human perception (see Section 1.1.1), when it comes to machine/computer based systems, it becomes increasingly apparent that efforts are still to be made. In particular, it has been reported non-conventional, but usually unobtrusive, biometrics lack in recognition capacity [17].

Following a similar approach as with physiological biometrics, behavioural multi-biometrics seek to alleviate some of the drawbacks encountered by

uni-biometric by consolidating the evidence presented by multiple biometric sources. This way, the gap with classical approaches will be narrowed, and existing unimodal approaches will be augmented with additional personal information.

In this direction, to the author's view, there is much stronger motivation towards the utilization of soft biometric, that can be derived by the same sensor as the hard biometric, as an extra biometric trait in multi-biometric approaches. The reason for this choice stems from the fact that some of the most common problems in deploying multimodal systems are the computational cost and the complexity of added sensors and the corresponding user interfaces. Moreover, it is also more difficult to control the acquisition environment simultaneously for several traits [14]. Namely, by incorporating sparse and not strictly distinctive characteristics of individuals that can be collected simultaneously with the regular recognition process, such as the colour of the eyes, the skin colour, etc.

Although the overall outlook of a human body may exhibit significant variations over time (e.g. weight), the anthropometric (e.g. limb size, height, etc.) and specific soft characteristics, such as gender, eye-colour, etc., of an adult person remain unchanged throughout his/her life³. Based on this assumption and on the numerous possible combinations of soft and anthropometric characteristics, useful outcomes can be derived either for reducing the search space (i.e. population) or augmenting the recognition results. This way, in order to compensate for the possible lower recognition performance of dynamic biometric information (compared to traditional intrusive biometric technologies), a novel probabilistic framework for the incorporation of supplementary biometric characteristics (i.e. soft and/or anthropometric biometrics) is also proposed herein.

1.1.3. Contribution in activity related biometric recognition

As indicated by its title, the topic addressed by the current thesis is a general study on activity related biometrics. Thus, the main aim of the thesis is the contribution will regard the advancement of activity related biometrics in general. In this context, prehension, an inherently activity related biometric trait, is introduced and studied as a promising mean for human

³Within rational time periods provided the fact of the aging of biometrics [20]

recognition. Despite its general applicability in everyday movements, the evaluation of prehension biometrics is limited in two scenarios (i.e. a phone conversation and reach & grasp activity). However, provided that the aim of the thesis is not only prehension, but activity related biometrics in general, new ideas for improving the performance of the latter are also explored. In this direction, multi-biometric approaches, based mainly on static anthropometric characteristics, are explored. Their efficiency is thus evaluated partially on top of the proposed prehension biometric and partially on a different activity-related system (i.e. gait recognition).

In this context, the biometric traits that will be considered in the current thesis are summed up below:

- Introduction of prehension biometrics traits.
- Introduction of a novel descriptor for prehension biometrics for the
- Evaluation of anthropometric traits of the upperbody
- Multi-biometrics based on soft biometrics traits
- Application of the same multi-biometric framework, based on soft biometrics, in combination with *3D* static facial biometrics.
- Multi-biometric application on purely activity related biometrics (i.e. prehension and gait related traits).

1.2. Problem Formulation

Existing biometric approaches obey to, more or less, a standard recognition procedure. The users requesting recognition (i.e. authentication or identification) are initially interacting either with the interface of the system or come in (close) contact with its sensors, so as to let their data (i.e. biometric traits) to be captured. The latter are then processed, so as the redundancy and all unimportant information is filtered out. Depending on the working mode (i.e. enrollment or recognition), the most significant extracted features are fed either to the training module for the generation of the signature of the users or to a classifier to decide about their identity, respectively.

Moreover, there is a number of 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 way and the frequency with which a given recognition procedure is asked to be performed, as well as the degree of approval of a certain technology by the society are also significant issues to be taken into account.

In addition to the above, the quote that “the ideal biometric system should offer high security combined with excellent user convenience” is nowadays becoming an inviolable guideline for researchers and designers of the future biometric systems [3]. Based on this, novel technologies should be proposed that can seamlessly incorporate the standard biometric requirements in an unobtrusive framework for the users, maintaining in parallel high performance rates, acceptable for biometric recognition systems.

Inspired from the analysis of periodic movements, such as gait, the current thesis tries to extend the concept of biometric recognition based on activity-related traits, by analyzing the way task-specific movements are performed. To address the aforementioned requirements, the current thesis deals with the marker-less tracking of humans, the extraction and evaluation of various static and/or dynamic behavioural characteristics, as well as the analysis and interpretation of the moving patterns of specific joints of the human body during certain activities.

The main challenge of the aforementioned approach is the identification of such movements that take place within AmI environments, the definition of a generic descriptor for them, the extraction of the appropriate features, as well as their classification based on pre-defined signatures. Moreover, provided the reported trade-off between unobtrusiveness and high recognition performance, the incorporation of supplementary biometric data (i.e. soft biometric traits or further static biometric information) or the exploitation of multi-biometrics and multiple authentication ⁴ will be studied in depth.

⁴In the current thesis two types of authentication will be addressed. Namely, **episodic** authentication refers to the cases where only a single authentication procedure is attempted at one instance. The authentication decision is then inferred based on the value of the matching probability that is produced. On the other hand, **multiple** or **continuous** authentication refers to the repetition of episodic authentication attempts, when allowed by the utilized scenario (i.e. when same movements are repeated

The expected outcome of the current thesis will be the thorough study of prehension biometrics regarding their recognition capacity and their potential for incorporation in future biometric systems. Moreover, within the framework of this thesis, a novel approach will be proposed for the analysis of human actions/movements, that can either provide food for thought for or be directly applied to other scientific domains (e.g. activity detection, human tracking, etc.), as it will be shown later. In particular, a novel descriptor will be proposed for the efficient exploitation of the aforementioned traits in problems of biometric recognition in Ambient Intelligence (AmI) environments, based on the utilization of spatiotemporal algorithms. This way, the dynamic nature (i.e. the transitions that are performed in space over time) of the latter will be effectively analyzed, enhancing thus, the current SoA.

Additionally to the above, the recognition capacity of static anthropometric features will be evaluated in short datasets, forming thus, the basis for further experimentation towards the augmentation of the recognition performance of the activity related biometric systems. This way, the combination of such biometric characteristics of static nature (i.e. anthropometric or in general soft biometrics), with hard biometrics (i.e. gait and face), will be attempted under the development of a generic probabilistic framework with high adaptability and integrability.

In general, it can be claimed that this thesis will provide high quality and novel research results regarding both the mathematical tools, as well as prototype use cases, where future biometric systems can have a significant impact.

1.3. Introduction to Prehension Biometrics

One of the advantages of activity-related biometrics is the absence of a particular predefined recognition scenario (i.e. in regular biometric recognition the user's actual work is interrupted by a specific recognition procedure, such as the scanning of their fingerprints). In other words the users are not obliged to undergo a certain (often annoying) recognition process, but

within the same session). In this case, the identity of the user is verified in a constant basis and the overall recognition outcome is derived as a merging of all aforementioned single attempts.

they are allowed to act normally performing their usual everyday activities, while their recognition process takes place transparently.

In addition to the dynamics of the face (i.e. facial dynamics) and body (i.e. gait) of the users, many of the activities performed in everyday life include the physical interaction between humans or between a subject and one or more objects. The latter still remains an unaddressed topic in the field of activity related biometrics. Herein, the focus is on the movement of the arm (i.e. **Reaching**) and on the movement of the fingers (i.e. **Grasping**). Thus, both movements are thoroughly studied during specific actions that include people manipulating objects and for ease of reference they would be described as *Prehension Biometrics* from now on and for the rest of the current thesis. Although the prehension biometric features have not been employed in the field of biometrics yet, and they form a completely novel topic in the literature, significant amount of research has been performed on various aspects of both arm movement for robotics [186] [187] [188] [190] and dynamic palm gestures [182] [183]. As with any other type of biometric modality, in the given case, the goal is to detect and to evaluate a series of stable, invariant, permanent over time and unique activity related biometric characteristics for each human.

Similarly 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 for and grasping an object in the vicinity of a user. Such activities may 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.

As shown before (see Section 1.1), the assumption is that all users have their own characteristic way of reaching, grasping or in general manipulating specific objects, while performing specific activities. In particular, it can be assumed 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.

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 [187].

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) [187]. According to Goodman et al. [188], some features, like the ones discussed hereafter (Section 4.1), 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. [189] 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 (Figure 1.1(a))
- (ii) a slow approach movement, whereby the final stage of the grasping scheme takes place (Figure 1.1(b)).

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 indirectly encoded (e.g. the relative mean or maximum distance values between the head and the hands during the reaching movement) [187]. This assumption can be easily extended to the movement of the fingers, 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.

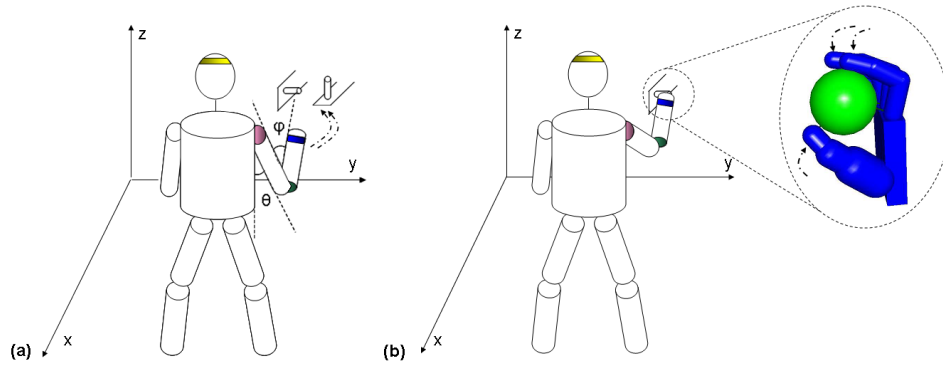


Figure 1.1.: (a) During a *Reaching Movement* shoulder and elbow angles change in a predefined way - (b) During a *Grasping Activity* the fingers' and the palm's angles are moving towards the hand's final posture

1.3.1. Reaching (Arm movement)

Regarding the reaching task, the most important so-called invariances have been analyzed in [188] and in [190]. Specifically, any reaching task of a human arm is characterized by the following common properties.

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 (e.g. to reach for an object by detouring a small obstacle on the path of the hand), 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 [191] and the fact that the users seek the "most convenient" and the least effort demanding way to perform each movement [186], [185] has also been proven.

One of the most important studies has been conducted by Rosenbaum et al. [190], 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 [192].
3. Variability of end positions is generally smaller than variability of movements towards those end positions [194].
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.

1.3.2. Grasping (palm/fingers movement)

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 [216], 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 [217] 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 recognition potential is encoded in the way one grasps an object. This claim can be also supported by the Rosenbaum's model [190], 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 [218]. 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).

1.4. Validity of prehension related features as biometric traits

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 [180] [181], 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, such as gait, have shown 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 [66]

- **Universality:** Each user should possess it.
- **Distinctiveness/Uniqueness:** 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.

In the context of the current study, the *Universality* requirement is satisfied by definition, since all users are expected to be able to perform such movements with their hands. Moreover, there are plenty of models which depict that humans seek the “most convenient” and the less effort demanding way of performing each movement. Specifically, there is the *Flash and Hogan’s Minimum Jerk Model* [185] which indicates that hand paths (i.e. the path drawn by the palm joint during movements) in 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* [186] assumes a hand movement according to the minimization of the torque during the movement. Based on these observations, on Turvey et al.’s [187] and Goodman et al.’s [188] findings, but also on the psychological background mentioned in Section 1.1, it can be claimed that the *Distinctiveness*, the *Reproducibility* and the *Permanence* requirements are fulfilled. This is to be justified by the fact that all these parameters are related to the user’s anthropometric variables, exhibiting significant variance within the population. The *Distinctiveness* requirement, in particular, will be the core study case of the current thesis.

Given the specificities of the current biometric trait, it should be noted that the accurate reproducibility of a prehension movement is highly depending on the environmental context (see Section 3.2.2 for a detailed discussion regarding the context in prehension related movements), especially in the cases that the movement regards an interaction with an environmental object. Specifically, in order to ensure maximum authentication capacity of a prehension movement, the repetition of the movement should take place in an almost identical contextual conditions (i.e. relative position of the user with respect to the interaction object), as when initially registered. For this reason, this restriction has been taken into account in the scenarios of the recorded datasets (see Section 3.2.2).

Similarly, within the acceptable frames regarding the aging of biometric traits, the *Permanence* requirement is preserved, given that the human body remains unchanged over the years, in terms of anthropometric proportions, like the distances between the joints. Of course, like with all biometrics, the issue of aging can only be overcome via the update of the biometric signature over time. However, it should be mentioned that expressions of behaviour are less vulnerable to sudden changes [113] (i.e. a fingertip has a direct and quick effect on the authentication than a change in speech-related facial motions).

Furthermore, in order to ensure the *Permanence* requirement in the experiments conducted in the current thesis (see Section 3.2.2), the generated biometric signatures were used for the verification of incoming prehension traits within rational time frames (i.e. maximum period between two recordings was 6 months), with insignificant influence in the aging of the biometrics. Significantly longer periods between successive recordings should be addressed by retraining of the users' signatures. 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, and thus, it is easily controllable to fulfill them.

1.4.1. Ergonomic factors in Prehension

In order to fulfill the repeatability/reproducibility requirement (Section 1.4) the same or almost the same environmental conditions should remain stable among different sessions. Moreover, the stylistic and behavioural analysis of a person's movements always refers to a relaxed state. Otherwise, unwanted artifacts may appear, which will act as noise to the measurements. In the following, a method based on ergonomic studies is presented, which can handle the "extreme" cases of movements.

Ergonomic Spheres

Due to restrictions set by the structure of the human body, it is easy to understand that there are regions around the human, where the movement of the hands is more convenient than in other regions. These assumptions have been scientifically formulated in [275]. Specifically, it has been proven that the area in front of a seated human can be divided in three different spheres, according to the easiness with which the user can reach an object within certain regions (Figure 1.2). It is suggested that the darkly grey area is the one where the user moves most convenient and is thus, called the "convenient zone". On the contrary, the light grey area indicates the "kinetosphere", whereby the user has to stretch or to bend his body in order to reach something. The white areas on Figure 1.2 are out of reach for the user.

Thus, it can be assumed that the user performs more relaxed movements within the "convenient zone" than in the "kinetosphere". During run-time, it can be claimed that the movements within the "convenient zone" reveal more information about the user's behavioral response, since they are performed under no pressure or with force. On the other hand, the movements within the "kinetosphere" can be considered as forced movements. Thus, the ergonomic zones taken into account are dependent on the distance between the user's torso and the interaction objects.

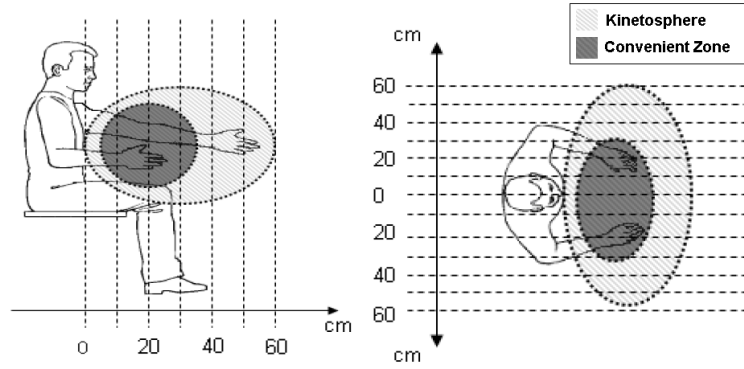


Figure 1.2.: Human convenience zones. The dark shaded area marks the so-called “*Convenient Zone*”, while the light shaded area marks the so-called “*Kinetosphere*”.

1.5. Originality Achievements of the Thesis

The current thesis deals with the development of novel technologies for the unobtrusive and if possible multiple authentication of the users within AmI environments.

The main contribution offered to the State-of-the-Art technologies by the current thesis can be summed up in the following bullets:

- A novel framework for biometric recognition of humans by analyzing the dynamic and static traits of their movement during common daily activities
- The introduction of a novel descriptor, namely the *Activity hyper-Surface*, for the analysis of the movements of the upperbody, by tracking the trajectories of the joints of the subjects, using as reference point the heads. This way, global, local, spatial and temporal based features can be extracted, providing a complete description of the most significant and personalized characteristics of the movement
- The efficient analysis of the extracted biometric features and their selection from specific authentication cases via the combinatory utilization of relative entropies and mutual information techniques in an iterative algorithm, targeting to maximize the authentication performance (e.g. by adjusting Error Rates such as the False Acceptance

Rate (FAR), the False Rejection Rate (FRR) and the Equal Error Rate (EER)).

- The utilization of soft biometrics for augmenting the accuracy of activity-related biometric recognition, via:
 - late fusion with recognition based on anthropometric biometrics,
 - the introduction of systematic error in the measurement of soft biometrics
 - the partitioning of the soft biometric feature space via clustering
- A highly unobtrusive multi-biometric study is performed, utilizing solely activity-related biometric modules (i.e. prehension and gait biometrics) in the framework of an on-the-move⁵ recognition scenario [12] and exhibiting its advantages.
- The evaluation of the aforementioned novelties in multiple datasets and under various scenarios, i.e. episodic and multiple authentication.

Hereafter follows the list of the produced publications that have been achieved during the current study:

Journals

1. A. Drosou, D. Ioannidis, D. Tzovaras, M. Petrou, “Activity Related Authentication using Prehension Biometrics”, Pattern Recognition (Elsevier), accepted with major revisions.
2. A. Drosou, K. Moustakas, D. Tzovaras, M. Petrou, “Systematic Error Analysis for the Enhancement of Biometric Systems using Soft Biometrics”, IEEE Signal Process. Lett., vol.19, no.12, pp.833 - 836, 2012, doi:10.1109/LSP.2012.2221701.
3. D.Tzovaras, A. Drosou, “Continuous Authentication using activity-related Traits”, SPIE Newsroom, 2012, doi:10.1117/2.1201204.004199.

⁵The term “*on-the-move*” is used to emphasize the fact that the recognition procedure is seamlessly integrated within the regular actions expected by the users and thus, the latter are not obliged to interrupt their ongoing action or to be diverged from their actual business.

4. D. Giakoumis, A. Drosou, P. Cipresso, D. Tzovaras, G. Hassapis, A. Gaggioli, G. Riva, "Using Activity-Related Behavioural Features towards more Effective Automatic Stress Detection", *PLoS ONE*, vol.7, no.9, e43571, 2012, doi:10.1371/journal.pone.0043571.
5. A. Drosou, D. Ioannidis, K. Moustakas, D. Tzovaras, "Spatiotemporal analysis of human activities for biometric authentication", *Elsevier Journal of Computer Vision and Image Understanding (CVIU) - Special issue on Semantic Understanding of Human Behaviors in Image Sequences*, vol. 116, no. 3, pp. 411 - 421, 2012, doi:10.1016/j.cviu.2011.08.009.
6. A. Drosou, D. Ioannidis, K. Moustakas, D. Tzovaras, "Unobtrusive Behavioural and Activity Related Multi-modal Biometrics: The ACTIBIO Authentication Concept", in *The scientific World - Special Issue on: Biometrics Applications: Technology, Ethics and Health Hazards*, 2011, doi: 10.1100/tsw.2011.51.
7. A. Drosou, G. Stavropoulos, D. Ioannidis, K. Moustakas, D. Tzovaras, "Unobtrusive multi-modal Biometric Recognition Approach using Activity-related Signatures", in *IET Comput. Vis.*, vol.5, no.6, pp. 367 - 379, 2011, doi:10.1049/iet-cvi.2010.0166.

Conferences

1. A.Drosou, P.Moschonas, D.Tzovaras, "Robust 3D Face Recognition from Low Resolution Images", in *Proc. of International Conference of the Biometrics Special Interest Group (BIOSIG)*, pp.289-296, 2013.
2. A.Drosou; N.Porfyriou; D.Tzovaras; , "Enhancing 3D face recognition using soft biometrics" in *Proc. of 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pp.1-4, 2012.
3. D. Giakoumis, A. Drosou, P. Cipresso, D. Tzovaras, G. Hassapis, A. Gaggioli, G. Riva, "Real-time Monitoring of Behavioural Parameters Related to Psychological Stress", in *Proc. of 17th Annual CyberPsychology & CyberTherapy Conference*, vol.181, pp.287 - 291, 2012.
4. A.Drosou, D.Ioannidis, K.Moustakas, D.Tzovaras, "Activity related biometric authentication using Spherical Harmonics", in *Proc of IEEE*

Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 25 - 30, 2011.

5. A.Drosou, K.Moustakas, D.Tzovaras, “Event-based unobtrusive authentication using multi-view image sequences”, in Proc. of *ACM Multimedia/Artemis Workshop ARTEMIS10*, pp. 69 - 74, 2010.
6. A.Drosou, D.Ioannidis, K.Moustakas, D.Tzovaras, “Activity Related Biometrics based on motion Trajectories” in Proc. of BIOSIG 2010: *Biometrics and Electronic Signatures*, pp. 127 - 132, 2010.
7. A.Drosou, D.Ioannidis, K.Moustakas, D.Tzovaras, “On the potential of activity related recognition”, in Proc. of *The International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, 2010.

Book Chapters

1. A.Drosou, D.Tzovaras, “Activity and Event Related Biometrics”, *Second Generation Biometrics*, E.Mordini, D. Tzovaras (Eds.), Springer, vol.11, pp. 129 - 148, 2012.
2. A.Drosou, D.Tzovaras, “Case Study - Biometric monitoring of behaviour” *Handbook on Ambient Assisted Living for Healthcare, Well-being and Rehabilitation*, J. C. Augusto, M. Huch, A. Kameas, J. Maitland, P. McCullagh, J. Roberts, A. Sixsmith, R. Wichert (Eds.), IOS Press, vol. 11, pp. 155 - 177, 2012.
3. A.Drosou, D.Ioannidis, G. Stavropoulos, K.Moustakas, D.Tzovaras, “Biometric Keys for the Encryption of Multimodal Signatures” *Recent Application in Biometrics*, Jucheng Yang and Norman Poh (Eds.), InTech, 2011.

1.6. Thesis Outline

The current thesis consists of 7 chapters. The thesis starts with introducing the reader to activity related biometrics. The motivational background for using activity related biometrics is stated, mainly on the basis of a series of psychological studies that not only indicate both the increased human

perception of inferring personal properties from one's movements, but they also prove a direct link between one's behaviour and motor responses.

Thereafter, in Section 1.3 a new concept of behavioural recognition, the Prehension Biometrics, is further analysed and motivated towards its exploitation for biometric reasons, while its validity is explained and justified theoretically.

Next, Chapter 2 starts with a detailed literature survey that introduces the need for biometrics for modern human-machine interaction applications. Thereafter, the evolution of biometric technologies through time is described, while the need for the transition from conventional static biometrics to the new concept of behavioural biometrics is thoroughly explained. In this respect, the pioneering and most important works in this field will be discussed thereafter and the open issues will be highlighted.

Chapter 3 provides a detailed presentation of possible use cases for the biometric approaches that the current thesis will propose. Moreover, the influence of the context in respect with activity related biometrics is thoroughly discussed, while the appropriate specifications and limitations for the selection of the utilized datasets are explicitly defined. The analytical description of the utilized datasets follow, presenting the selected actions/movements that will be used for human recognition.

Following this, Chapter 4 presents a novel approach for biometric feature extraction from everyday prehension-related movements. In particular, the extraction of a series of features is performed herein, followed by the corresponding analysis towards dimensionality reduction, in terms of their discrimination capacity. Moreover, relevant classification methods are selected and customized accordingly, so as to meet the needs of the specific data.

In order to leverage the recognition performance of the aforementioned prehension based biometric approach, a multi-biometric approach that utilizes anthropometric and/or other soft biometrics is studied in Chapter 5. In this respect, the anthropometric biometrics of the upperbody are initially studied (i.e. modelled) as a stand alone modality via the novel concept of anthropometric graphs, that is evaluated in Chapter 6 both separately, as well as fused with the aforementioned prehension biometric modality.

However, since the anthropometric graphs can not describe the general case of soft biometrics, a generic probabilistic framework for augmenting the recognition performance of any biometric system via the utilization of other

soft biometric traits is proposed in Chapter 5. In particular, the theoretical background of the framework is explained in detail in the beginning of Chapter 5, while its superiority over state-of-the-art methods is illustrated via two practical applications (i.e. face- and gait-related biometric system) in Chapter 7. The vision based techniques for measuring the utilized soft biometric are also described herein.

Specifically, Chapter 6 and Chapter 7 mainly deals with evaluation of the biometric methods presented in the previous chapters. Hereby, each approach is exhaustively tested, while the advantages compared to the state-of-the-art approaches are highlighted. Additionally, in Chapter 6 a section is dedicated to the combination of several modalities in a combined multi-modal scenario for “on-the-move” recognition, based on a score-level fusion approaches, while the corresponding improvements are presented and discussed. Among others, some slight improvements over the SoA regarding the gait recognition method, used in the experimentation with multi-biometric systems, are presented in Annex B.

Finally, the novelties and the most significant outcomes of the current thesis are summarized in Chapter 8, along with some interesting suggestions for a future research.

2. Literature Survey

Human identification has always been a field of primary concern in applications such as access control in secure infrastructures. Contrary to old fashioned methods, such as ID cards (“what somebody possesses”) or passwords (“what somebody remembers”), which can be easily lost, stolen, forgotten or shared, biometrics offer a reliable solution to the problem of identity management. By using biometrics, it is possible to confirm or establish an individual’s identity based on who she is, rather than by what she possesses (e.g., an ID card) or what she remembers (e.g., a password). Depending on the application context, biometrics are called to provide answers to either the verification (Is the user Mr. X?) or the identification problem (Who is the user?), by measuring the distinctive characteristics of individuals as a means to reveal their identity.

Thus, biometrics have recently gained significant attention from researchers, while they have been rapidly developed for various commercial applications, ranging from surveillance and access control against potential impostors to the management of voters to ensure no one votes twice [7], [8]. Moreover, the aforementioned traditional personal recognition tools (i.e. passwords and PINs) are not useful at all for negative recognition applications (e.g. employee background checks or terrorists prevention from boarding airplanes via identification). On the contrary, although biometric systems may not yet be extremely accurate to support large-scale identification applications, they are the only choice for negative recognition applications [66].

These systems require reliable personal recognition schemes to either confirm or determine the identity of an individual requesting their services. In this concept, a number of approaches have been described in the past to satisfy the different requirements of each application such as reliability, unobtrusiveness, permanence, etc.

In general, biometric traits can be divided in two main categories. Namely, **soft biometrics**, which do not predict a deterministic identity, but only

output certain human characteristics, can be divided into **continuous** (e.g. height, weight, stride length, anthropometrics, etc.) and **discrete** properties (i.e. gender, race/ethnicity etc.) [106]. On the other hand, **hard biometrics** include both the common **physical biometrics**, like finger- [143], hand- or palmprint [172] and retina, iris [108], or facial characteristics [109], and the **behavioural** ones, which describe activity-related patterns of the user, such as signature [110], speech [168], keystroke pattern [111], gait [176], etc.

Provided that the current thesis will tackle and explore significant open issues regarding all most of the aforementioned types of biometric characteristics, it is worth describing some of the latest and most significant works in this area.

2.1. Anthropometric characteristics & Soft Biometrics

Soft biometrics, whose automated recognition is a growing area of research, are characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals [222] [66]. However, they can straightforwardly provide useful additional information towards user identification in large datasets, by verifying hypotheses or by reducing the search space in typical biometric systems [114], or in augmenting authentication processes [6]. In the same context, Cordea et al. in [277] and Jahanbin et al. in [276] proposed two alternative schemes for augmenting facial recognition via the extraction and incorporation of facial anthropometric characteristics.

The increasing importance and reliability of soft biometric traits has already been proven by Dantcheva et al. [115]. Their importance in biometric recognition systems becomes even more evident when considering that automatic soft biometric based inference outperforms human observation, since cognitive biases sometimes make it difficult to come to accurate decisions regarding the face of a person [116] (e.g. people are generally better at recognizing and characterizing those of their own race and approximate age).

Among a series of potential carriers of soft biometric characteristics [117], such as palm geometry (i.e. ratio between finger lengths), facial characteris-

tics (i.e. age [126] [127], gender [128] [129] [130] or race/ethnicity [131] [135] [137], geometry of the chin, lips, nose, eyebrows, creases, lines, sagging, the loss of muscle tone [129], wrinkles or other face components), gait-related recordings are among the most popular. In this respect, efficient algorithms have been proposed for the accurate extraction of gait-related soft characteristics like weight, height and stride length [138] [139].

2.2. Hard Biometrics

As mentioned above, hard biometrics can be divided in two broad categories, namely the *Physical* and the *Behavioural* (i.e. activity-related) ones.

2.2.1. Physical Biometrics

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 [105], which is widely used in law enforcement for identifying criminals [143]. Further, static biometrics include DNA, facial characteristics [168], iris [169] and/or retina [170], and hand geometry [171] or palm print [172] recognition.

Despite their high accuracy, physical biometric traits exhibit a couple of shortcomings that can be proven significant per case. Firstly, the obtrusive process it is required of obtaining the biometric signature. In particular, 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. Moreover, 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 [240]. Last but not least, dealing with very specific features of the human body (e.g. fingerprint, iris, etc.) in detail, physical biometric systems are hard to implement, since they require the exploitation of high precision sensors, efficient algorithms for demanding data processing and their transparent integration in a wide range of environments.

Alternatively, emerging biometric technologies, which have recently at-

tracted the attention of researchers, resemble more natural ways of recognizing people. Similarly to the ways or techniques humans utilize to recognize each other (dynamic face, grimaces, gait, movements, etc.), they detect liveness. Specifically, they utilize behavioural and activity related biometrics and thus they can potentially allow the non-stop (i.e. “on-the-move”) authentication or even identification in an unobtrusive and transparent manner regarding the subject and become part of an Ambient Intelligence (AmI) environment.

Although physiological biometrics have enjoyed more attention than behavioral biometrics and have consequently become more integrated into commercial products, behavioral biometrics exhibit several qualities that make them attractive for security applications. For instance, whereas an adversary can passively extract physiological biometrics (i.e. by lifting a fingerprint from a keyboard), behavioral biometrics do not lend themselves as easily to surreptitious capture as they require a user to consciously perform an action (i.e. speaking a specific phrase) [166]. Thus, it becomes evident that while physiological biometrics cannot change, behavioral biometrics naturally change with the action that is performed. This property is useful for security applications such as key generation, where key compromise necessitates the creation of a new key.

In this respect, the need for the transition from the classic biometrics to the new concept of activity related biometrics is stated and the latest advances in the field of behavioural biometrics are discussed.

2.2.2. Behavioural Biometrics

Recent technologies in biometric recognition resemble more natural ways of recognizing people. Similarly 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 (e.g. fingerprint, iris, etc.). In this respect, behavioural biometrics are related to specific actions and the way that each person executes them.

However, an important question still to be answered is the extent to which behavioural (i.e. activity related) biometrics can be considered as

valuable informational traits with high recognition capacity. To this respect, significant psychological studies are presented below, so as to form a solid theoretical basis to support the use of movements for human identification.

Psychological background of behavioural recognition

Following the guideline that Artificial Intelligence in Human Machine Interaction (HCI) is built on the principles of more human-centered designs, (i.e. made for humans and based on naturally occurring human interactive behaviour) [43], the main motivation for investigating activity-related and thus, behavioural biometrics stems from both the fields of human perception and human psychology.

In order to obtain a solid base and a strong motivation for investigating new types of behavioural biometrics, the human nature should have been explored in terms of distinctive (behavioural) properties that can differentiate a personality from the crowd.

As inherently social beings, humans have developed the ability to perceive and interpret the actions of others as the fundamental prerequisite for successful social interaction. In this respect, it has been observed that they show remarkable ability in recognizing movements of a biological origin, even in complex visual scenes [295, 318]. In particular, psychophysical research has demonstrated that the human visual system is finely tuned to the social cues available in human movement.

In an initial approach, movements highlighted by a few point-lights [297] allowed the researchers to investigate the perception of biological motion without the influence of biological form. Later work confirmed that a human figure can be perceived very easily from such moving displays and more recently [298] it has been shown that point light displays are sufficient for the discrimination of different types of motion such as jumping and dancing. Furthermore, observers have been still shown capable of identifying a point-light one's emotional state, deceptive intent, motor effort, vulnerability, gender [301–304], respectively, or even their sexual orientation from brief and degraded displays of their actions [305]. Moreover, a study regarding the features used by people towards the recognition of style in movements has been delivered in [323].

The aforementioned findings have been explained in the literature on a

neural basis. Starting with single-cell recording results from monkeys [318] and subsequent brain imaging experiments [319] studies have revealed a specific brain area in the superior temporal sulcus (STS) that appears active when human movement is observed. This area in the STS has been implicated to be part of a larger system that is involved in social cognition [320]. In addition to these visual and social regions, research into the production of human movement has found that certain brain areas traditionally thought of as motoric also serve a visual function. In particular the cells termed mirror neurons [321] are activated by both producing a movement, as well as seeing the same goal-directed movements performed. All of these results are beginning to have an impact on research into the perception of human movement since they hold suggestions for what processing constraints are apparent in the human system and what distinct subsystems a general purpose ability to perceive human movement might fall into [322].

Extending these findings, Thompson et al. has noticed the behavioral ability of humans to recognize even more subtle changes in another's movements [294], by triggering different parts of their brains for the tracking of different bodyparts [293] each time. This impressive visual sensitivity to human movement is mainly justified by two theories. Namely, the first refers to an indirect coupling of the perception and reproduction of actions performed by others [306], while the second deals with the fact that humans have a lifetime of experience watching other people moves [307]. On top of these, Rosenbaum et al. proved in [317], that different outcomes of the motor behaviour are in direct correlation with inherent individual differences.

In this direction, a further implication of this improved human perception is the identification of others, based on their moving patterns. In this respect, the first experimentally rigorous study of identity perception from motion was performed by Cutting et al. in [308], where a group of friends managed to accurately identify each of them in recorded videos, just by viewing their moving point-lighted joints. Qualitative good results showed also a similar study of Beardsworth et al. conducted in 1981 [310].

The background assumptions for this ability (e.g. what are the roots of this improved perception ability [311], whether motor experience influences the visual analysis of action [312], whether view-dependent visual experience determines visual sensitivity to human movement [308], etc.) have been

addressed in several psychological studies, towards the recognition of people based on their movements [296]. Similarly [299] has shown that humans can also recognize people by their gait. The study proved that humans have the ability to identify their familiar persons, based on their movement, even under adverse viewing conditions.

Consequently, it can be claimed that psychology presents a strong base for considering action/movement patterns valuable for perceiving individual differences.

Advantages of Behavioural Biometrics

On the whole, behavioural biometrics are less obtrusive and simpler to be integrated in existing systems and scenarios [66] [173], although they are less reliable than physiological biometrics in most cases. This way, integral drawbacks of regular biometrics can be lifted; for instance, inborn physiological characteristics may be mixed with stylish and behavioural ones, so that even twins can be separated.

Provided that the imposed obtrusiveness by non-behavioural biometrics lies in both the utilized sensors (e.g. fingerprint or iris reader) and the recognition procedure which the users are subjected to, Table 2.1 lists a series of different types of behavioural biometric traits [112] that have been presented in the literature up to date and that can be potentially used for recognition reasons, along with the environmental objects/sensors required for the capturing of activity-related traits. Hereby, it becomes evident that no special procedure has to be followed by the user requesting recognition. On the contrary, the subject is let free to act as he would normally do in his everyday activities. Moreover, the reader can also notice that no special hardware equipment (if any) is required in most cases, other than commonly used environmental objects (e.g. Phone, Computer, Credit Card, Pen, etc.), while the recording needs can be easily covered in most cases by a single low cost camera.

Despite the large number of reported behavioural biometrics (see Table 2.1), only a few of them have shown adequately high recognition results, for both person verification and reliable large scale person identification, i.e. signature/handwriting and speech. Yet, some other types have exhibit increased recognition capacity and high potential, exploiting in parallel a

Table 2.1.: Classification and properties of behavioural biometrics [112]

Group	Classification of the various types of behavioural biometrics	Direct HCI	Indirect HCI	Motor Skill	Purely Behavioural	Recognition Type	Required Hardware
Generic Comp. Usage	Audit logs [270] [258] Network traffic [257] Storage Activity [324]			• • •			Computer Computer Computer
Personalized Computer Usage	e-mail behaviour [282] Calling Behaviour [273] Call-stack [261] [264] System Calls [264] Command line Lexicon [274] [278]		•	• •		• •	Computer Phone Computer Computer Computer
Habitual Style	Credit Card use [267] Registry Access [290] GUI Interaction [262]			• •		•	Credit Card Computer Computer
Body Dynamics	Blinking [272] Dynamic Facial Features [113] [177] Lip Movement [288] Gait/stride [176] Handgrip [255] Haptic [285]	•			• • • • • •		Camera Camera Camera Camera Gun Sensors Haptic
Cognit. Behav.	Game Strategy [254] [284] Programming Style [263] Car driving style [266] [265]		• •			• • •	Computer Computer Car Sensors
HCI I/O	Keystroke dynamics [286] Mouse dynamics [268] [269] Tapping [260]	• • •			• • •		Keyboard Mouse Sensor
Writing Style	Biometric sketch [271] Painting Style [289] Text Autorship [313] Signature/Handwriting [325] Voice/Speech/Singing [291] [292]	•			• •	• • •	Mouse Scanner Computer Stylus Microphone

combination of cognitive, motor and habitual patterns, namely the group of body dynamics (see literature review below and Section 1.3), i.e. Blinking, Dynamic Facial Features, Lip Movement, Gait, Handgrip/Haptic.

As mentioned before, recent research trends have been moving towards the vision-based methods (i.e. [176]) aiming at decreased intrusiveness, contrary to sensor-based recognition [174] using body dynamics related signals. Additionally, the majority of recent works and efforts on human recognition are mainly focusing on two specific biometric traits, i.e. the extraction and processing of facial dynamics features [113], as well as the well-known gait related features, either in the form of human body shape dynamics [178] or joints tracking analysis [179].

In this respect, an analysis of the existing methodologies in these scientific areas is presented in the following paragraphs of the current Section, implicitly exhibiting the advantages of dynamic behavioural traits over traditional biometrics.

At this point, a list of some indicative survey publications has been included in the following table per biometric approach, targeting to the convenience of the reader.

Table 2.2.: Biometric Approaches and Relevant Surveys

Biometric Approach	Relevant Literature Review
General (e.g. fingerprint, iris, etc.) Biometrics	[66], [7], [173], [134]
Static Face Biometrics	[68], [44], [118], [119], [120]
Behavioural Biometrics	[67], [112]
Facial behavio-metrics	[121], [122], [123], [44]
Gait biometrics	[309], [132], [125]
Soft Biometrics	[133], [13], [117]
Multi-Biometrics	[68], [14], [15], [93]

Person Recognition using Facial Dynamics

The term “*facial dynamics*” refers to the way one moves both his/her head and his/her facial parts (e.g. mouth eyebrows, etc.). In general, facial dynamics belong to the same category of behavioural movements, and are thus, defined by the same generic principles regarding the psychological-motor dependencies and neural and cognitive perception procedures, as the ones mentioned in Section 2.2.2.

However, since the face is the primary identification mean between humans, its expressions (i.e. facial dynamics) have been given special attention by psychologists and neurologists. In particular, [162–165] indicate that both facial structure and behavioral characteristics provide valuable information to face analysis in the human visual system.

Among some significant findings derived from the aforementioned psychological works is that (i) both static and dynamic facial information are useful for face recognition and analysis, (ii) static information has higher recognition capacity, (iii) facial dynamics provide added value to the recognition under degraded viewing conditions (iv) facial dynamics are more difficult to be learned, (v) familiar faces are easily identified in animation sequences, contrary to unfamiliar ones and (vi) facial dynamics is better for gender identification.

For decades human face recognition has been an active topic in the field of object recognition. Recently, the potential of recognizing a human based on his/her behavioural information from face videos was examined [113, 150, 154, 156–158]

Most of algorithms have been proposed to deal with individual images, also called image-based recognition, where both the training and test set consist of individual face images. However, with existing approaches, the performance of face recognition is affected by different kinds of variations: for example, expression, illumination and pose changes. Thus, researchers have started to look at video-based recognition, in which both training and test sets are video sequences containing the face.

Person recognition using videos has some advantages over image-based recognition. Firstly, the temporal information of faces can be exploited to facilitate the recognition task (e.g. the user’s specific dynamic characteristics like the motion of the head, the evolution of the pose or the mimic of the face, etc.) and secondly, more effective representations (i.e. 3D face models and super resolution images) can be obtained from the video sequence and used to improve the performance of the systems. Last but not least, video-based recognition allows learning or updating the subject model over time. In this respect, aiming to exploit the temporal information or the human behaviour embedded in video sequences, the following categorization of the recent attempts can be considered:

1. **Holistic Methods** This family of techniques analyze the head as a whole, by extracting the head displacements or the pose evolution. In 2002, Li and Chellappa [167] were the first to develop a generic approach to simultaneous object tracking and verification in video data, using posterior probability density estimation through sequential Monte Carlo methods. Huang and Trivedi in [153] describe a multi-camera system for intelligent rooms, combining PCA based subspace feature analysis with Hidden Markov Models (HMM). Three classification schemes are applied to the videos. In the first one each frame is classified and a majority rule is applied to the entire sequence. In the second scheme a Discrete HMM is created by using several training sequences for each person using Baum-Welch estimation, and then a test sequence is classified by the forward method.

Following the same principle, Zhou et al. [151] propose a similar probabilistic framework, where a time series state space model is applied to fuse temporal information. An exemplar-based learning strategy to automatically select video representatives has been also developed therein, serving as mixture centers in an updated likelihood measure. Liu and Chen [155] propose a recognition system based on adaptive Hidden Markov Models (HMMs). They first compute low-dimensional feature vectors from the individual video frames by applying a Principal Component Analysis (PCA); next they model the statistics of the sequences and the temporal dynamics using a HMM for each subject. Then, in the classification part, identification is achieved by using the likelihood scores provided by the HMMs, which are then automatically adapted with the test video sequence, towards better modelling over time.

Aggarwal et al. have modelled the moving face in [150] as a linear dynamical system using an autoregressive and moving average (ARMA) model. The system starts by a pre-processing step which crops the face to a fixed size, then the nose tip is tracked by a KL tracker and pose is estimated by a rough edge based technique. Recently, Lee et al. [154] developed a unified framework for tracking and recognition based on the concept of appearance manifold. In this approach, the tracking and recognition components are tightly coupled: they share the same

appearance model, a low dimensional piecewise linear approximation of the appearance manifold. The connectivity among the subspaces is represented by a transition matrix, which is directly learned from training sequences by observing the actual transitions between pose states. Finally, recognition is achieved through Bayesian inference and a maximum likelihood estimate.

In [156] and [157] a new recognition system based on head motion is proposed, exploiting the temporal information and the human behaviour embedded in video sequences. Head motion is firstly analysed by retrieving the displacements of the eyes, nose and mouth in each video frame; then, the raw signals are transformed and normalized in order to obtain video independent feature vectors. Finally, in order to extract the behavioural information related to each individual and use it for identification and verification purposes, they train individual Gaussian Mixture Models (GMMs) and achieve classification through a Bayesian classifier. In [113] the idiosyncrasies (i.e. uniqueness and permanence of facial actions) of facial motions for person identification are investigated whether these can be used as a behavioral biometric via an efficient pattern recognition algorithm based on dynamic time warping (DTW) concluding that emotional expressions (e.g., smile and disgust) are not sufficiently reliable for identity recognition in real-life situations, contrary to speech-related facial movements that show promising potential.

2. Feature-based methods These methods typically exploit the individual facial features, like the eyes, nose, mouth and eyebrows. One of the first attempts to exploit facial motion for identifying people is presented by Chen et al. in [152]. In their work, they propose to use the optical flow extracted from the motion of the face in order to create a feature vector used during the identification. More precisely, they concatenate the optical flows belonging to each frame into a high dimensional feature vector, which is subsequently reduced using a Principal Component Analysis (PCA) followed by a Linear Discriminant Analysis (LDA). Recently, Choi et al. proposed a face recognition based on the adaptive fusion of multiple face features acquired from a sequential video frames. Towards the fusion the weights for each of

the facial features are computed based on fuzzy membership function and an additional quality measurement for facial images [204].

Using a combination of Hidden Markov Models, Tistarelli et al. manage to capture both the face appearance and the face dynamics, forming thus, a dynamical face model [203]. Interestingly, they automatically map the number of states via unsupervised clustering of expressions of faces in video sequences, resembling in a way the neural patterns activated in the perception of moving faces.

Extending their work on Local Binary Patterns (LBP) for combining appearance and motion for dynamic texture analysis, Hadid et al. investigated in [202] the combination of static and facial dynamics via an extended set of volume LBP features and a boosting scheme (EVLBPL operator), that selects only the personal specific information related to identity while discards any information related to facial expression and/or emotions towards dynamic face and gender recognition.

3. Hybrid methods Colmenarez et al. in [149] have proposed a Bayesian framework which combines face recognition and facial expression recognition to improve results; it finds the face model and expression that maximizes the likelihood of the test image. The face is divided into 4 feature region images containing 9 facial features which are detected and tracked automatically. The feature region images are modelled using Gaussian distributions on principal component subspace. Recently in [158], Saeed et al. have augmented the feature vector of a previous system [156] [157] with features extracted from the motion of the mouth. They exploit the rough localization of the mouth provided by the head feature extractor and develop a simple algorithm based on a colour transform that enhances the mouth and then detects the edges using Sobel edge detector with Otsu's thresholding, followed by a series of morphological operations to improve the shape of the detected mouth. Then features such as the area contained in lip, major / minor axis of lip are extracted and arranged in the feature vector and used for recognition.

Conclusively, current work and efforts on human recognition have shown that the human behaviour (e.g. extraction of facial dynamics features) and motion (e.g. human body shape dynamics during

gait), when considering activity-related signals, may be useful for discriminating people. This is the first step in the exploration of such signals and their potential use in real applications.

In the following, Table 2.3 provides an overview of the recognition results and the size of the utilized dataset for some of the most recent works regarding the facial dynamics related human recognition.

As it was seen by the most of the methods which use facial dynamics for face recognition from videos, promising results can be seen. However, there are still some drawbacks in suggesting them as robust biometric systems. First of all, the relative small size of most utilized dataset, indicates the further investigation needed, unless they are targeting at specific application scenarios for authentication, with small size of registered subjects. Moreover, the utilization of mostly global features is a another drawback, since local ones are also of high recognition capacity, when biometric recognition is the question, as indicated in [124]. Finally, the holistic approaches that lack from a feature selection preprocessing, leading this way to the utilization of redundant and noisy information to be co-processed towards person recognition. Last but not least, it should be mentioned that most of the presented methods exhibit great difficulties when dealing with not-aligned faces, making them very sensitive in most cases of practical scenarios and uncontrolled contextual conditions.

In this respect, the aforementioned drawbacks form also the challenges remaining to be solved in the current domain, so as to make facial dynamics related biometrics a robust identifier.

Person Recognition using Gait Biometrics

Gait is a common, periodic human movement and very indicative representative of activity related biometric traits. The field of gait related recognition has become an attention from researchers, due to its property to reveal significant evidence for one's gender [309], behaviour (e.g. emotional state [62], etc.) and even identity [308], as indicated by significant works in psychology of movement and perception.

Among them, most researchers have focused on gait related human recognition, as the most challenging and perspective domain. Most of the recent gait analysis methods can be divided into two categories of complementary

nature [58]; namely the model-based and the feature-based (i.e. model-free) ones. Model-based approaches use the human body structure [23] [24], while model-free methods employ the whole motion pattern of the human body [25].

Human model-based approaches, that represent an action or a gait with body segments, joint positions, or pose parameters. In general, model-based approaches create models of the human body from the input gait sequences. Previous work on these approaches has shown that they can guarantee good degrees of view- and scale-invariance. They study static and dynamic body parameters of the human locomotion [26], like stride length, stride speed and cadence [28].

The authors of the latter presented in [27] a gait recognition method that fused static and dynamic body biometric features with the constraint that people were walking parallel to the image plane. Static features were extracted from a contour, while dynamic ones were obtained with a model-based tracking approach. In particular, human body is modeled as 14 rigid parts connected to one another at the joints. The whole model has 48 degrees of freedoms (DOFs). The tracking results, namely joint-angle trajectories signals, are considered as gait dynamics for identification and verification. They also obtain static information of body based on Procrustes shape analysis of the change of moving silhouettes, which can has the potential improve the recognition performance.

Bouchrika et al. presented the effects of covariates, such as footwear, load carriage, clothing, and walking speed, for gait recognition [175], while a year later, the same authors proposed a model-based method to extract the joints of the lower limb from lateral walking sequences. The adaptive sequential forward floating selection search algorithm was then employed to select the discriminative features for gait recognition [36]. A noise resistant method has been presented in [29], whereby the proposed gait model contained a pendular motion model and a structural model. The lower limb was modeled as two interconnected pendulums.

A view invariant human action recognition with inputs of projected 2D joint positions was suggested in [35], similarly with [30], whereby an invariant analysis regarding human actions on the use of anthropometry towards constraints on matching was presented. Thereby, a point-light display like representation was used, where a pose was presented through a set of 3D

points. The accurate matching between similar actions performed at different rates was ensured by nonlinear time warping. Further, a realtime action segmentation and recognition algorithm with given 3D joint positions was developed in [31] based on a multi-class AdaBoost algorithm and an HMM classifiers so to improve the overall accuracy. However, it was found out that 2D models are more suitable for motions parallel to the image plane, particularly for gait recognition, while 3D models are capable of tracking movement that is more complex. In [24] the previous human model-based pose recovery approaches are thoroughly reviewed.

A common limitation of gradient descent approaches is the use of a single pose or state estimate that is updated at each time step [32], while other approaches use annealed particle filtering [33] or calibrated cameras [34] to track the full body

Contrary to model based approaches, model-free (i.e. feature based) ones seem to be more attractive to researchers, since they do not rely on the assumption of any specific model of the human body for gait analysis. They directly represent human motion using image information, such as a silhouette, an edge, and an optical flow.

Most of works in the current domain have as origin the pioneering idea of Davis and Bobick [197] pioneered the idea of temporal templates for appearance-based action representation. In particular, they used the two well-known spatiotemporal templates, motion energy images (MEIs) and motion history images (MHIs), to represent action sequences.

Usually simple methods are employed, such as temporal correlation, linear time normalization [37], full volumetric correlation on partitioned silhouette frames [38] and Dynamic Time Warping (*DTW*) [39]. For instance, in [38] the extraction of features was performed on whole silhouettes, in [61] an angular transform was applied on silhouette sequences, while in [40] and [45] gait recognition based on Hidden Markov Models (HMM).

Contrary to the fronto-parallel view assumption or other view dependent approaches like [46], some recent approaches deal with non-canonical view gait recognition, or view-invariant recognition of gait sequences, including model-based schemes with self-camera calibration [47]. Similarly, view transformations based on planar silhouette approximation are presented in [48] while tracking of body parts' trajectories are studied in [49] that can be also used to reconstruct the articulated full body motion [50].

More recent approaches in this domain have intensively dealt with the view invariance of the gait sequence. In this context, Hu et al. presented in [51] a novel multiview gait recognition method that combines the enhanced Gabor (EG) representation of the gait energy image and the a variance of discriminant analysis method for dimensionality reduction in the CASIA gait dataset, so as to cope with the variations due to surface, shoe types, clothing, carrying conditions, etc.

Similarly, Kusakunniran et al. proposed an approach using multiple regression-based view transformation model (VTM) to address this challenge of changing view angles during gait, by transforming gait feature from the source viewing angle into the target one [56]. A year later, the same authors presented in [53] a new view-invariant feature based on procrustes mean shape (PMS), via invariant low-rank textures procrustes shape analysis (PSA) towards cross-view gait recognition.

In order to add contextual invariance, two novel gait descriptors are proposed (i.e. the shifted energy image and the gait structural profile, with increased robustness to some classes of structural variations), have been proposed in [57], along with a new method for the simulation of walking conditions. The same authors combined a year later holistic and model-based features for capturing general gait dynamics and more detailed sub-dynamics, respectively, by refining upon the preceding general dynamics [52].

Last but not least, Matovski et al. the conclusion that elapsed time does not affect recognition significantly in the short-medium term was drawn by the experiments carried out in [54]/

Apparently, the main challenges in the current domain include (i) mainly the preservation of view invariance in the most efficient way, in terms of computational cost and amount of utilized sensors, (ii) the context invariance and (iii) the optimal modelling of the input information, towards improved recognition performance. In this respect, the current thesis aims to address challenges (i) and (ii) by applying the algorithms proposed in Annex B on top of existing SoA algorithms, as far as the uni-biometric gait recognition is concerned.

Table 2.3.: Recognition Performance and Dataset Size of significant works on Facial Dynamics related Biometric Recognition

Publication	Dataset Size	Recognition Rate
[150]	MoBo (24 subjects) - Honda/UCSD (20 subjects) - CRIM (20 subjects) subjects	93.4% - 84.9% - 80.0%
[154]	20 subjects	98.8%
[156]	9 subjects	94.7%
[157]	9 subjects	95.8%
[158]	9 subjects	97.0%
[202]	MoBo (24 subjects) - Honda/UCSD (20 subjects) - CRIM (20 subjects)	97.9% - 96.0% - 98.5%
[203]	21 subjects	95.23%
[155]	MoBo (24 subjects) - Honda/UCSD (20 subjects) - CRIM (20 subjects)	92.3% - 84.2% - 85.4%
[113]	94 subjects	~ 96%
[204]	VidTIMIT (43 subjects) - YouTube celebrity (47 subjects) - Honda/UCSD (20 subjects) - CMUMoBo (25 subjects)	~ 97% - ~ 87% - 100% - 98.2%

2.3. Multi-biometric Approaches

The assumption that humans have the ability to recognize each other based on motion patterns has been initially psychologically proven by Cutting et al. in 1977 [308], while it was neurologically justified by Thompson et al. in [294]. Yet, a work presented by Hong et al. a few years earlier than the latter extended this assumption to multiple evidence and suggested that multiple traits could augment the identification performance. Specifically, they also proved the validity of this claim mathematically in [16]. In the same context, Ross et al. claimed that humans have the ability to recognize one another, based on the evidence presented by multiple biometric characteristics (behavioral or physical) in addition to several contextual details associated with the environment. In particular, the recognition process itself can be viewed as the reconciliation of evidence pertaining to these multiple traits and/or multiple modalities.

Thus, following the analysis that was initiated in Section 1.1.2, a summary of the most important works in the field of multi-biometrics will follow in the current section, discussing as well the pros and cons of such approaches over uni-biometric ones.

In general, from an academic perspective, research in multi-biometrics has several different facets: identifying the sources of multiple biometric information; determining the type of information to be fused; designing optimal fusion methodologies; evaluating and comparing different fusion methodologies; and building robust multimodal interfaces that facilitate the efficient acquisition of multi-biometric data.

Uni-biometric systems have to contend with a variety of problems such as noisy data, intra-class variations, restricted degrees of freedom, non-universality, spoof attacks, and unacceptable error rates. On the other hand, a multimodal system can combine any number of independent biometrics and overcome some of the limitations arising when using just one biometric as the verification tool. For instance, it is estimated that approximately 3% of the population does not have legible fingerprints [103] [104], a voice could be altered by a cold and face recognition systems are susceptible to changes in ambient light and the pose of the subject.

According to Jain et al. [66], multimodal biometric systems can be designed to operate in the following five scenarios:

1. Multiple sensors: the information obtained from different sensors for the same biometric are combined (e.g. optical, solid-state, and ultrasound based sensors are available to capture fingerprints).
2. Multiple biometrics: multiple biometric characteristics such as fingerprint and face are combined. In the verification mode, the multiple biometrics are typically used to improve system accuracy, while in the identification mode the matching speed can also be improved with a proper combination scheme.
3. Multiple units of the same biometric (e.g. fingerprints from two or more fingers, images from the two irises of the same person, etc.).
4. Multiple snapshots of the same biometric: more than one instance of the same biometric is used for the enrollment and/or recognition (e.g. multiple impressions of the same finger, multiple images of the face, etc.).
5. Multiple representations and matching algorithms for the same biometric: a verification or an identification system uses such a combination for recognition, or an identification system uses such a combination for indexing.

As stated in [66] scenarios 2 and 3 are considered to provide the larger improvement in recognition accuracy. However, the aforementioned improvement is bound to come with an increase in the inconvenience to the user in providing multiple cues and a longer acquisition time, while scenarios 1, 4 and 5 apart from combining strongly correlated measurements, they also require increased storage and processing resources. Thus, the next chapters will focus on the overview of approaches that either combine multiple biometric characteristics, i.e. different hard biometrics (Section 2.3.1) or multiple traits of the same biometric, like the hard and soft biometrics extracted by the upperbody, or the face of the full body in gait (Section 2.3.2).

On the whole, multi-biometric approaches can significantly improve the recognition performance of a biometric system besides improving population coverage, deterring spoof attacks, and reducing the failure-to-enroll rate. Although the storage requirements, processing time and the computational demands of a multi-biometric system can be significantly higher, the

above mentioned advantages present a compelling case for deploying multi-biometric systems in systems requiring very high accuracies (see proposed application scenarios in the end of Section 3.1).

2.3.1. Combination of Hard Biometrics

Additionally to the aforementioned fact that the combination of more biometric traits of the same identity improves the recognition performance and the reliability of the biometric systems, it also provides reduced discrimination to people, whose biometrics cannot be recorded well (e.g. due to certain disabilities, etc.). For instance, a major issue with biometrics is that there is a group of humans who do not satisfy all requirements for specific biometrics (i.e. 3% of the population does not have legible fingerprints). Furthermore, there is a demand in improving the performance of the uni-biometric systems, in order to broaden their scope and deployment in real scenarios. Moreover, multimodal biometric systems are much more invulnerable to fraudulent technologies, since multiple biometric characteristics are more likely to resist to spoof attacks than a single one [59]. Additionally, multi-biometric systems are difficult to simultaneously be spoofed, while they can efficiently search large databases, by utilizing pruning methods.

According to recent state of the art works, there are four major levels at which, fusion of multimodal biometrics can take place, namely the sensor level, the feature level, the score level and the decision level fusion. However, score level fusion is mainly preferred when dealing with biometric authentication [93].

In this respect, a lot of work has been carried out in the last decade by the scientific community on multimodal biometrics. Following many works in the domain of fixed biometrics solutions that have already been investigated and implemented various multimodal systems [9–11, 18, 19, 21, 22, 64, 65, 117, 220]. Indicatively, an “on-the-move” gait recognition system has been proposed in [102], whereby gait traits have been combined with face recognition in a controlled environment with fixed cameras. Other multimodal approaches have combined face images and speech signals [10], [11], while face and fingerprint have been combined [18].

Of high significance is also the multimodal approach of the research team of S.Nixon, an expert in gait recognition, who presented the so-called multi-

Biometric tunnel, as able to contact-less acquire a variety of biometrics (i.e. face, an ear and gait) via the utilization of eight synchronised cameras [102], targeting at applications related to “on-the-move” recognition in airports and other high throughput environments.

Some more recent works in the domain of multi-biometry that are based on the combination of purely hard biometrics are presented below, so as not only to exhibit the recent advantages in the field but also to point the direction towards which the research trends are moving.

In [99] a negotiating agent computational architecture towards achieving the desired performance of the multimodal system is presented, while the work in [100] proposes the adaptive combination of real multiple biometric data to determine the optimal fusion strategy and the corresponding fusion parameters and thus, to ensure the optimal performance for the desired level of security.

In particular, great achievements have been performed recently by J.Kittler and his group. For instance, a benchmarking study has been carried out in [94] involving face, fingerprint, and iris biometrics towards person authentication of ~ 500 persons. Namely, quality-dependent and cost-sensitive experiments have been performed thereby, assessing how well fusion algorithms can perform under changing quality of raw biometric images principally due to change of devices and how well a fusion algorithm can perform given restricted computation and in the presence of software and hardware failures, respectively.

The same group proposed a kernel based approach, for substituting a missing modality - at the kernel level- by an unbiased one, the so-called a neutral point, achieving good performance compared to simple baseline fusion methods (e.g. sum rule fusion) [96]. Unlike common missing-data substitution methods, calculation of neutral points may be omitted.

Another two works of the same research team regard again quality measures (e.g., image resolution) derived from the raw biometric data. In [95] the performance degradation that can be due cross device matching is compensated by device-specific quality-dependent score normalization, based on several alternatives such as direct score modelling, by modeling via the cluster index of quality measures, etc. Furthermore, a general Bayesian framework for quality-based fusion of multimodal biometrics dynamically combines the outputs of several biometric classifiers [97].

In a similar context, Fernandez et al. proposed a quality-based conditional processing of the multi-biometric traits to be fused, allowing thus, the easy and efficient combination of matching scores from different devices assuming low dependence among modalities [101]. This way, biometrics traits with low quality data are rejected during the fusion.

Yet, Monwar et al. attempted to lift the limitations induced by noise, intraclass variability, and low data quality [98] via a multimodal fusion scheme. Thereby, the results of different classifiers for face, ear, and signature biometrics are consolidated via the rank-level fusion integration method.

Below, the few available works on the multi-biometric fusion of hard and soft biometric traits is presented, while the pros and cons of each method are discussed. Similarly, the decision for choosing one of these two approaches in the current thesis is justified.

2.3.2. Combination of Soft with Hard Biometrics

As mentioned earlier, soft biometric traits do not have the discriminative capacity to function as stand-alone biometric systems. However, since their first introduction as potential identifiers, they have been effectively integrated in a series of applications. In particular, it has been reported that automated gender identification and age estimation has potential applications in a multitude of areas, such as in filtering digital photo albums or in customizing advertisements. Similarly, iris-based soft biometrics can benefit Unique Identification Number (*UID*) programs for large populations (e.g. in India) [116].

The most important property of soft biometrics, however, is their contribution to a much smaller candidate pool and the fact that they allow the overall query to be answered more accurately and faster by minimizing the size of comparisons for both single uni- and multi-biometric identifications. In the context of authentication processes, Moustakas et al. have proposed a Bayesian framework for boosting the authentication rates from behavioral gait traits with gait-related soft biometrics [6], while Marcialis et al. and Jain et al. proposed a combinatorial approach for facial recognition and fingerprint in [220] and [13], respectively. Soft biometrics have been also combined in another work of A.Jain's group; namely [18]. Thereby, a hybrid multi-biometric approach that combines face and fingerprint with

gender, ethnicity, and height as the soft characteristics, towards enhanced person recognition of 263 users under a bayesian framework.

Similarly, a soft biometric (i.e. the color of the iris) is integrated within a multi-biometric system that combines hard biometrics, such as fingerprint and iris texture in a framework where steerable pyramid filters and multi-channel log-Gabor filters are employed for the proper feature extraction [140]. Moreover, Niinuma et al. proposed in [141] a method for fusing face colour and clothing colour with conventional authentication schemes for multiple user authentication.

Although not addressing combinatorial approaches of multi-biometrics, two works performed by Dantcheva et al. are of high interest, in order not only to formulate an opinion of how soft biometric are recently preferred to be exploited, but also to find commonalities with combination with solely hard multi-biometric approaches.

Initially, the reliability of certain biometric traits (i.e. weight and color of clothes) is studied [106], while person identification is attempted based only on a “bag” of facial, body and accessory soft biometric traits, so as to produce a first evaluation of their recognition capacity. Following this work, statistical analysis on the reliability of soft biometrics systems which employ multiple traits for human identification is performed in [107], emphasizing on the setting where identification errors occur mainly due to the fact that two or more subjects may share similar facial and/or bodily characteristics.

As it can easily become apparent from Section 2.3.1, but also from some recent works on soft biometrics (e.g. [107]), the recent trends indicate a direction towards exploitation of quality factors in modern techniques for combining multi-biometric traits. In other words, the challenge of recent trends tend to assess the effectiveness of a certain biometric trait according to the systematic noise (i.e. reduction in quality) that has been induced on it.

In this direction, to the author’s view, there is much stronger motivation towards the utilization of soft biometric, that can derived by the same sensor as the hard biometric, as an extra biometric trait in multibiometric approaches. The reason for this choice stems from the fact that some of the most common problems in deploying multimodal systems are the computational cost and the complexity of added sensors and the corresponding user interfaces. Moreover, it is also more difficult to control the acquisition en-

vironment simultaneously for several traits [14]. Namely, by incorporating sparse and not strictly distinctive characteristics of individuals that can be collected simultaneously with the regular recognition process, such as the colour of the eyes, the skin colour, etc.

Thus, being in alignment with recent multi-biometric trends (i.e. quality and systematic noise of biometric traits as fusion factors), a novel probabilistic framework that will deal with the induced systematic error in multi-biometric recording will be presented in Section 5.2, apart from the exploitation of classic sum rule based approaches at score level (see Section 6.1.3 and Section 6.2.3). This will then be compared in terms of recognition performance with the state of the art multi-biometric approaches based on soft characteristics of [6] and [141] (see Section 7.1.1 and Section 7.1.2. respectively).

2.4. Summary

In the current chapter, a detailed overview of the research field of biometrics has been presented. Once the motivational background for prehension biometrics in the general framework of behavioural biometrics has been justified, based on relevant psychological works, the most significant works in each relevant sub-domain has been included in a state of the art review. Moreover, the most typical and recent approaches towards human recognition have been analyzed, indicating thus, to which direction next generation biometric efforts should be put to.

Additionally, the current chapter is further extended in the field of multi-biometrics as a common approach to increase the accuracy and the performance of recognition. In this respect, the most significant works in the field of multi-biometrics are cited and shortly described, so as to make clear the experimental setup and the contextual settings. A general conclusion of the current analysis is that soft biometrics will be preferred for integration in multi-biometric systems for the reason that they allow their seamless integration in most of the existing biometric system by being extracted from the same sensor as the basic biometric trait and by requiring less computational resources than regular hard biometrics.

Based on the findings of the relevant literature, the aim of the current work is bilateral. Initially, the current thesis aims to propose and to evaluate

through experimental validation the introduction of a novel behavioural biometric trait that is based on prehension related movements of humans. Provided both the facts that there is a direct connection, behaviour wise, between one's movements and his/her identity and that the majority of these movements are performed by the upperbody, prehension movements can potentially lead to human identification.

Furthermore, this pioneering work will offer significant added value to the field of behavioural biometrics. In particular, extending the limited applicability of gait recognition (i.e. only when the user is walking), the concept of activity-related recognition will be significantly enhanced by the introduction of prehension biometrics, allowing thus, e a wide variety of recognition scenarios to be defined per case and per contextual environment. Moreover, the methodology that will be followed can be easily applied to a series of relevant scientific fields, such as activity detection or feature classification.

Apart from these, the current thesis will also have significant impact in the field of multi-biometrics. By evaluating the recognition potential of static anthropometric and their combination with existing physiological and behavioural biometric systems, but also by experimentally evaluating the performance of a purely behavioural multi-biometric system¹, the long standing problem of limited recognition performance of activity related biometrics will be seamlessly addressed.

Summarizing, the following chapters of the current thesis attempt to address the core of the aforementioned topics. In particular, a new concept of behavioural biometrics related with movement of the arm and the palm is proposed, its potential is studied and evaluated, while it is also combined with other existing unobtrusive biometric modalities (i.e. gait) in a multimodal human recognition framework. Moreover, static anthropometric characteristics are exploited and evaluated as for their recognition capacity and performance, so as to deliver increased accuracy in the recognition performance and ad-hoc reduction in the multi-dimensionality of the feature space.

Following the aforementioned challenges of several subdomains in the field of behavioural biometrics, the challenges posed for the rest of this disserta-

¹Some slight improvements over state of the art gait recognition systems will be also proposed.

tion thesis are listed below:

- The exploitation of novel activity related biometrics that are at least comparable, if not superior, with the current state-of-art behavioural biometrics, in terms of recognition performance and applicability.
- The development of a beyond state of the art framework for multi-biometrics that is in line with recent research trends.
- The definition of application scenarios that ensure context independency.
- The delivery of algorithms, methodologies and technologies that could be utilized in other scientific fields, will open new horizons in the research community and will ignite relevant future work in the same field.

3. Activity related application scenarios

The current chapter serves to familiarize the reader with the application scenarios of the biometric approaches that are proposed in the next chapters.

Definition: *Scenario is a sequence of interactions happening under certain conditions (i.e. context), to achieve the primary actor's goal, and having a particular result with respect to that goal. The interactions start from the triggering action and continue until the goal is delivered or abandoned, and the system completes whatever responsibilities it has with respect to the interaction.*

In this respect, a thorough description of the contextual specifications for the designed scenarios will follow, along with the specifications and the analytic description of the utilized datasets.

3.1. Context 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 [314] or by affecting the planning and execution of the prehension movement of the person to be identified [315] [316].

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 $W4$ (where, what, when and who) or even better via the $W5+$ (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 *W4*-context-sensitive manner is virtually unexplored area of research [43]. Thus, without loss of generality, in the current work only the apparent perceptual aspect of the context (*W4*) in which HCI takes place, will be dealt with.

Based on the *W4* 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, etc.)
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 are listed below:

- Restricted areas in military bases
- Sensitive infrastructures where classified data are stored (e.g. personal data, medical data, war plans, etc.)
- Highly secure areas in nuclear power plants (e.g. control rooms, etc.)
- Control rooms of surveillance/security companies
- Administrator rooms in the central servers of companies managing big amounts of data.

3.2. Activity related datasets

Provided the scope of the current thesis, the specifications of the utilized datasets have to be explicitly defined. Namely the following list contains the most important aspects that should be fulfilled by the datasets:

1. Prehension biometrics should be seamlessly incorporated in regular everyday movements, so that the recognition process to be transparent and thus, unobtrusive to the user.
2. Multiple authentication within the same time session should be supported.
3. Apart from prehension, further activity related movements should be recorded as well (e.g. gait).
4. Gait recordings should cover most real case walking scenarios.
5. Anthropometric and soft biometric traits should be able to be extracted.
6. In case of missing data from one dataset, virtual subjects should be able to be generated either via statistic analysis or via merging of different datasets.
7. The size of the datasets should be adequately large, while the recorded subjects should form a representative sample of real population.
8. Recording in different time session should be performed, so as to verify the permanence of the biometric traits in time.
9. It would be beneficial if contextual information could be extracted.

3.2.1. Discussion on the recorded datasets

Prehension biometrics, just like most of the known activity related biometrics, are still at their infancy. Thus, unlike established biometric traits like fingerprint, they are initially meant to be validated in small datasets. Additionally, provided the fact that context plays a significant role, it has been decided that a good experimental application use case of such biometric traits would be the AmI environment of high security areas (e.g. military

bases, nuclear plants, etc.) or companies (e.g. surveillance/security companies, etc.), where only a few people (i.e. small dataset) would be allowed to have access to certain installations.

In this respect, both ACTIBIO datasets (i.e. one for prehension based recognition DB.P.1 and one for gait recognition DB.G.2) that were recorded within the framework of the European funded ACTIBIO project [136] have been the guideline also for the rest of the dataset created within the current thesis.

In particular, both ACTIBIO dataset was created only after a thorough survey that was conducted on ~ 64 professionals from different disciplines, i.e. Company representatives, Security controllers and Control Room operators, regarding the acceptability of vision based prehension/gait biometrics, their usability and the level of realism, the easiness in use and the level of unobtrusiveness, as well as the integrability and the impact they would have in the aforementioned working environments. Thus, both ACTIBIO datasets were recorded simulating a real case working environment, according to their indications, as described below.

Apart from the HUMABIO gait related dataset DB.G.1, which has been recorded by following the same procedure (e.g. a distinct survey was conducted on professionals, etc.), as with the ACTIBIO dataset, all the recordings of the other datasets mentioned below have remained consistent to the same principles, in terms of application scenarios, demographics and utilized movements.

Specifically, regarding the prehension related recordings, the movements that better resembled a full prehension movement were chosen (i.e. phone conversation and interaction with an office panel) out of a wide variety of office related ones. Further, regarding the gait recordings, as many as possible scenarios of real case contextual walking alternatives have been simulated, while the walking tempo was considered constant. Last but not least, regarding the demographics issues of the recorded datasets, attention has been given to the following issues:

- All recorded subjects were healthy, did not report any disabilities and were capable of performing all prehension- and gait- related actions/movements.
- A balanced proportion between male and female participants was at-

tempted to be kept.

- Unavoidably, the age related demographics did not cover a wide range of ages. Namely, most of the recorded participants were within 24 – 40 years old, while there were some exceptions of 5 persons being at their fifth decade of their life.
- Anonymised data policy was followed in all recordings

It should be noted, that despite being relatively small - due to time, money and effort consuming issues - the size of the recorded datasets can be considered adequate for the case of the simulated application scenario. However, in order to cover this issue in prehension related analysis, a virtual dataset has been generated based on statistical analysis of real ones. Moreover, both ACTIBIO and HUMABIO datasets include recordings in different time sessions.

Finally, it should be noted that the case of recognition retrials is covered, as well, in the Proprietary Multiple Reaching Dataset. Last but not least, the face related dataset DB.F.1, was selected not only due to the challenging fact that it contained low resolution colour and depth images, captured by a low cost commercial depth sensor, but also in order to demonstrate the general applicability of the proposed soft biometric framework on physiological biometric systems. Demographics, were more balanced this time, containing an almost uniform distribution of ages between 20-55 years of 25 males and 20 female users.

3.2.2. Datasets description

For the prehension related biometric approach of Chapter 4, three distinct datasets have been used, i.e. two average sized real ones and a large one consisting exclusively of virtual subjects, while for the exhibiting the generic applicability of the probabilistic framework of Chapter 5, two large gait-related datasets and one relatively small but demanding 3D face related dataset have been utilized.

In this context, the prehension related datasets follow below:

1. **ACTIBIO Reaching Dataset (DB.P.1):** This database was captured in an ambient intelligence indoor environment. There are two

available sessions that were captured with a difference of almost six months. The first session consists of 35 subjects, while the second includes 33 subjects, that have also participated in the first session. The collection protocol defines that each person seats at the desk and act naturally, as if he/she is working. In this respect, "NORMAL" activities, such as answering the phone, drinking water from glass, writing with a pen, interacting with a keyboard panel, etc. and "ALARMING" ones, such as raising the hands, etc. were performed repeatedly by the user, according to the ongoing environmental triggering (e.g. phone ringing) and captured by the surrounding sensors. Although five calibrated cameras have been constantly recording the scene, only one zenithal USB cameras and a Bumblebee (Point Grey) stereo camera for office activity recognition.

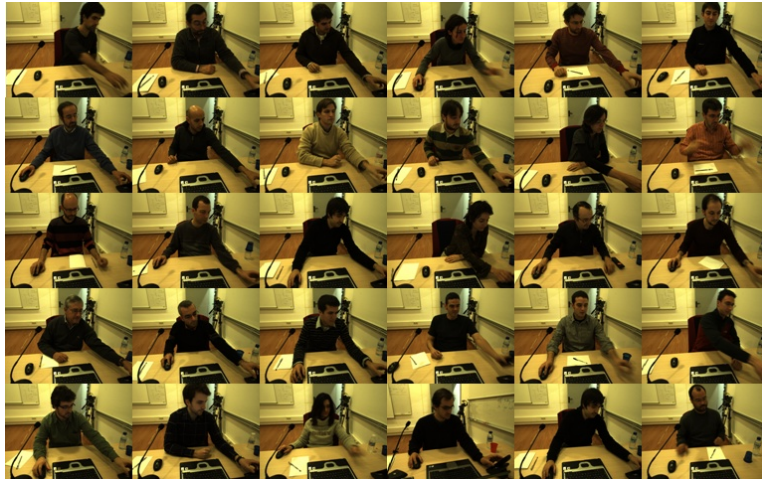


Figure 3.1.: Screenshots of several subjects from the ACTIBIO Reaching Dataset (DB.P.1) performing the "Phone Conversation" activity.

Following the findings of the EU-funded research project *ACTIBIO* [136], it should be noted that among a series of possible office related movements that have been studied, the "Phone Conversation" and the "Interaction with the Microphone Panel" have exhibited the highest recognition potential. Thus, the current work will mainly deal with recognition results from these activities. Namely, the "Phone Conversation" activity (Figure 3.1) includes the picking of the phone (Fig-

ure A.2¹), its narrowing to the ear and its placing back on the base. Similarly, the “Interaction with the Microphone Panel” includes the leaning of the user towards a fixed microphone, the pressing of the ON button and the returning of the user to its initial pose. Some other scenarios include the typing of a short text in the keyboard, an interaction with the mouse, etc. However, such activities do not exhibit high recognition potential, since they do not request significant movement from the user (i.e. during typing there is only a short movement of the fingers) and they do not indicate an exact moving pattern (i.e. moving the mouse can be executed in infinitely many ways). This way, inadequate and non standardized activity-related information is delivered.

A propriety annotator tool was developed, so as to generate complete activity segments and to perform single and multiple searches for all annotated activities sequences, using any possible criterion.

2. **Proprietary Prehension Dataset (DB.P.2):** This dataset was based on two scenarios of the ACTIBIO dataset, including two testing activities performed by 29 subjects. The difference, herein, is that the current dataset contains information captured by the CyberGlove during the grasping activity. In particular, each user has been asked to perform an activity denoted as a raw “Reaching and Grasping” activity (Experiment 1): The user had to lean forward and grasp a lamp standing on the desk. This activity is identical with the “Interaction with the Microphone Panel” activity described in DB.P.1 above, as far as the movement of the arm is concerned. 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 then place 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

¹The figures and equations labelled as “0.X” refer to the Annexes in the end of the current manuscript.

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 later time session (i.e. 1 week after the gallery recordings), 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.

3. **Virtual Prehension Dataset (DB.P.3):** Finally, the proposed method was evaluated in a new 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 Figure 3.2). Given the estimated first and second order statistics among all users’ activity curves ($m^{inter}(t)$, $\sigma^{inter}(t)$) from the Proprietary Prehension Dataset (DB.P.2) above, 100 new “base-features” were created, as indicated below.

$$\mathbf{s}_l^*(\mathbf{t}) = \mathbf{m}_{s,l}^{inter}(\mathbf{t}) + \mathbf{n}_{s,l}(\mathbf{t}) , \quad (3.1)$$

where $n_{s,l}(t)$ was a random number drawn from a normal distribution with 0 mean and standard deviation $\sigma_{s,l}^{inter}(t)$, regarding the l^{th} 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 Figure 3.2) 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. $\mathbf{v}_{v,l}^*(\mathbf{t}) = \mathbf{m}_{v,l}^{inter}(\mathbf{t}) + \mathbf{n}_{v,l}(\mathbf{t})$ and the corresponding intra-parameters.

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 4.1 and 4.2. In Figure 3.2, 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.

In total, 5 virtual activity curves that have been created for each of the 100 virtual subjects. Namely, 4 of them have been used as the gallery set, while the remaining one has been utilized for the generation of the testing signature.

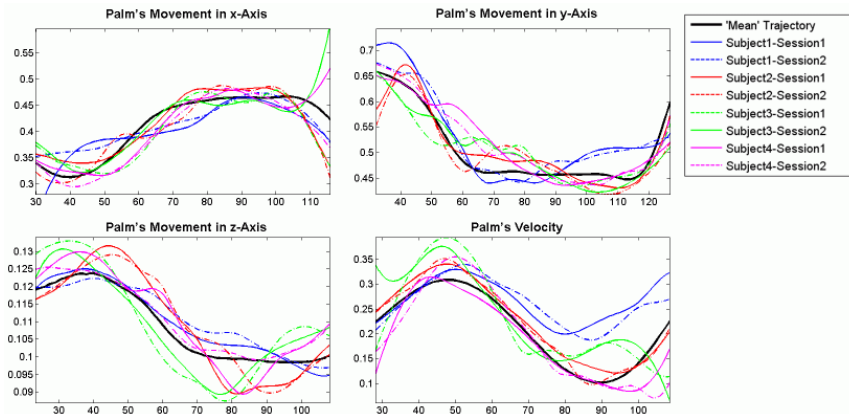


Figure 3.2.: Multiple Repetitions of several features of the Virtual Subjects.

4. **Proprietary Multiple Reaching Dataset (DB.P.4):** This refers to a proprietary dataset that has been created within the Information Technology Institute (ITI) of the Centre for Research and Technology Hellas (CeRTH). The database was captured in an indoor environment and consists of 25 healthy subjects playing a customized version of the Stroop Colour Word test [243].

In particular, during the experiment, a Microsoft Kinect sensor has been constantly recording the movements of each subject, while the pointing gesture was translated in screen coordinates, so that the user was able to control the mouse cursor along the whole range of the screen. In parallel, GSR and ECG sensors were attached on the fingers and chest of each subject, respectively, so as to record the corresponding physiological and emotional traits, as shown in Figure 3.3.

During the whole gameplay time (i.e. 15 minutes), a vast number

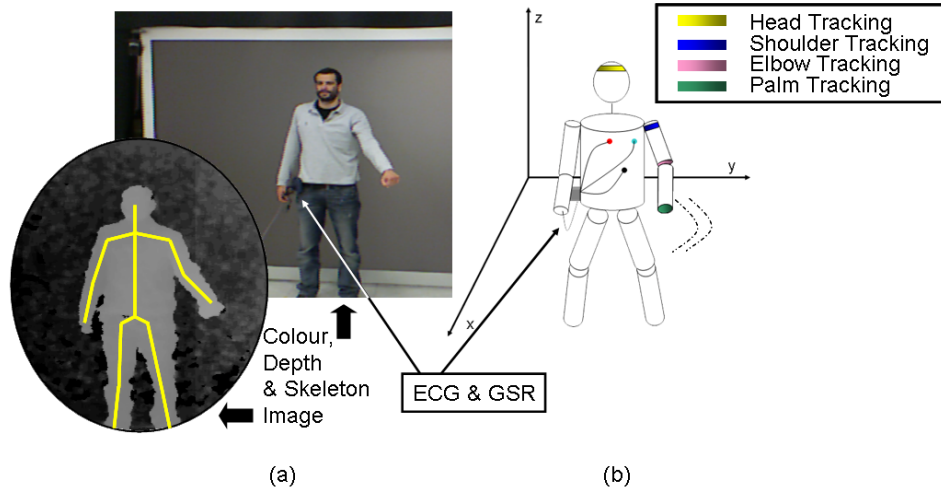


Figure 3.3.: A user playing the Stroop test. The real world coordinates of the his/her joints are automatically provided by the Microsoft SDK [5] (skeleton), while the Galvanic Skin Response (GSR) and the ElectroCardioGram (ECG) signal are concurrently recorded via the attached sensors.

of signatures for each subject was extracted. Regarding the affective state estimation, 1064 intervals of 20sec each were recorded. Half of these recordings can be used for modelling affective response of the average user (i.e. train-dataset) during the game, while the remaining time-intervals have been used for the evaluation of the recognition capabilities of the system (i.e. test-dataset).

The Stroop test, as well as variations of it have been recently used for stress induction purposes [41] [244]. Herein, a modified version of the Stroop colour word test has been used, resembling as accurately as possible a series of movements that are actually performed in commercial games, such as pointing and manipulation gestures, as shown in Figure 3.4. In particular, the current version of the Stroop test utilized five colours (i.e. red, green, blue, yellow and pink), in a question-answer game with 3 Sessions, each consisting of a number of Stroop questions that had to be answered.

At each question, five buttons, each labeled with one of the aforementioned colours, appeared in distinct regions of the screen Figure 3.5.

At the same time a colour name appeared in the middle of the screen, whose font colour varied within the aforementioned colour set. The scope of the game was to pick the button that matched the font colour of the colour name that was displayed in the middle of the screen, just by driving the cursor on it. The question was only then answered, when the gamer managed to press the correct button within a given time frame, which varied among different rounds.

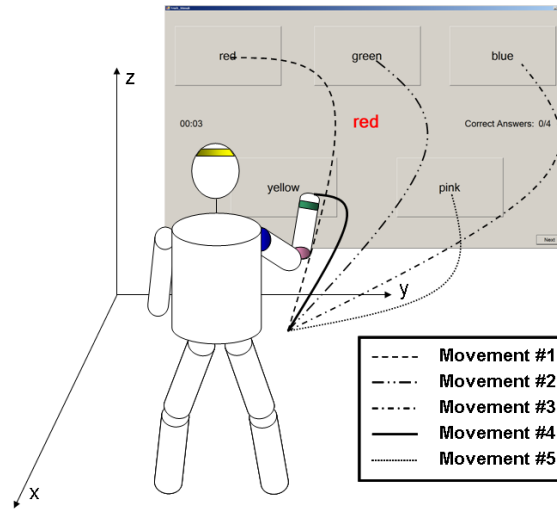


Figure 3.4.: Illustration of the 5 different possible manipulation gestures of the user during the game. The screenshot of the Game Layout refers to the (1st Trial): The buttons are fixed and the reference font colour matches the displayed colour name (waiting time=4 seconds). In the current snapshot, the correct answer is selected by Movement #1. It should be noted that the presented curves do not necessarily represent real movements, in terms of curvature, but they are drawn so for illustrative reasons.

As it can be seen on both Figure 3.4 and Figure 3.5, the buttons were placed in such positions, that the user was forced to actively interact with the game, in order to achieve a high score. In order to make sure that the user would perform distinct and whole movements (i.e. the movement starts when the hand of the user lies relaxed next to his/her body, goes on until the cursor is moved to the indicated button and ends when the hand is back to its initial position), a short break of two seconds was forced between two successive questions.

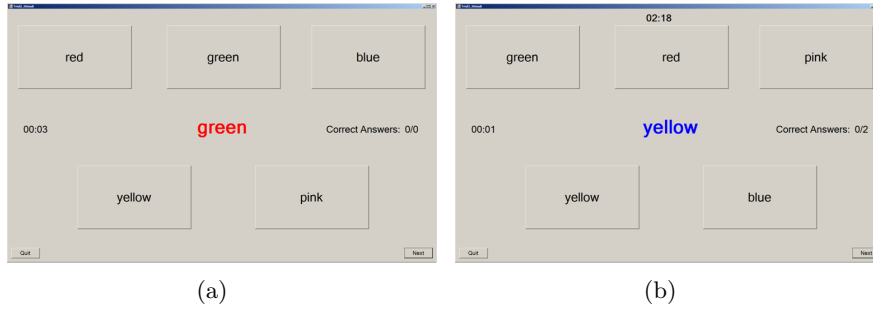


Figure 3.5.: Game instances - (a) Trial 2: The buttons are fixed but the reference font colour does not match the displayed colour name (waiting time=3 seconds) - (b) Trial 3: The buttons are not fixed and the reference font colour does not match the displayed colour name (waiting time=2 seconds).

5. **Proprietary Anthropometric Dataset (DB.A.1):** In order to acquire the appropriate recordings for the OpenNI algorithms, a custom dataset has been recorded by the Microsoft Kinect[®] range sensor. This dataset consists of 35 (testing) and 14 (training) subjects performing the activities indicated in the two selected scenarios of the ACTIBIO Reaching Dataset (DB.P.1) in 3 repetitions. 200 frames from each of the first two repetitions have been used for extracting the user’s anthropometric profile, by which each user was registered to the database. Similarly, the anthropometric profile that has been used for authentication was formed by averaging the results of 200 sequential frames from the third repetition.

In the same respect, the gait and face related datasets, used for the evaluation of the probabilistic framework proposed in Chapter 5, follow below:

1. **HUMABIO Gait Dataset (DB.G.1):** The HUMABIO database, extensively described in [178], was captured in an indoor environment. Briefly, it consists of 75 subjects in the first and 51 in the second capture session. The collection protocol had each person walk multiple times naturally along a predefined straight path, so that the view is approximately fronto-parallel. The differences between different experiments mainly include variations in the clothing of the subjects (e.g. wearing jackets, high heels, carrying a bag, etc.).

2. **ACTIBIO Gait Dataset (DB.G.2)**: The ACTIBIO database is a proprietary activity recognition dataset that also includes two sessions of gait sequences with the same 29 subjects each, captured in two months difference. The subjects were asked to walk several times following patterns of increased complexity (e.g. fronto-parallel, small or progressive deviation in walking direction).

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

3. **Proprietary Activity & Gait Dataset (DB.P.G.3)**: Similarly to the *ACTIBIO Gait Dataset - DB.G.2* dataset, in the current one, each of the *14-subjects* included is walking again in a random path and stops for performing a dual activity. Specifically, each user should type a pin in a panel and then apply a card on a card reader. Both the gait sequences and the rest activities (prehension activities, such as lifting the hands so as to “type on a keyboard panel” and to “insert a ID card” to the appropriate slot) are also included in the current dataset) have been captured by stereo cameras. Hereby, the size and quality of the gait recording were identical to the ACTIBIO dataset *DB.G.2*, while the recorded images were of lower resolution ($320 \times 240 \times 24BPP$) with $15fps$ compared to the ACTIBIO dataset *DB.P.1*.
4. **BIOTAFTOTITA 3D Face Dataset (DB.F.1)**: This dataset was captured in an indoor environment and consists of video sequences (i.e. successive frames) from 54 subjects. It includes, various poses, angles (e.g. $-90^\circ, 45^\circ$), and grimaces (e.g. neutral, smile, scream, etc.) of the *3D* recorded faces, under different lightning conditions, for both

enrollment (“gallery”) and authentication (“probe”) procedures, all annotated and split in different time sessions on the same day (Figure 3.6). All 3D related recordings have been exclusively performed via the Microsoft Kinect Sensor[®].



Figure 3.6.: Screenshots of several subjects from the BIOTAFTOTITA 3D Face Dataset (DB.F.1) during the probe recordings of their faces under several poses/angles/conditions.

Concluding, Table 3.1 summarizing some important information regarding the aforescribed datasets.

Table 3.1.: Summary of the important characteristics of the utilized datasets

Dataset ID	Dataset Name	Size	Properties	Public Availability
DB.P.1	ACTIBIO Reaching Dataset	35(33)subjects	prehension	not public
DB.P.2	Proprietary Prehension Dataset	29 subjects	prehension	to be published
DB.P.3	Virtual Prehension Dataset	100 subjects	prehension	not public
DB.P.4	Proprietary Multiple Reaching Dataset	25 subjects	prehension & affective	to be published
DB.A.1	Proprietary Anthropometric Dataset	35+14 subjects	soft biometric	not public
DB.G.1	HUMABIO Gait Dataset	51 subjects	gait & soft biometric	not public
DB.G.2	ACTIBIO Gait Dataset	29 subjects	gait & soft biometric	not public
DB.P.G.3	Proprietary Activity & Gait Dataset	14 subjects	prehension, gait & soft biometric	not public
DB.F.1	BIOTAF/TOTITA 3D Face Dataset	54 subjects	3D facial recordings	to be published

3.3. Conclusion

In the current chapter, a link between the possible application scenarios of the activity related technologies that will be proposed in the current thesis and the utilized datasets, on which these technologies will be experimentally evaluated. In this respect, the definition of the context has been given, while all requirements for lifting the context dependency of the recorded datasets have been specified. The main scope herein is both the setting up of the framework of the work that will be presented in the next chapters, as well as the familiarization of the reader with some real case scenarios to which the core of the thesis is targeting to have an impact.

After a short discussion regarding the properties of the utilized datasets (i.e. demographics, the size, the selected actions/movements, etc.) and the pros and cons they exhibit, a detailed description of each dataset follows. The chapter ends with an overview presentation of the main properties of the utilized datasets (Table 3.1).

4. User Recognition using Prehension Biometrics

The current chapter deals an extensive study on prehension-based dynamic features and their use for recognition 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 behaviours would produce distinguishable, and thus valuable for biometric verification, activity-related traits.

As mentioned in Section 1.3, following Hoff's principle [189], a prehension activity will be analyzed hereafter as the temporal succession of two distinct movements: (i) A novel approach for analyzing the arm movement is presented herein, based on the generation of an activity related manifold, the so-called *Activity hyper-Surface* (Section 4.1), and (ii) the finger/palm movement will be analyzed complementarily, in terms of the so called spatiotemporal *Activity Curves* formed at the finger/palm joints (Section 4.2). Finally, the authentication capacity of the extracted features is evaluated in terms of their relative entropy and their mutual information (Section 4.3) within a complete framework targeting user verification (Section 4.4).

4.1. Reaching - Feature Extraction using Activity hyper-Surfaces

Provided that the open access Microsoft Kinect Tracker was released, only after the current study was conducted and the databases were captured, a proprietary upper-body tracker has been developed, so as to cover the needs of feature extraction from the collected data. In this context, the current section introduces a novel descriptor for the representation of the

arm movement, based on the trajectories formed by the arm’s end effector (i.e. location of the palm) keeping as reference the initial position of the head. The aforementioned trajectories are unobtrusively detected via the proprietary tracker described below that requires a depth sensor camera (i.e. in our case the stereo Bumblebee camera of Point Gray Inc. [159], while the PrimeSense[®] advanced depth-sensor in combination with the OpenNI [160] library has also been utilized providing improved results, due to the improved depth accuracy it offers.)

4.1.1. Tracking of Reaching Movement

A detailed description of the tracking algorithm can be found in Annex A, while it is briefly described in the following for the convenience of the reader. In the same concept an overview of the building blocks of the proposed tracker is presented in the block diagram of Figure 4.1.

The movements of the users are recorded by a stereo camera and the raw captured images are processed, in order to track their head and hands via the successive application of filtering masks on the captured image. In particular, given the n^{th} frame F^n of the recorded image sequence, a skin-colour mask $S(F^n)$ [195] combined with background extraction $B_{head}(F^n)$ with respect to the position of the head can offer an initial approximation of the possible location of the palms. The head can be efficiently tracked via the Viola Jones based head detection algorithm [201] enhanced by a mean-shift object tracking [212]. Thus, it can be written that the derived filtered image $D(F^n)$ is given as $D(F^n) = S(F^n) \cap B_{head}(F^n)$. In all case, the origin of the axes at each repetition, is the head’s initial location.

Then, by defining as $M(F^n)$ the pixel-wise subtraction of two sequential filtered images $D(F^n)$ and $D(F^{n+1})$

$$M(F^n) \equiv D(F^n) - D(F^{n+1}) \quad (4.1)$$

the remaining blobs on the image I_f^n provide a good estimation of the palms’ positions under the assumption that the right/left hand are expected to be found on the right/left side of the head, respectively.

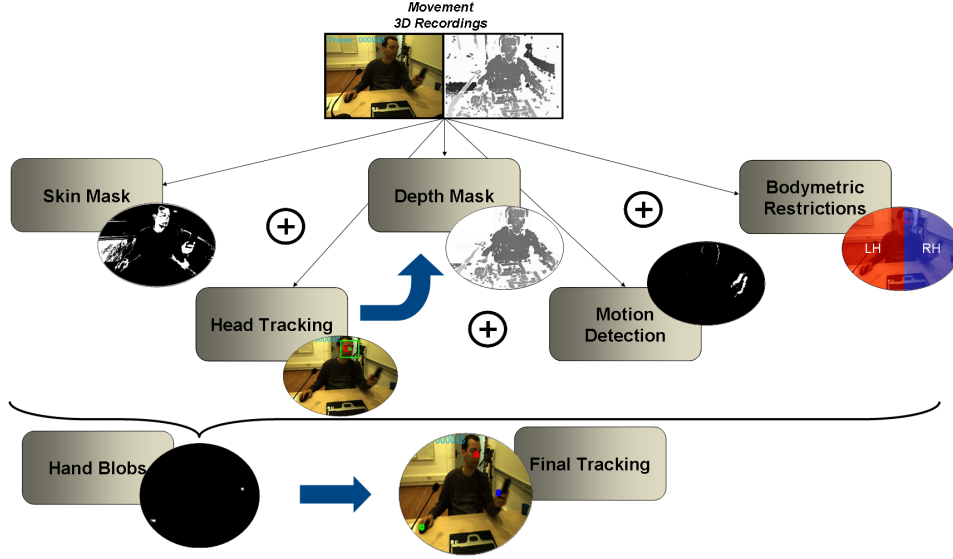


Figure 4.1.: The block diagram architecture of the tracker, that is realized by the sequential application of filtering masks and an enhanced Viola-Jones based face detection algorithm. Initially, the input frame is filtered by a skin mask and the face of the user is localized within the captured image. The input of the face localization is fed to the background removal building block, while a motion detection is applied, so as to exclude all remaining non-moving (i.e. background) objects. Finally, the application of specific bodymetric restrictions (i.e. the right/left hand are expected to be found on the right/left side of the head) results to the localization of the hands as the last remaining blobs on the incoming frame.

$$I_f^n(x, y) = \begin{cases} 2, & \text{if } M(F^n(x, y)) = 1 \\ \max(0, I_f^{n-1}(x, y) - 1), & \text{otherwise} \end{cases} \quad (4.2)$$

As it can be easily perceived from the analysis above, it should be noted that a smooth function of the proposed tracker can be verified only when the following list of assumptions is fulfilled:

1. The users are allowed to turn their head to any direction during the movement. However, they have to look straight to the camera for at least one frame, so that their face is detected by the utilized Viola-

Jones algorithm.

2. If both hands are detected simultaneously the left/right one is assigned to the left/right hand of the user. If only one hand is detected, then it is assigned to the left/right corresponding hand of the user based, on which side of the image it has been detected with respect to the (see Body Restriction related block in Figure 4.1 and Figure A.2(e))
3. The palms of the user (i.e. the skin colour) have to be visible to the camera, in order to be detected by the tracker.
4. If no hands are detected, their last valid detected position is considered as valid.

Trajectory Filtering

The real $3D$ values are acquired taking into account the disparity estimation from the aforementioned $2.5D$ head and palm locations via the intrinsic parameters of the utilized cameras. In order to reduce the effect of noise in the calculation of sensitive high order derivatives, to filter out unwanted artifacts and to make the signature robust, the motion trajectories undergo a multiple pre-processing steps. Moreover, by applying a set of trimming algorithms the length of these trajectories can be improved in terms of homogenization among different users, and the shape can be smoothed in order to represent natural movements (Figure 4.2).

First, short-term fluctuations or perturbations of the exact spot of the hands, due to the shift of the center of gravity of the remaining hand-blob (see Figure A.2(e)), to possible occlusions and/or increased shattering, are discarded via a smoothing window based on the Moving Average Window:

$$s_l(n) = \frac{\mathbf{s}_l^{\text{raw}}(\mathbf{t}_n) + \dots + \mathbf{s}_l^{\text{raw}}(\mathbf{t}_{n+k})}{k} \quad (4.3)$$

whereby $\mathbf{s}^{\text{raw}}_l(t) = (x_l(t), y_l(t), z_l(t))$ with $t \in t_1, \dots, t_N$ represents the $3D$ coordinates of a tracked point of a single trajectory and k is the width of the window utilized.

Next, each s_l undergoes a Kalman filtering process [161]; a very powerful recursive estimator performing low-pass filtering. Since only two points are of interest on each frame, (i.e. the head and the hand), a six-dimensional

(nine-dimensional in case of simultaneous tracking of both hands according of the utilized scenario) Kalman filter is needed. It has to be noted that depending on the utilized scenario/case either the combination of one hand and the head or both hands and the head are taken into account.

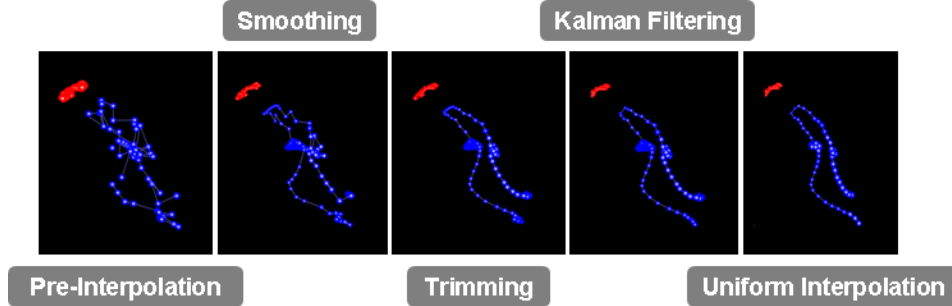


Figure 4.2.: Processing of the tracked locations of the user's palm (i.e. blue colour) and head (i.e. red colour) towards the extraction of continuous and smooth motion trajectories. The radius of each coloured circle on the images is proportional to the distance of the current location from the camera.

Last but not least, each set of trajectories is resized to a pre-defined length by utilizing a modified cubic Hermite spline algorithm, which among others preserves the initial temporal information of the signal.

As shown in Figure 4.3, the horizontal axis stands for time t and has been divided in equally long periods of Δt . The blue spots stand for the raw points that have been actually detected by the tracker while the green circles form the temporally uniform resampled signal. In particular, this is achieved, by applying the following polynomial equation for the estimation of each resampled point $p(t)$

$$p(t^*) = h_{00}(t^*)p_0 + h_{10}(t^*)m_0 + h_{01}(t^*)p_1 + h_{11}(t^*)m_1 \quad (4.4)$$

where the elements h_{00} , h_{10} , h_{01} and h_{11} are given as following

$$\begin{aligned} h_{00}(t) &= (1 + 2t^*)(1 - t^*)^2 \\ h_{10}(t) &= t^*(1 - t^*)^2 \\ h_{01}(t) &= t^{*2}(3 - 2t^*) \\ h_{11}(t) &= t^{*2}(t^* - 1) \end{aligned} \quad (4.5)$$

while m_k is estimated as shown below

$$m_k = \frac{p_{k+1} - p_k}{2(t_{k+1}^* - t_k^*)} + \frac{p_k - p_{k-1}}{2(t_k^* - t_{k-1}^*)} \quad (4.6)$$

It has to be noted that the used time variable t^* refers to the normalized for each set of $\{p_{k-1}, p_k, p_{k+1}\}$ as $t_k^* = \frac{\Delta t}{t_{k+1} - t_k}$.

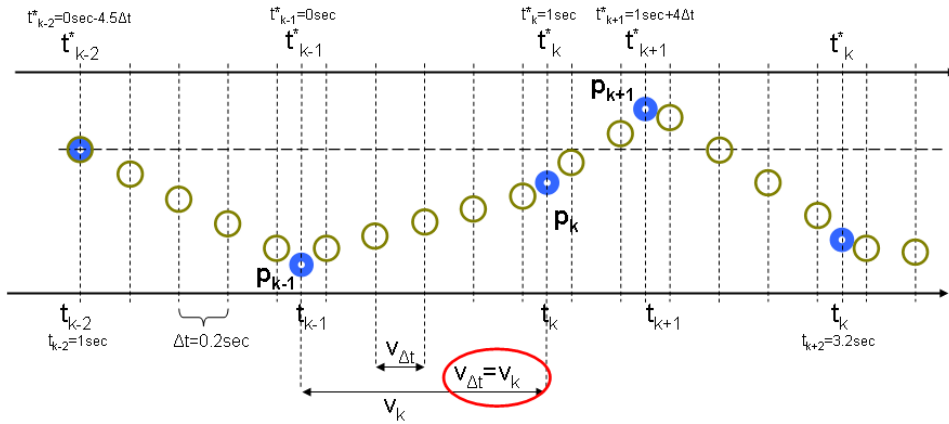


Figure 4.3.: A practical example of a raw and uniformly re-sampled trajectory is shown herein. The blue points refer to the initially tracked ones, while the green circles refer to the re-sampled trajectory. It becomes obvious that the velocity/acceleration/jerk information at any time remains intact.

To process these signals with the most rich content and assess them, in order to quantify the shared information between two distributions, uniform trajectories are created. The uniform interpolation ensures a temporally uniform distribution of the points in the final signature. The re-sampled points of the final trajectory are drawn in such a way that the velocity vector between two actually detected locations is also preserved. This results in optimized and clean trajectories with a slightly different signature data set, without loss of the initial motion information [215]. With this, the output signals have the form of a continuous trajectory description, rather than a sequence of discretely sampled points.

Tracking Evaluation

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 Figure A.2 in Annex B and Figure 4.4, during the execution of the experiments. Simultaneously, the user’s head and hand were tracked by the aforementioned colour/depth sensors.

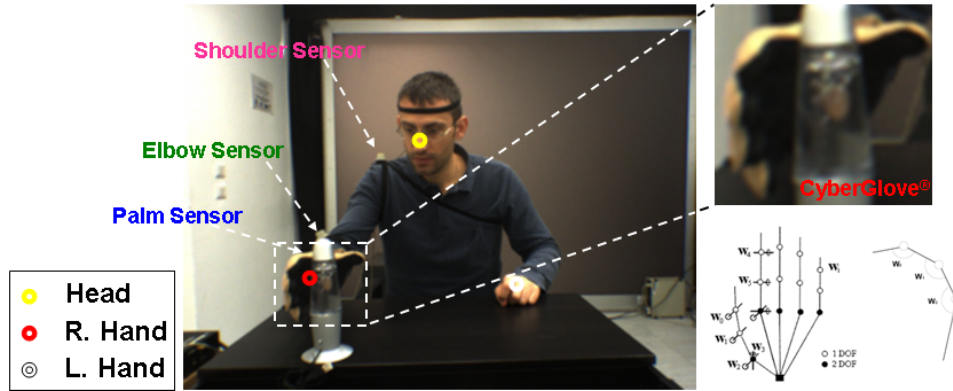


Figure 4.4.: Simultaneous data recording from three tracking sources during a *Reach and Grasp* activity.

The comparison between the derived motion trajectories in Figure 4.5 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 blob 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 also be shown in Section 6.1.

In order to provide a measure for this assumption, the Mean Squared Errors between these two signals (i.e. magnetic tracker and vision-based tracker) for five different persons, as measured by the corresponding sensors are indicatively presented in Table 4.1, below. The low values derived prove the accuracy of the vision based tracker lies within acceptable ranges.

4.1.2. Activity hyper-Surfaces & Feature Extraction

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 in-

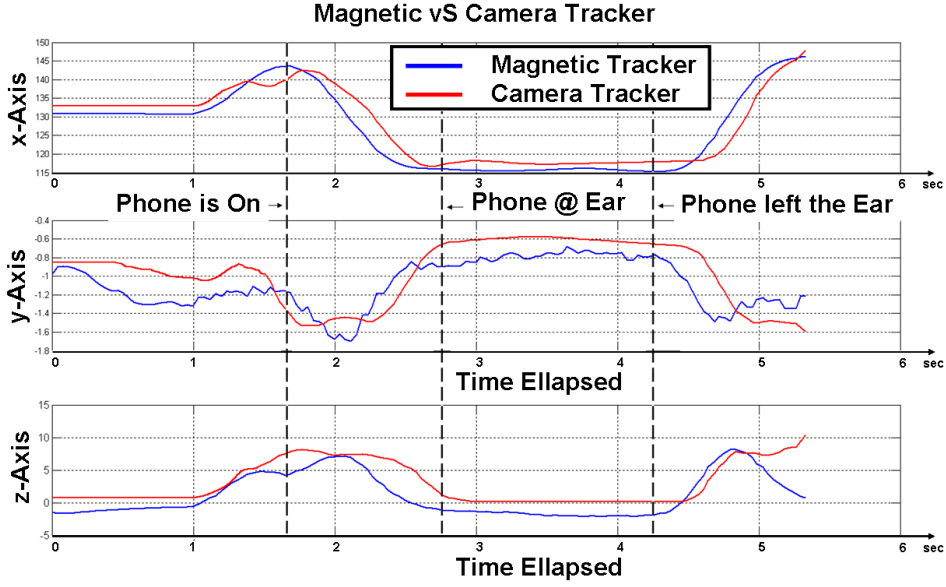


Figure 4.5.: Comparison between vision-based tracker and ground truth (magnetic tracker)

to produce 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 position vector $\mathbf{s}_l(t)$ of the limbs in the 3D space:

$$\mathbf{s}_l(t) = (x_l(t), y_l(t), z_l(t)) \quad (4.7)$$

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 forms the corresponding 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 $0 \leq \mu \leq 1$ (see Figure 4.6(a)).

The area A of such a surface is given by

$$A = \int_{AhS} [(1 - \mu)\mathbf{s}_h(t) + \mu\mathbf{s}_l(t)] ds \quad (4.8)$$

Table 4.1.: The Mean Squared Error values when comparing the tracking accuracy of the vision based tracker, having as ground truth the response of the magnetic tracker.

Subject ID	Mean Squared Error (MSE)			
	X-axis	Y-axis	Z-axis	Overall
Subject 1	0.7932	0.48832	0.5156	1.0646
Subject 2	0.6354	0.61771	0.78512	1.1839
Subject 3	0.8752	0.65879	0.39782	1.1654
Subject 4	0.8348	0.51478	0.52798	1.1138
Subject 5	0.7425	0.43955	0.67421	1.0950

where $ds = dt d\mu$ 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. Without loss of generality, the term AhS will refer only to the head-to-palm for the rest of the thesis. This simplification is justified by the facts that the movement of the elbow’s joint is significantly correlated with the palm’s movement ([191]) and that the shoulder’s and head’s movements exhibit high dependency on each other.

The proposed AhS is a four dimensional ($4D$) manifold that is not only difficult to be perceived visually, but also inappropriate to be analysed by the approach presented in the next paragraphs of the current Section. Thus, following the assumption that during a prehension activity, the movement is mainly concentrated on a $2D$ plane [185], 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 together with the time axis as an extra dimension form a $3D$ space in which the original AhS manifold is represented by a surface, characteristic of the ongoing activity, as illustrated in Figure 4.6(b).

Definition: *The simplified Activity Surface \mathbf{AS} comprises the union of all points that lie on the lines connecting these representative points of the head and hand trajectories, that are closest neighbours with respect the time dimension 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*

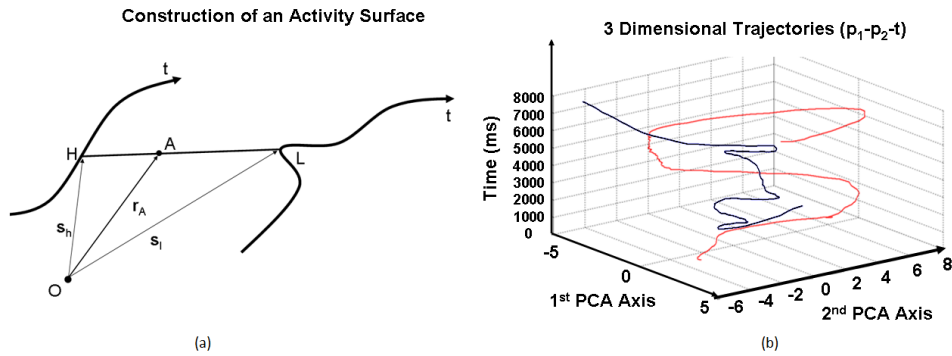


Figure 4.6.: (a) Construction of an Activity Surface. O is the origin of the axes. The two black curves represent the trajectories of vectors $s_h(t)$ and $s_l(t)$ parameterized by time t . Point A is any point of segment HL defined by the position of the head H and the limb L at a specific time, with position vector 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.

the activity's duration (Figure 4.7).

As already mentioned, the utilization of the PCA algorithm on the 3D trajectories has been chosen, so as to simplify the complexity induced by the relative position the user-camera position during the movement. This is a necessary step in order to downgrade the visual complexity of the Activity hyper Surface with negligible information loss, based on Flash et al. work, to form a surface perceivable by the human eye and most important to be able to process it with existing 3D modelling techniques. Although any linear dimensionality reduction algorithm would fit herein, PCA has been preferred due to its general application and its relative low computational complexity.

This way, the view-variance of each trajectory drawn by a single body joint (i.e. head or hand) is lifted. However, since the Activity Surface (**AS**) is a combination of two trajectories, its view in-variance can not be guaranteed at this step. For this reason, the utilization of a view-invariant method (i.e. Spherical Harmonics Analysis is utilized, herein, as it will be described later in the current chapter) for the analysis of the generated surface is needed.

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. 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 Figure 4.7, the reader can notice pairs (vertically) of the Activity Surface of the different users. The intra-similarities and the inter-variances of the shape of the Activity Surfaces between several users are visually illustrated. Inspired by this, an analysis of the shape of the surface is expected to have good potential in providing discriminative features for recognizing between different users.

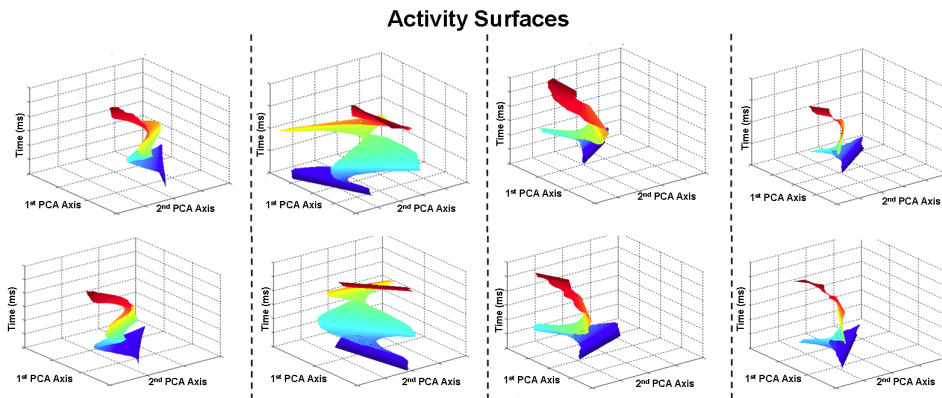


Figure 4.7.: Eight Activity Surfaces exhibiting visually intra-similarity and inter-variances. The surfaces in different columns correspond to different people, all executing the same action. The labels of the axes are similar to the ones of Figure 4.6(b) and are omitted here for reasons of visual simplicity.

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) [213].

At this point it should be noted that the analysis in terms of Spherical Harmonis and not a different tool for 3D shapes, has been preferred due

to the following reason. Namely, due to the fact that the tracked trajectories (i.e. forming the boundaries of the Activity Surface) are already noisy, as explained before, gradient analysis and high order derivatives in general have experimentally shown to amplify the noise. Moreover, despite the fact that high order Spherical Harmonic Coefficients do also encode noisy information, only the lower, and thus, noise-free bands/ranks are kept within the proposed approach. Last but not least, the lower bands/ranks of Spherical Harmonic Coefficient which are practically noise-free contain the higher amounts of energy.

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 Coefficients (SHC) via Spherical Harmonic Analysis (SHA). For their calculation, the Activity 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 the aforementioned reference point R in our coordinate system, so as to minimize the amount of 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 explained in Table 4.2 and shown in Figure 4.8. 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, $f_{R_i}(\bar{\omega}) : \{\bar{\omega} \in \mathbb{R}^3 : f_{R_i}(\bar{\omega}) \in \mathbb{S}^2\}$, whereby $\bar{\omega}$ is a point $\{P_1, P_2, t\}$ of the AS in spherical coordinates.

$$f_{R_i}(\theta, \phi) = \min_{k=1, \dots, K} \{d(R_i, v_k)\} \quad (4.9)$$

whereby $v_k = v_k(R_i, \theta, \phi)$ is the k^{th} 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

Table 4.2.: (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
R₁	CoG_{P_2t}	CoG_{P_1t}	$min(AS_t)$
R₂	CoG_{P_2t}	$min(AS_{P_2})$	$CoG_{P_1P_2}$
R₃	$min(AS_{P_1})$	CoG_{P_1t}	$CoG_{P_1P_2}$
R₄	CoG_{P_2t}	CoG_{P_1t}	$CoG_{P_1P_2}$
R₅	CoG_{P_2t}	CoG_{P_1t}	$max(AS_t)$
R₆	CoG_{P_2t}	$max(AS_{P_2})$	$CoG_{P_1P_2}$
R₇	$max(AS_{P_1})$	CoG_{P_1t}	$CoG_{P_1P_2}$

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 calculated by utilizing the orthonormalized spherical harmonics, multiplying them with the aforementioned function, and integrating the product over the solid angle $d\theta d\phi$.

$$c_l^m = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} \int_0^{2\pi} \int_0^\pi f_{R_i}(\theta, \phi) P_l^m(\cos \theta) e^{jm\phi} d\theta d\phi \quad (4.10)$$

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 l^{th} SHB 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 SHB, the polynomials are orthogonal with each other, unless neither $m = m'$ nor $l = l'$ holds [184].

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

$$c_l^* = \sum_{m=-K_l}^{K_l} |c_l^m| \quad (4.11)$$

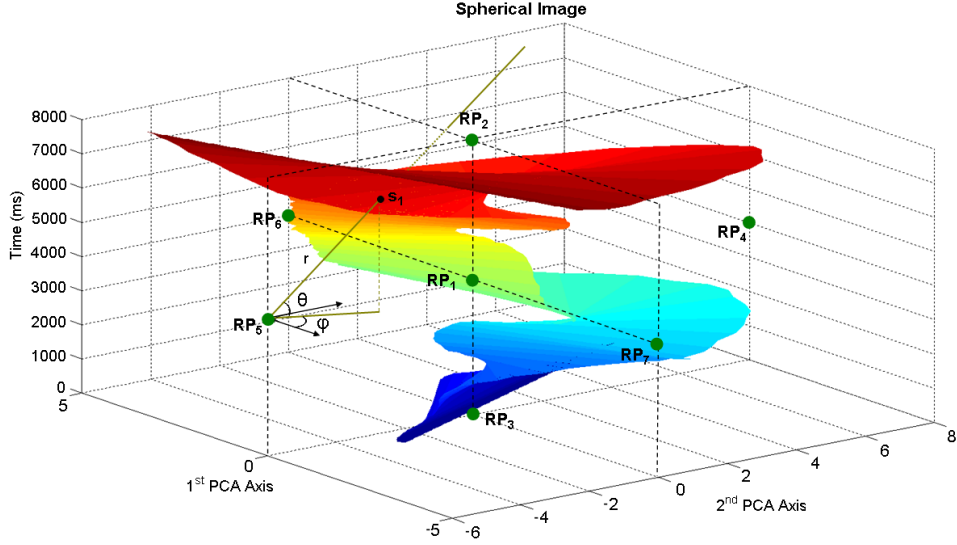


Figure 4.8.: Each Activity Surface (AS) is analyzed with respect to the seven highlighted different reference points. Hereby, the sampling of the surface with respect to $R_5 \equiv RP_5$ is illustrated.

whereby c_l^* denotes the l^{th} 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 V. Schonefeld [214].

Orientation as biometric descriptor

By further studying the previously constructed Activity Surface, two new invariant descriptors can be derived during each movement. Namely, these are the state vectors $\theta_{arm,l}$ and $\phi_{arm,l}$ that can be produced for each joint of the arm that performs the movement. These two state vectors describe the orientation of the joint l with respect to the head of the user (i.e. origin of the axis), for the whole duration of the movement, as shown below

$$\theta(t) = \arctan\left(\frac{y_{head}(t) - y_{limb}(t)}{x_{head}(t) - x_{limb}(t)}\right) \quad (4.12)$$

$$\varphi(t) = \arccos\left(\frac{z_{head}(t) - z_{limb}(t)}{r(t)}\right) \quad (4.13)$$

$r(t) = \sqrt{(x_{head}(t) - x_{limb}(t))^2 + (y_{head}(t) - y_{limb}(t))^2 + (z_{head}(t) - z_{limb}(t))^2}$
and t a certain time instance within the duration of the movement.

The meaning of this transformation is graphically explained in Figure 4.9. Thereby, the straight line (i.e. vector in the 3D space) that connects the points H (i.e. the position of the head at time t) and L (i.e. the position of joint l at time t) has a certain orientation in the 3D space. In particular, it forms the relative angle with which the joint l is positioned in the space that has its origin in the current position of the head. The orientation of the axis (θ_0, ϕ_0) is selected so that they are aligned to the direction, where the limb l is found at the beginning of the movement.

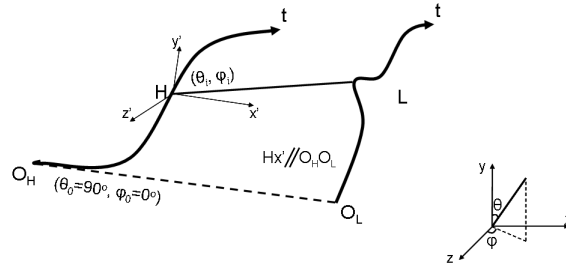


Figure 4.9.: The extraction of the orientation vectors of each joint of the arm is calculated with reference to the user’s head. The origin of the axis is aligned with the initial pose of the user (i.e. $Hx'//O_HO_L$). As such, two orientation vectors (θ, ϕ) are generated for each joint of the arm, describing the changes of the movement’s orientation in the 3D space.

This way, some sample orientation vectors of the horizontal θ -angle vectors and the vertical ϕ -angle vectors are presented in Figure 4.10.

Speed, Acceleration and Jerk

In [191] 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 they 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,

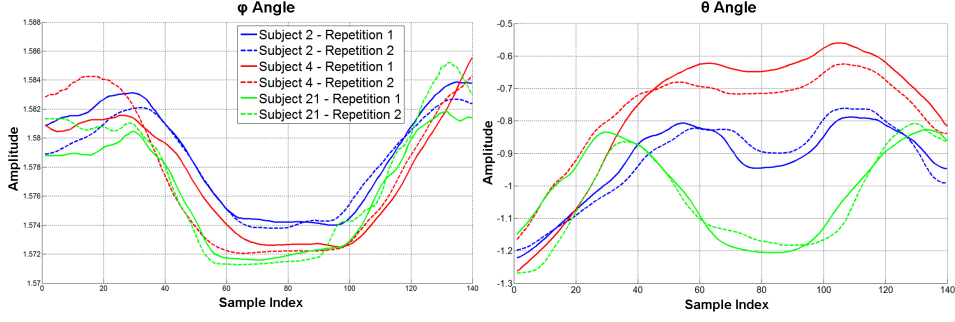


Figure 4.10.: The distinctiveness between the orientation vectors of different users is clearly exhibited by illustrating several recordings of (a) the θ -angle vectors and (c) the vertical ϕ -angle vectors.

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

$$\mathbf{v}_{x,y,z}(t) = \frac{\mathbf{s}_{x,y,z}(t+1) - \mathbf{s}_{x,y,z}(t-1)}{2\Delta t} \quad (4.14)$$

$$\boldsymbol{\alpha}_{x,y,z} = \frac{\mathbf{v}_{x,y,z}(t+1) - \mathbf{v}_{x,y,z}(t-1)}{2\Delta t} \quad (4.15)$$

$$\mathbf{j}_{x,y,z} = \frac{\boldsymbol{\alpha}_{x,y,z}(t+1) - \boldsymbol{\alpha}_{x,y,z}(t-1)}{2\Delta t} \quad (4.16)$$

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

$$\kappa(t) = \frac{\mathbf{s}(t) \times \ddot{\mathbf{s}}(t)}{\|\dot{\mathbf{s}}(t)\|^3}; \quad \tau(t) = \frac{(\mathbf{s}(t) \times \ddot{\mathbf{s}}(t)) \cdot \dddot{\mathbf{s}}(t)}{\|\dot{\mathbf{s}}(t) \times \ddot{\mathbf{s}}(t)\|^2} \quad (4.17)$$

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

$$\kappa_s(t) = \frac{d\kappa(t)}{ds} = \frac{d\kappa(t)}{dt} \frac{dt}{ds} = \frac{d\kappa(t)}{dt} \frac{1}{\|\dot{\mathbf{s}}(t)\|} \quad (4.18)$$

$$\tau_s(t) = \frac{d\tau(t)}{ds} = \frac{d\tau(t)}{dt} \frac{dt}{ds} = \frac{d\tau(t)}{dt} \frac{1}{\|\dot{\mathbf{s}}(t)\|} \quad (4.19)$$

The intra-similarities and the inter-dissimilarity of some extracted curvature and torsion traits from three arbitrary experiments (Section 6.1) are illustrated in Figure 4.11(a) and Figure 4.11(b).

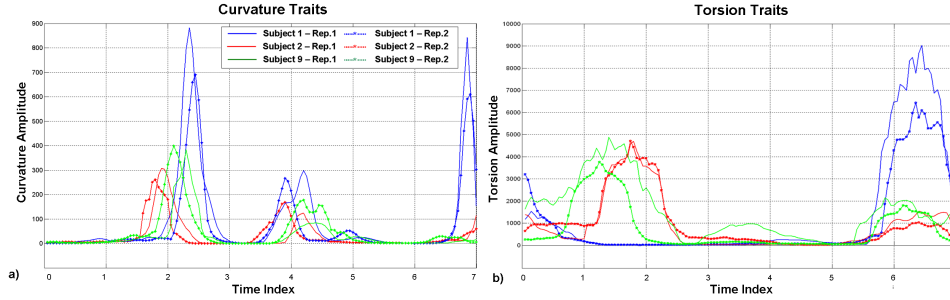


Figure 4.11.: The distinctiveness between the (a) Curvature (*Phone Conversation* Experiment) and (b) Torsion (*Reach & Grasp* Experiment) vectors of different users is clearly exhibited by the illustrated traits.

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 [215].

Dynamic Spatial Cost

According to Rosenbaum et al. [190], 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

$$dsc_p = \sqrt{(X_o - X_c)^2 + (Y_o - Y_c)^2 + (Z_o - Z_c)^2} \quad (4.20)$$

where (X_o, Y_o, Z_o) 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 distance at a given time-instant, as described by the following recursive equation:

$$dsc_p(t) = dsc_p(t-1) + \sqrt{(x_l(t) - x_l(t-1))^2 + (y_l(t) - y_l(t-1))^2 + (z_l(t) - z_l(t-1))^2} \quad (4.21)$$

In this respect the $dsc_p(T)$ 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 equation (4.21) 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 6.1.1). 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.

Activity Curves

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

$$\mathbf{S}(t) = (\mathbf{s}_{head}(t), \mathbf{s}_{shoulder}(t), \mathbf{s}_{elbow}(t), \mathbf{s}_{hand}(t)) \quad (4.22)$$

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 ($\mathbf{V} = \{\mathbf{S}(t), \mathbf{c}_l^*, \mathbf{v}(t), \boldsymbol{\alpha}(t), \boldsymbol{\theta}_{arm,l}, \boldsymbol{\phi}_{arm,l}, \mathbf{c}(t), \mathbf{dsc}(t)\}$).

The aforementioned set \mathbf{V} of biometric feature vectors was constructed empirically in order to describe the person's movement as it contains temporal-related features (i.e. velocity $\mathbf{v}(t)$ and acceleration $\boldsymbol{\alpha}(t)$), spatial-related features (i.e. spherical harmonics \mathbf{c}_l^*) in terms of reconstruction of the surface, view-invariant global features (i.e. orientation of the movement $\boldsymbol{\theta}_{arm,l}$, $\boldsymbol{\phi}_{arm,l}$) and local features such as the dynamic spatial cost of the curves (i.e. $\mathbf{dsc}(t)$).

4.2. Grasping - Feature Extraction using Activity Curves

In order to provide a complete analysis of the prehension movement, an approach to the study of the palm and finger movement for recognition reasons, during a grasping movement is presented in the current section. Provided that no robust, vision based and unobtrusive method for the tracking of these movements is proposed up to now in the literature, and considering the fact that its development would be out of the scope of the current thesis, the tracking on the grasping movement has been exclusively based on the a wired glove sensor. Similarly to the approach followed in Section 4.1, the movement of the palm's base and the finger's phalanxes will be studied via the introduction of the corresponding Activity Curves, as described hereafter.

4.2.1. Tracking the Grasping Movement

In order to cover the second part of a prehension movement (i.e. grasping), the tracking of the fingers during the grasping activity is required. The device that has been utilized for this scope is the CyberGlove [252]. It provides the current angles between the phalanxes of the hand and their shifts

over time by translating the changes of the flowing current on the surface of the glove, caused by the bent deformations of integrated thin metallic layers, into real angle values, according to linear transformation. In this respect, the 3D reconstruction of the hand is possible for visual verification of the tracking. Specifically, each finger has been assigned 4 Degrees of Freedom (DoF), while they consist of 3 phalanxes, as it is illustrated in Figure 4.13(a). A further short illustration of them is also provided in the right side of Figure 4.12 and Figure 4.4, as well as in Figure 4.13(b)). Similarly, another 3 DoFs have been assigned to the joint in the base (i.e. wrist) of the palm.

At this point, it should be noted that according to the authors knowledge, there is no available vision-based tracker for detecting 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 authentication potential of the aforementioned arm based movements.

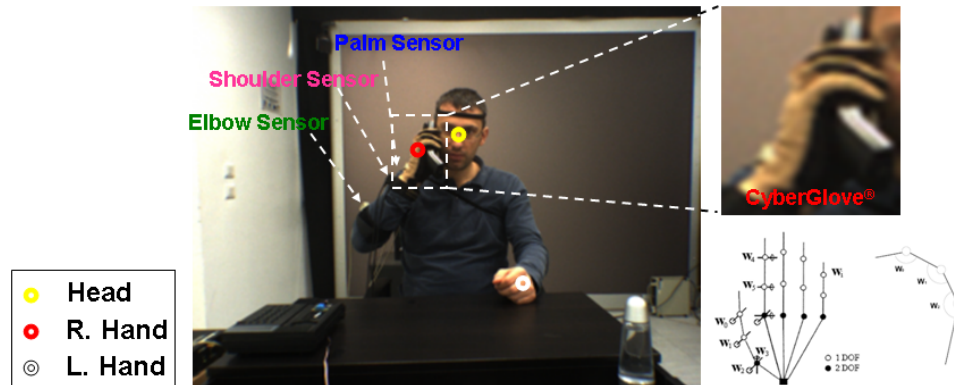


Figure 4.12.: Simultaneous data recording from three tracking sources during a *Phone Conversation* activity.

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.

The most typical 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 $\mathbf{P} = \{\theta_1, \theta_2, \dots, \theta_N\}$, where N stands for the total number of the utilized fingers' joints.

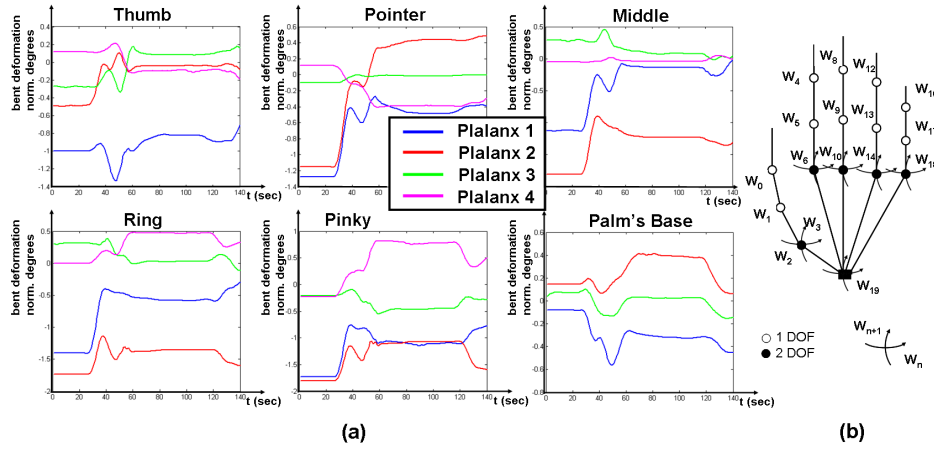


Figure 4.13.: (a) Raw Angles of the Finger Phalanxes, (b) Notation of Finger Names.

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

$$\mathbf{P}(t) = \{\theta_1(t), \theta_2(t), \dots, \theta_N(t) | t \in [t_e - t_0, t_e]\} \quad (4.23)$$

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 4.2.1), $N = 23$.

4.2.2. Angular Speed, Angular Acceleration and Angular Jerk

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

$$\omega_\theta = \frac{d\theta}{dt} \quad \alpha_\theta = \frac{d\omega}{dt} = \frac{d^2\theta}{dt^2} \quad \beta_\theta = \frac{d\alpha_\theta}{dt} = \frac{d^2\omega}{dt^2} = \frac{d^3\theta}{dt^3} \quad (4.24)$$

4.2.3. Dynamic Travel Cost

Rosenbaum et al. [190] introduced in their study the total travel cost $\mathbf{dtc}_p(t) = \sum_{j=1}^{N=23} dtc_j(\theta_j(t), T_j)$, where θ_j is the angular displacement of the j^{th} 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) $dtc_j(\theta_j(t), T_j)$ is going to be utilized as a descriptor of the hand movements.

Specifically, the cost $dtc_j(\theta_j, T_j)$ 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:

$$dtc_j(\theta_j(t), T_j) = k_j \theta_j (1 + [T_j - T_j^*(\theta_j)]^2) \quad (4.25)$$

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

$$T_j^*(\theta_j(t)) = k_j \ln(\theta_j + 1), \quad k_j \geq 0 \quad (4.26)$$

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 4.3).

It is expected that distinctive variations will be extracted by equation

(4.25) due to the unique finger size and articulation characteristics of each user. Thus, the *Hard to circumvent requirement* will be also satisfied (see Section 4). More evidence on this issue will be also presented in the recognition capacity analysis that follows in Section 4.3.

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 $\mathbf{W} = \{\boldsymbol{\theta}(t), \boldsymbol{\omega}(t), \boldsymbol{\alpha}_\theta(t), \boldsymbol{\beta}_\theta(t), \mathbf{dtc}(t)\}$.

4.3. Evaluation of Prehension related Features Authentication Potential

This section deals with the evaluation of the authentication potential of the extracted features. Specifically, two measures are presented herein, 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.

4.3.1. 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 [219] between the inter-individual (f_i^{inter}) and intra-individual (f_i^{intra}) probability distributions of an entire population S is defined as follows:

$$D(f_i^{intra} || f_i^{inter}) = \int f_i^{intra} \log \frac{f_i^{intra}}{f_i^{inter}} dr \quad (4.27)$$

For the relative entropy, also known as Kullback and Leibler divergence [219], $D(f_i^{intra} || f_i^{inter})$ 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,

$$\text{i.e } D(f_i^{intra}||f_i^{inter}) \neq D(f_i^{inter}||f_i^{intra}) , \quad (4.28)$$

a notion of symmetry is usually inserted by the mean relative entropy:

$$D_{sym}(f_i^{intra}||f_i^{inter}) = D_{sym}(f_i^{inter}||f_i^{intra}) = \frac{D(f_i^{intra}||f_i^{inter}) + D(f_i^{inter}||f_i^{intra})}{2} \quad (4.29)$$

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:

$$I^{norm}(F_i^{inter}, F_j^{inter}) = \frac{I(F_i^{inter}, F_j^{inter})}{H(F_i^{inter}) + H(F_j^{inter})} \quad (4.30)$$

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

$$I(F_i^{inter}, F_j^{inter}) = \sum_{f_i^{inter} \in \mathbb{R}} \sum_{f_j^{inter} \in \mathbb{R}} f_{i,j}^{inter} \log \frac{f_{i,j}^{inter}}{f_i^{inter} f_j^{inter}} \quad (4.31)$$

4.4. Classification and User Authentication

The current section deals with the algorithms that will be utilized in Section 6, in order to realize the final step of a recognition process, i.e. the comparison of the incoming signature with the ones stored in the database (i.e. enrollment), towards the classification of the corresponding user to a client or an impostor.

Two very commonly used modelling methods in machine learning and in robotics are the Gaussian Mixture Models (GMM) and its spatiotemporal extension, the Hidden Markov Models (HMM), are presented in Section

4.4.1 and Section 4.4.2. In general, model-based recognition and classification has been extensively used in applications such as action recognition [205], sports video analysis [206], speech/speaker recognition [207], etc. In this section, we present the GMM-based modeling, wherein the probability density function (PDF) corresponding to the data matrices \mathbf{V} and \mathbf{W} (see Section 4.1.2 and Section 4.2.3) for each activity is represented using Gaussian mixtures. The successful static PDF estimation using GMMs is further extended to robustly model temporal variations using continuous-density HMMs.

The extracted features for each user render traits of his behavioural characteristics towards a certain action and can thus include discrimination capacity. In order to utilize features from different enrolments and thus, to produce a more robust and more invariant signature, statistical models will be trained from the activity-related traits. Similarly, other signatures can later be evaluated and classified when compared with the trained statistical model. The cluster signature has two merits. First, it is a model-based abstract representation of a motion class. Thus, a motion pattern can be described more efficiently in terms of a signature model. Second, the cluster signature is not sensitive to trajectory length and thus the preprocessing step for . Even the matching of two full signatures requires that the metric must be able to account for the inconsistency in trajectory length and point distribution.

Herein, although GMM based signature classification has not been exploited, since it does not efficiently handle the temporal information of the extracted features, it is briefly described in Section 4.4.1, since it forms an integral part of the HMM algorithm.

On the other hand, clustered signatures may exhibit drawbacks, such as increased processing time and overfitting issues. For this reason an alternative efficient comparison method that requires no training and takes into account the spatiotemporal information of the input signals is the Dynamic Time Warping, as presented in 4.4.3.

4.4.1. Gaussian Mixture Model

An efficient way to cluster two or more signatures in a statistical model is provided by the Gaussian Mixture Model method [208], which has been

mobilized in similar case studies in [215] and in [209]. Gaussian mixture models (GMM) are used to learn a model for a motion class by the density estimation for a cluster of full or optimized signatures.

Following the principle of a model-based method, we model each motion class by a probability distribution model. Moreover we classify the trajectories into different classes. The word “*class*” refers to a type of activity (represented by its full set of trajectories) for each user, for which we have sufficient number of samples to train the system.

Assuming that the number of all motion classes for all users is C , C models will be learned via training which will be characterized by respective model parameters $\{\Theta_i\}_{i=1}^C$. Moreover, S_i contains M signature samples ($S_i = \{S_{i,m}\}_{m=1}^M$), which serve as the training samples to train an individual model i . After we have rearranged these samples in the form of $S_i = [S_{i,1}, S_{i,2}, \dots, S_{i,m}, \dots, S_{i,M}]$, the underlying probability density function (pdf) of S_i can be modeled by a mixture of Gaussian functions in the following form:

$$P(S_i|\Theta) = \sum_{k=1}^K w_k \mathbb{N}(S_i; \mu_k, \Sigma_k) \quad (4.32)$$

whereby S_i the input trajectory signal, K is the number of mixing Gaussian components, w_k are the mixing weights with $w_k = P(k|\Theta_i)$ obeying $\sum_{k=1}^K w_k = 1$, and $\mathbb{N}(S_i; \mu_k, \Sigma_k)$ denotes the multivariate Gaussian function.

The mixture is completely specified by parameters $\Theta = \{w_k, \mu_k, \Sigma_k\}_{k=1}^K$. Since the parameter estimation phase is identical for each class and the training is performed on the disjoint dataset of these classes, the class indexing subscript will be omitted hereafter from our notation for brevity.

Now, given a training set of trajectories with length T for a particular class $\{\mathbf{s}_t\}_{t=1}^T$, the mixture parameters can be estimated using the maximum likelihood:

$$\Theta^* = \arg \max \left[\prod_{t=1}^T P(\mathbf{s}_t|\Theta) \right] \quad (4.33)$$

This estimation problem can be solved using the *EM* algorithm [210]. The initial parameters of the *EM* algorithm, which is an iterative algorithm, are provided by the utilization of the *k – means* method [208]. The

EM algorithm transfers a single and difficult optimization problem into a sequence of smaller and simpler problems. Specifically, it seeks to maximize the likelihood function by the gradient descent technique. Here instead of the likelihood function, the log-likelihood is used:

$$L(S_c|GMM_{n,k}) = \sum_{t=1}^T \log \left(\sum_{k=1}^K w_j P(j|\mu_j, \Sigma_j) \right) \quad (4.34)$$

T the size of the incoming trajectory vector, the w_j the weight factor of the j^{th} cluster, μ_j and Σ_j are the mean value and the variation of the distribution in the j^{th} cluster.

Once the GMMs for all activities have been learnt, the classification of new trajectories can be performed by computing the likelihood for each GMM. For this purpose, each new incoming vector of the input trajectory is posed as an observation sequence to each GMM. During this computation, the likelihood is computed as shown in Eq. (4.34) and the corresponding weights are applied to generate the likelihood of the Gaussian mixture. The trajectory is declared to belong to the subject represented by the GMM with the highest likelihood.

4.4.2. Hidden Markov Models

Despite their usefulness and their simple implementation, Gaussian Mixture Models (GMM) have been proven in the past inadequate for capturing the temporal relations and ordering of the successive tracked locations during a movement. For this reason, they have not been used in the current thesis for classification means. However, they form an integral part of the Hidden Markov Model (HMM) algorithm, and thus they have been presented in Section 4.4.1. In this respect, given that the extracted trajectories $s_l(t)$ exhibit a strong dependence on temporal ordering, the HMM algorithm has been utilized for both the training and the authentication/recognition session of the current recognition module, as described below.

It should be noted that due to the similarities of the experiments carried out in the evaluation Sections of the thesis, whenever an experiment engages the training of an HMM model, the parameters defined at this stage remain unaltered.

In this respect, the first parameter specified for an HMM is the number of

states. The number of states of the utilized HMM has been set equal to the maximum number of changes in the direction of the palms and head during the performance of an activity. Thus, a five-state, left-to-right, fully connected HMM is trained from several enrollment sessions of the same user for the given activity. In other words, the observation data, which are no other than the 3D points of the detected positions of the head/arm/elbow/palm during the performed movement, are assigned to one of the hidden states of the Hidden Markov Model. This way, not only the spatial position is evaluated, but also the transitions between these states over time (i.e. temporal information).

In order to verify the aforementioned assumption also 3-, 4-, 5- 6- and 7-state HMMs have been trained and experimentally utilized in subsets of the available datasets (see Section 3.2.2). The results have not only shown that the 5-state formulation of the HMM algorithm achieved the best performance in terms of distinctiveness between different HMM models, for the same incoming genuine trajectory, but they also exhibited the a sparse distribution of matching probabilities between incoming genuine and impostor trajectories. For instance, the following table (i.e. Table 4.4.2) exhibits the results derived by producing Hidden Mark Models of different amount of states at their training phase, as biometric signatures. The experiment that has been carried out to prove the aforementioned statement is the same with the one described in Section 6.1.3.

Table 4.3.: Authentication Performance, measured as EER scores for different amount of utilized states in the training of the HMM based signature.

Movement ID	3 States	4 States	5 States	6 States	7 States
Phone Conversation	37.4%	22.6%	16.7%	18.3%	25.0%
Interaction with the Microphone Panel	31.7%	19.2%	10.3%	13.3%	19.8%

Once the number of states is fixed, the complete set of model parameters describing the HMM is given by:

$$\lambda = \{\pi_j, \alpha_{ij}, b_j\} \quad (4.35)$$

where π_j is the probability of the j^{th} state being the first state among all the trajectories, α_{ij} denotes the probability of the j^{th} state occurring immediately after the i^{th} state, and b_j denotes the PDF of the j^{th} state.

A Gaussian mixture-based representation is used for each state PDF as indicated in equation 4.32. Then, the state variable q_t , which corresponds to the t^{th} state of the utilized HMM, takes one of T values $q_t \in \{s_1, \dots, s_T\}$. Since a Markovian process is assumed, the probability distribution of q_{t+1} depends only on q_t . This is described by the state transition probability matrix A whose elements a_{ij} represent the probability that q_{t+1} corresponds to state s_j given that q_t corresponds to s_i . The initial state probabilities are denoted by π_j , namely that the probability q_1 corresponds to state s_1 .

The observational data O_t from each state of the HMM are generated according to a PDF dependent on the instant of t^{th} state, denoted by $b_j(O_t)$. This state-conditional observation PDF is modeled as a Gaussian mixture as indicated below:

$$b_j(O_t) = \sum_{k=1}^K w_{jk} \frac{1}{(2\pi)^{P/2} |\Sigma_{jk}|^{1/2}} \exp \left\{ -\frac{1}{2} (O - \mu_{jk})^T \Sigma_{jk}^{-1} (O - \mu_{jk}) \right\} \quad (4.36)$$

whereby w_{jk} , μ_{jk} and Σ_{jk} denote the scalar mixing parameter, P -dimensional mean vector and $P \times P$ covariance matrix of the k^{th} Gaussian component in the j^{th} state. Here, each Gaussian component is a multivariate normal distribution of the same dimensionality, since all trajectories are described with three dimensions. The parameters of the HMM are initialized to random values and the BaumWelch algorithm is used for estimation using the forward-backward procedure [211].

At this point, the selection of the fully connected HMM should be justified. The main reason for this choice is to allow the training algorithm to form those transition for each HMM that best describe the movement of the joints over time. However, in typical movement of the hand during any of the scenarios in the utilized datasets (see Section 3.2.2), only a few of the interconnections between the different states of the HMMs are enabled, as shown in the transition matrices in Eq. (4.37), Eq. (4.38 and Eq. (4.39) for

the average user, indicatively.

$$T_{phone} = \begin{bmatrix} 8.96 * 10^{-1} & 1.03 * 10^{-1} & 0.00 * 10^{-0} & 0.00 * 10^{-0} & 0.00 * 10^{-0} \\ 0.00 * 10^{-0} & 9.42 * 10^{-1} & 5.81 * 10^{-2} & 0.00 * 10^{-0} & 0.00 * 10^{-0} \\ 0.00 * 10^{-0} & 0.00 * 10^{-0} & 9.27 * 10^{-1} & 7.32 * 10^{-2} & 0.00 * 10^{-0} \\ 0.00 * 10^{-0} & 0.00 * 10^{-0} & 0.00 * 10^{-0} & 9.41 * 10^{-1} & 5.89 * 10^{-2} \\ 0.00 * 10^{-0} & 0.00 * 10^{-0} & 0.00 * 10^{-0} & 0.00 * 10^{-0} & 9.99 * 10^{-1} \end{bmatrix} \quad (4.37)$$

$$T_{panel} = \begin{bmatrix} 8.37 * 10^{-1} & 1.62 * 10^{-1} & 0.00 * 10^{-0} & 0.00 * 10^{-0} & 0.00 * 10^{-0} \\ 0.00 * 10^{-0} & 8.22 * 10^{-1} & 1.78 * 10^{-1} & 0.00 * 10^{-0} & 0.00 * 10^{-0} \\ 0.00 * 10^{-0} & 0.00 * 10^{-0} & 8.59 * 10^{-1} & 1.40 * 10^{-1} & 0.00 * 10^{-0} \\ 0.00 * 10^{-0} & 0.00 * 10^{-0} & 0.00 * 10^{-0} & 9.08 * 10^{-1} & 9.17 * 10^{-2} \\ 0.00 * 10^{-0} & 0.00 * 10^{-0} & 0.00 * 10^{-0} & 0.00 * 10^{-0} & 1.00 * 10^{-0} \end{bmatrix} \quad (4.38)$$

$$T_{combined} = \begin{bmatrix} 9.59 * 10^{-1} & 4.05 * 10^{-2} & 0.00 * 10^{-0} & 0.00 * 10^{-0} & 0.00 * 10^{-0} \\ 0.00 * 10^{-0} & 9.39 * 10^{-1} & 6.00 * 10^{-2} & 0.00 * 10^{-0} & 0.00 * 10^{-0} \\ 0.00 * 10^{-0} & 0.00 * 10^{-0} & 9.50 * 10^{-2} & 4.94 * 10^{-2} & 0.00 * 10^{-0} \\ 0.00 * 10^{-0} & 0.00 * 10^{-0} & 0.00 * 10^{-0} & 9.19 * 10^{-1} & 8.10 * 10^{-2} \\ 0.00 * 10^{-0} & 0.00 * 10^{-0} & 0.00 * 10^{-0} & 0.00 * 10^{-0} & 1.00 * 10^{-0} \end{bmatrix} \quad (4.39)$$

Once the training phase has been completed, new trajectories are categorized as one of the learned users for the specific activity based on the maximum likelihood criterion (ML) principle. Given HMMs for the L enrolled subjects, $\lambda_1, \lambda_2, \dots, \lambda_L$, and the new trajectory vectors $s'_i(t)$ of the incoming trajectory vectors from the new recording (i.e. the observation sequence) O_1, O_2, \dots, O_m , we assign user label m as the HMM that maximizes the likelihood given the new trajectory [211]:

$$m = \arg \max_{i \in [1, \dots, L]} \sum_j P(O_{t+1:k} | q_t^i = j, O_{1:t}) P(q_t^i = j, O_{1:k}) \quad (4.40)$$

The above computation can be efficiently performed using the forward

recursion procedure in the BaumWelch algorithm [211].

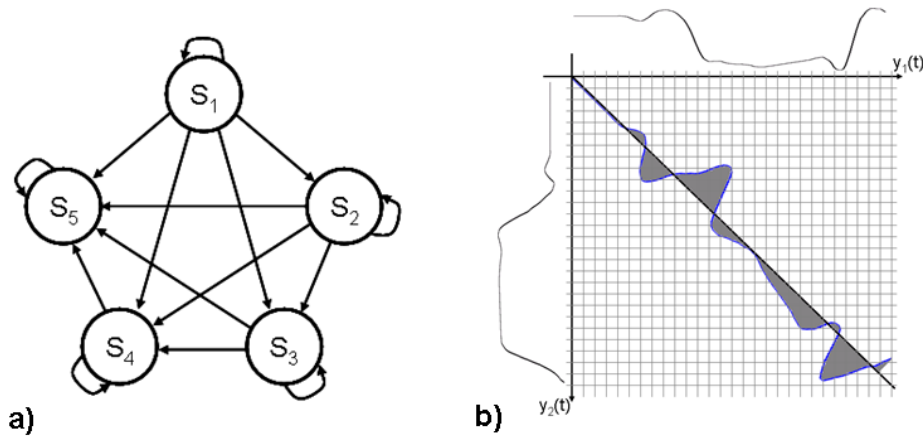


Figure 4.14.: a) Graphical representation of the utilized 5-state, left-to-right and fully connected Hidden Markov Model. b) Estimation of the matching score between the “gallery” and the “probe” vectors using a DTW-Grid. The plotted diagonal presents the (optimal) path with the least difference cost, 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 [4].

4.4.3. Dynamic Time Warping

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 (SHC), 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 [4] 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 $L1 - norm$.

Used for calculating a metric about the dissimilarity between two (feature) vectors, DTW is based on the difference cost that is associated with the

matching path computed via dynamic programming, namely the Dynamic Time Warping (*DTW*) algorithm. The *DTW* algorithm can provide either a valuable tool for stretching, compressing or aligning time shifted signals ([4]) or a metric for the similarity between two vectors ([237]). Specifically, it has been widely used in a series of matching problems, varying from speech processing ([4]) to biometric recognition applications ([39]). A possible implementation basis on estimating the closed area formed by the path around the diagonal of the rectangular *DTW*-grid (Figure 4.14(b)). 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 [4]. Its main advantages are its simple implementation and its satisfactory performance given the required processing time.

A short description of the functionality of *DTW* algorithm for comparing two one-dimensional vectors (probe & gallery signal) is presented below:

The probe vector \mathbf{p} of length L is aligned along the X-axis while the gallery vector \mathbf{g} of length L' is aligned along the Y-axis of a rectangular grid respectively. In our case $L \equiv L'$ as a result of the preprocessing steps (Section 4.1.1 or Section 4.1.2). Each node (i,j) on the grid represents a match of the i_{th} element of \mathbf{p} with the j_{th} element of \mathbf{g} . The matching values of each $\mathbf{p}(i), \mathbf{g}(j)$ pair are stored in a cost matrix C_M associated with the grid. $c(1, 1) = 0$ by definition and all warping paths are a concatenation of nodes starting from node (1, 1) to node (L, L).

The main task is to find the path for which the least cost is associated. Thus the difference cost between the two feature vectors is provided. In this respect, let (x_k, y_k) represent a node on a warping path at the instance k of matching. The full cost $D(x_k, y_k)$ associated to a path starting from node (1, 1) and ending at node (x_k, y_k) can be calculated as:

$$D(x_k, y_k) = D(x_{(k-1)}, y_{(k-1)}) + c(x_k, y_k) = \sum_{m=1}^k c(x_m, y_m) \quad (4.41)$$

Accordingly, the problem of finding the optimal path can be reduced to finding this sequence of nodes (x_k, y_k) , which minimizes $D(x_k, y_k)$ along the complete path.

As stated by Sakoe and Chiba in ([4]), a good path is unlikely to wander very far from the diagonal. Thus, the path with minimum difference cost, would be the one that draws the thinnest surface around the diagonal as shown by the dashed lines in Figure 4.14(b). In the ideal case of perfect matching between two identical vectors, the area of the drawn surface would be eliminated. The closed area around the diagonal can be calculated by counting the nodes between the path and the diagonal at every row ([110]) as indicated by the following equation.

$$V(p_i, q_j) = \begin{cases} 1, & \text{if } (i > j) \text{ of } N(p_i, q_j) \\ & \text{for } j = j, j + 1, \dots, j + d, \text{ where } d = i - j \\ 1, & \text{if } (i < j) \text{ of } N(p_i, q_j) \\ & \text{for } i = i, i + 1, \dots, i + d, \text{ where } d = i - j \\ 1, & \text{if } (i = j) \text{ of } N(p_i, q_j) \\ 0, & \text{otherwise} \end{cases} \quad (4.42)$$

Thus, the value $V(p_i, q_j) = 1$ to these nodes. On the contrary, all other nodes lying outside the closed area will be assigned the value $V(p_i, q_j) = 0$. Then, the total area S created by the path is mathematically stated as following:

$$A_c = \sum_{i=1}^T \sum_{j=1}^L V(p_i, q_j) \quad (4.43)$$

whereby

Finally the total dissimilarity measure d_{DTW} between vector \mathbf{p} and \mathbf{g} (Equation 4.43) can be computed as the product of area size A_c and the minimum full cost $D(T, T)$ (Equation 4.41):

$$d_{DTW} = A_c \cdot D_{min}(T, T) \quad (4.44)$$

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 are 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:

$$D_{tot} = \sum_{j \in \{C, M, G\}} w_j d_{j, DTW} \quad (4.45)$$

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

$$w_j = \frac{\text{bits of information for all features of device } j}{\text{total Number of bits for all utilized features}} \quad (4.46)$$

4.5. Summary

Summarizing, in the current section a novel biometric module has been proposed, exploiting the dynamic characteristics of the movements of the arm and the finger. The feature extraction procedures for each of the two body-parts can be schematically seen in Figure 4.15 and Figure 4.16, respectively.

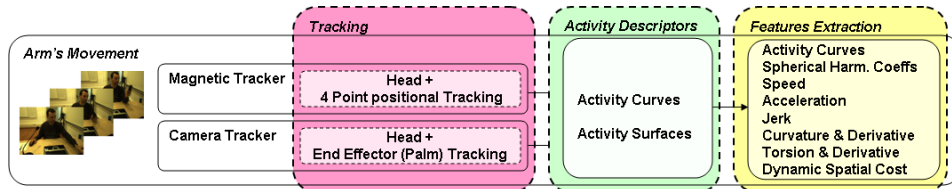


Figure 4.15.: Flow Chart diagram of the procedure followed for the extraction of dynamic features from the movement of the arm.

Specifically, the movement of the arm of the user is initially recorded by two different types of trackers, i.e. an proprietary vision-based tracker that manages to effectively capture the movement of the head and the palm in the $3D$ space, and a 4-point wired sensor-based tracker that accurately detects the location of the head, shoulder, elbow and palm of the user in short timesteps (Figure 4.15). Next, the aforementioned locations are used for the description of the performed movement via generating the so-



Figure 4.16.: Flow Chart diagram of the procedure followed for the extraction of dynamic features from the movement of the fingers.

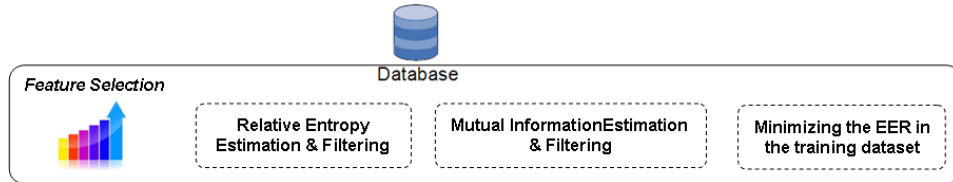


Figure 4.17.: Flow Chart diagram of the procedure followed for the selection of the most indicative activity related features.

called “Activity Surface” and “Activity Curves” descriptors. Finally, the last building block refers both to the direct processing and the processing via transformations of these descriptors towards the extraction of these activity related features that are indicative for user recognition.

Similarly, the movement of the fingers of the hand of the user are tracked in means of angles between the phalanxes (i.e. via the translation of deformation of the integrated thin metallic layers on the surface of the glove, into real angle values). In the next step, the so-called “Activity Curves” are generated from the successive tracked angles and are used as the descriptors of the movements of the hand. Lastly, the activity related features that are characterized by adequate recognition capacity, are extracted via the corresponding processing.

Following the aforementioned activity-related feature extraction approach, the proposed feature selection methodology is illustrated in Figure 4.17. Initially, each feature is evaluated in terms of its relative entropy value, while the mutual information between all possible pairs of features are estimated in a confusion matrix. Based on these values, an iterative algorithm, run on a training dataset (i.e. a subset of the utilized dataset), is applied, in order to exclude the redundant and noisy features by aiming at the lowest EER value.

The proposed end-to-end recognition approach is depicted in Figure 4.18

for both enrollment (i.e. training) and recognition modes, respectively. In particular, the high level flow chart diagrams of the followed process consist of the aforementioned tracking module (i.e. Figure 4.15 and Figure 4.16) and are followed by the filtering out of the least indicative features for user recognition, as concluded in the module in Figure 4.17. The final decision, regarding the validity of the identity of the user is taken in the last building block via the utilization of two efficient classification algorithms that effectively handle the comparison between the incoming signature and the stored template.

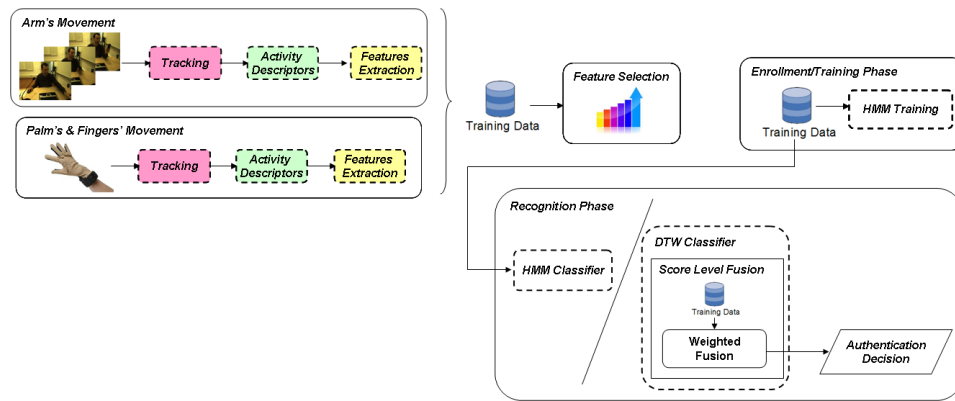


Figure 4.18.: High Level flow chart diagram of the procedure followed for the enrollment/training and recognition phase of the novel biometric system.

An analytical evaluation of the proposed methodology regarding both the validity of the feature selection methodology and the classification results will be presented in Chapter 6.

5. Enhancement of Biometric Systems using Anthropometric and Soft Characteristics

As it has already been mentioned, activity related and behavioural biometrics systems lack in recognition performance when compared to traditional ones. This way, however, the advantages they impose in terms of unobtrusiveness and easy integration characteristics are compensated by their relatively limited accuracy. Thus, innovative solutions have to be found out, so as to transparently narrow the gap with traditional biometrics.

In this respect, following the literature review performed in Chapter 1 and specifically in Section 2.3, multi-biometrics have the potential to offer a promising solution in the aforementioned direction. In particular, the current Chapter deals with the enhancement of existing biometric systems via the incorporation in the recognition process of static anthropometric traits (Section 5.1) and other soft biometric (Section 5.2) that can be extracted simultaneously during the recognition process of the user. This way, no extra sensors are required, while the fact that principally there is no correlation between soft biometrics and the originally extracted behavioural biometric traits, no redundant information is processed, improving thus, both the recognition performance and the accuracy of the initial system.

Hereafter, Section 5.1 deals initially with the evaluation of the static anthropometrics (i.e. length of the sections of the upperbody) of the user and then with their contribution to the original activity related recognition process presented in Chapter 4.

Moreover, in order to avoid the concept of late fusion in multi-biometrics, which can be both computationally expensive and rather ineffective, and in order to propose a generic approach for the incorporation of soft biometrics in any any biometric system, a probabilistic framework for enhancing the

performance of the latter via the utilization of continuous soft biometric traits is proposed in Section 5.2.

5.1. Static Anthropometric Profile

A significant enhancement to the authentication performance of the system described in Section 4.1 can be achieved by exploiting the static anthropometric information of each user, i.e. a user-specific skeleton model. At this point, it should be clarified that the development or the improvement of an gesture recognition technique is out of the scope of the current thesis. Specifically, the goal herein is to exhibit the potential of static anthropometric features towards biometric recognition.

Thus, two state-of-the-art methods are utilized in the current section for the extraction of the users' static biometric profile. The first is described in [42], whereby hierarchical particle filtering is utilized towards the accurate shape adjustment of an articulated model to the user's body. The multi-camera environment requested by this approach is provided by two calibrated cameras: a stereo frontal camera and a usb-simple camera, which is placed on top of the user.

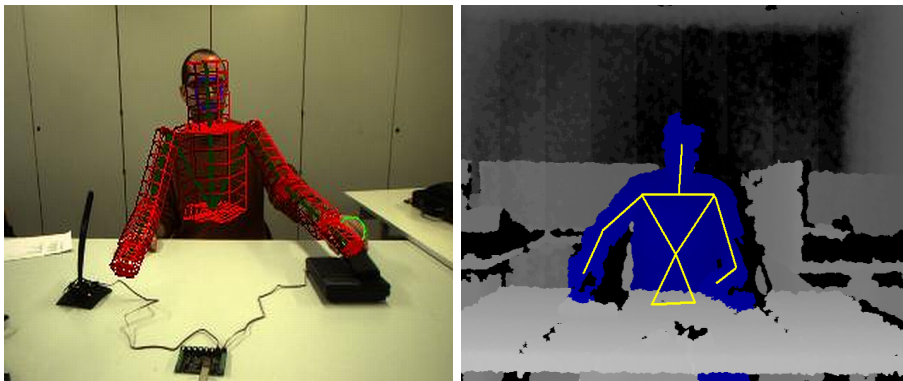


Figure 5.1.: Adjusted skeleton model based on: a) hierarchical filtering, b) OpenNI algorithms

Alternatively to the aforementioned method, a faster and more accurate method has been lately released. The latter utilizes the PrimeSense[®] advanced depth-sensor in combination with the OpenNI [160] library. Thus, the human form is segmented automatically from the high precision depth

image, while 48 essential points of the human body are simultaneously tracked in the 3D space.

The core of the OpenNI library is a machine learning algorithm that has been statistically trained by millions of images of people in different poses. The statistical compilation of all these data allows OpenNI to adjust the most appropriate skeleton model to each human body in terms of size and pose. The implemented methodology is covered by an international patent and is described in [199].

When comparing these two approaches, one could notice that in the current setting the particle filtering algorithm utilized in [42] requires ~ 15 seconds for the processing of a single shot ($1 \text{ shot} \equiv 1 \frac{\text{frame}}{\text{camera}}$). However, it has been found out that an initial approximate manual annotation of the user's joints may significantly increase the performance of the algorithm with respect to the achieved accuracy.

On the other hand, the OpenNI algorithm exhibits much lower computational requirements (30fps), with a slight decrease in accuracy. A comparison in terms of biometric recognition performance between the aforementioned methods, as well as their contribution to the carried out experiments follow in Section 6.1.3.

Once the location of all body's joints have been estimated, the extracted user's skeleton model can be represented by an Attributed Relational Graph (ARG) $G = \{V, E, \{A\}, \{B\}\}$ [200], whereby V are the nodes, E the edges, and A and B the corresponding attributes, respectively. The nodes and the edges stand for the joints and the limbs of the actual body, respectively, as shown in Figure 5.2. Attribute matrix \mathbf{A} is not used, since no attributes for the joints are utilized in the current framework, while attribute matrix \mathbf{B} corresponds to the lengths of the limbs (\equiv distances between the adjacent joints).

5.1.1. Attributed Graph Matching

Possible noisy estimation of the limbs' lengths is compensated when calculating the mean value of each anthropometric attribute among several enrollment sessions. However, there are some cases, where partially connected anthropometric graphs may be generated. This may be due to either partial occlusions of specific limbs from other foreground objects or low confidence

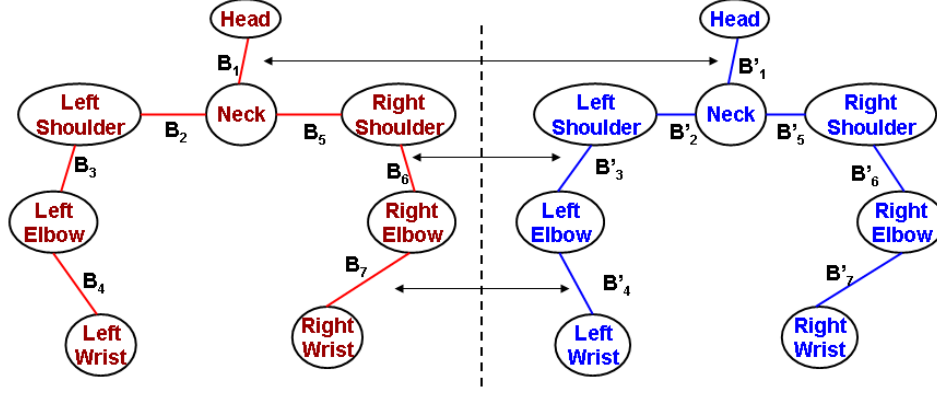


Figure 5.2.: Anthropometric Graphs' Comparison.

tracking (i.e. bad illumination). The Attributed Graph Matcher (AGM) based on Kronecker Graphs [200] has been utilized, whereby comparison between fully and partially connected graph is possible.

Let us assume two random anthropometric Graphs G and G' as shown below:

$$\begin{aligned} G &= \{V, E, \{B\}_{i=1}^n\}, \text{ where } n := |V| \\ G' &= \{V', E', \{B'\}_{i=1}^{n'}\}, \text{ where } n' := |V'| \end{aligned} \quad (5.1)$$

where B_k carries the lengths of the user's upper-body limbs.

The case of $n \neq n'$ indicates a Sub-Graph Matching (SGM), while $n = n'$ a Full-Graph Matching (FGM). In any case, Graph G is claimed to match to a sub-graph of G' , if there exists an $n \times n'$ permutation sub-matrix P so that the following equation is fulfilled.

$$B_j = P_0 B'_j P_0^T + (\varepsilon M_j), \text{ where } j = 1, \dots, r \quad (5.2)$$

where M_j is an $n \times n$ noise vector and ε is related to the noise power and is assumed to be independent of the indices i and j .

To accommodate inexactness in the modelling process due to noise, the AGM problem can be expressed as the combinatorial optimization problem of equation:

$$\epsilon = \min_p \left(\sum_{j=1}^s W_{j+r} \|B_j - P B'_j P^T\|^2 \right) \quad (5.3)$$

where $\|\cdot\|$ represents some norm $\mathbf{P} \in Per(n, n_0)$ denotes the set of all $n \times n_0$ permutation submatrices and $\{\mathbf{W}_i\}_{k=1}^{r+s}$ is a set of weights satisfying $0 \leq \mathbf{W}_k \leq 1$, $k = 1, \dots, r + s$ and $\sum_{k=1}^{r+s} \mathbf{W}_k = 1$.

In this respect, the minimum error ϵ stands for a metric for the similarity between the graphs under comparison.

5.2. Systematic Error Analysis of Soft Biometrics

Similarly to the anthropometric traits, soft biometrics belong to this category of human characteristics that are representative of individuals, but are not yet unique and discriminative enough to distinguish them within a large group. Following the previous chapter, where

This section presents a more generic probabilistic framework for augmenting the recognition performance of biometric systems with information from continuous soft biometric traits. In particular, by partitioning the soft biometric feature space in proximity-related similar cluster and by modelling the systematic error induced by the estimation of the soft biometric traits, a modified efficient recognition probability can be extracted including information related both to the hard and soft biometric traits.

Inspired from the works presented in [6] and [220], the author of the current thesis attempts to extend and generalize the idea of quantizing the multidimensional soft biometric feature space, that has been initially proposed in [221], into a generic framework for boosting the matching score of the basic biometric recognition system, via a probabilistic approach and a more efficient clustering of the feature space.

Improving the aforementioned concept, the current thesis addresses some serious open issues that have been raised in both aforementioned works. In particular, despite its seemingly smooth function, the bayesian framework proposed in [6] is based on the false assumption of independent conditional probabilities of the geometric trait when multiple soft biometrics are available. Moreover, only a naive separation of the feature space takes place thereby.

It should be emphasized that the proposed approach does not aim at the fusion of different traits [220], including also soft biometrics, at the score level as performed in [222] that exhibits some disadvantages like the need of the computation of a soft biometric score or weighting functions for fusion

at the score level based usually on posterior probabilities.

5.2.1. Modelling Soft Biometrics

The present thesis proposes a novel and highly efficient probabilistic framework for the integration of one or more soft biometric traits in any biometric recognition system, by taking advantage of the error induced by the system when measuring each soft biometric characteristic.

At this point it should be emphasized, that, similarly to [220] and [6] and contrary to [222], no fusion between the conventional biometric recognition score and the soft biometric matching score takes place. As such, there is no need for computation of a soft biometric score, weighting functions or posterior probabilities.

Let Ω be the set of all identities in the M -sized user population $\Omega = \{\omega_1, \omega_2, \dots, \omega_M\}$, x_c be the hard biometric information (e.g. geometric gait) and x_s be a continuous soft biometric trait (i.e. the height of the user) from a set \mathbf{X} with N available soft biometrics $X = \{x_{s_1}, x_{s_2}, \dots, x_{s_N}\}$. As such, $p(\omega|x_c) = 1 - p(\bar{\omega}|x_c)$ is the matching score of the conventional biometric system.

Partitioning the feature space

In both previous referenced works [220] and [6], the boosting is only augmenting the final recognition performance when applied to specific user groups. Specifically, in [220], users are categorized into “minority” and “majority” groups, according to the frequency of appearance of their soft biometric traits, while in [6], only extreme cases of soft biometric traits are boosted. Another drawback of these approaches lies in the fact, that an “extreme” or “minority” case is only defined in a single dimension. This leads to a uniform, linear quantization of the feature space, which is not the case in most real scenarios.

Contrary to the simple 3-stage partitioning (i.e. small-, normal- and large-sized population) [6] [220], a more sophisticated spatial partitioning of the feature space F in N_C clusters C_i , that exhibit notable variation in terms of their defining soft biometrics, is proposed herein. This way, the authentication probability of a client user is augmented, when the incoming soft biometric traits refer to the same cluster as the claimed ID. In all other

cases, the matching probability $p(\omega|x_c)$ remains untouched and is solely based hard biometric trait of the user. Although there is no actual limitation in the dimensionality of the clusters, the simple case of $2D$ clusters will be studied herein, without loss of generality.

In this respect, a cluster C_i of the multidimensional soft biometric feature space, associated to a subset $S_i \in \Omega$ of the set of identities Ω , is characterized as a valid cluster iff the following hold.

In particular, the a-priori probability of an identity ω to belong to a cluster has to be low:

$$0 < p(\omega \in C_i | x_{s_1}, \dots, x_{s_N}) = p_i(\omega) \ll 1$$

and there should exist a subset S_i of Ω , so as

$$\exists S_i \subset \Omega \begin{cases} \forall \omega \in S_i, p(x_s \in C_i | \omega) > \alpha \\ \forall \omega \notin S_i, p(x_s \in C_i | \omega) \approx p_i(\omega) \end{cases}$$

where α is a minimum non-zero value, C is the union of all clusters whose number N_C should be significantly lower than the size $|\Omega|$ of the identity set Ω :

$$\begin{aligned} C &= \cup C_i, \forall i = 1, \dots, N_C \\ N_C &\ll |\Omega| \end{aligned}$$

Three different partitioning alternatives have been implemented herein, all of which fulfil the requirements of a cluster:

Orthogonal Grouping (OG): A linear and possibly the simplest way of clustering the feature space its partitioning into uniform orthogonal clusters (Figure 5.3a). This kind of partitioning has been proposed in [220] and [6].

Using a brute force iterative algorithm on an adequately large reference soft biometric feature dataset, the dimensions of the prototype orthogonal cluster can be optimally defined. However, the major drawback of the current clustering method, as it has been implemented in [220] and [6], is the fact that it does not consider combined extreme cases of soft biometrics. On the contrary, it deals with each biometric feature separately. This way, some clusters are expected to be “left empty”, while others will possibly be “overcrowded”.

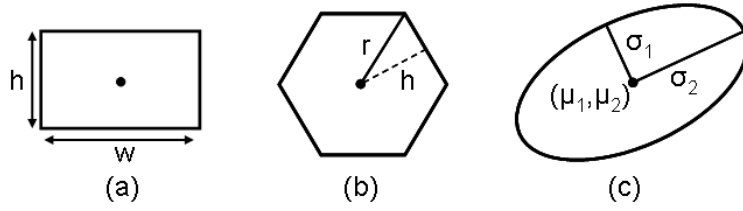


Figure 5.3.: a) Orthogonal Cluster - b) Hexagonal Cluster - c) Gaussian Cluster

Hexagonal Cell Grouping (HCG): A more efficient alternative for clustering the feature space is to partition it into adjacent identical hexagonal cells (Figure 5.3b). This way, isotropy is preserved along the whole feature space, while increased nonlinearity introduced, compared with the orthogonal grouping in Section 5.2.1.

In this case, the only parameter that has to be estimated and optimized is the hexagon's radius. Similarly, to the orthogonal grouping case, this can be experimentally specified on an adequately large reference soft biometric feature dataset of the same dimensionality. Hereby, all soft biometric data at each dimension have to be normalized using their corresponding standard deviation before being assigned to the hexagonal cluster, in order to conform to the isotropy of the current grouping.

Similarly to Section 5.2.1, neither the issue of possible empty nor that of overcrowded clusters is solved hereby, despite the increased non-linearity.

Gaussian Grouping (GG): Theoretically, the less linear the clustering is the more efficiently it will cover the feature space. In this respect, creating multidimensional gaussian clusters on the feature space is expected to provide increased flexibility in grouping similar users.

To this direction, an unsupervised clustering approach is utilized. Initially, the optimal number of clusters is estimated by utilizing the ISODATA clustering algorithm [223]. Then again, by exploiting the expectation-maximization (*EM*) algorithm [224], the soft biometric feature space can be easily described as a mixture of multidimensional (Figure 5.3c) Gaussian, whereby each Gaussian is described as

$$\mathcal{N}(\mathbf{f}_\omega | \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{Z/2} |\Sigma_k|^{1/2}} e^{-\frac{1}{2}(\mathbf{f}_\omega - \mu_k)^T \Sigma_k^{-1} (\mathbf{f}_\omega - \mu_k)} \quad (5.4)$$

Vector \mathbf{f}_ω includes the soft biometric trait values, $\mathbf{f}_\omega = \{x_{s_n,1}(\omega), \dots, x_{s_n,Z}(\omega)\}$, while μ_k and Σ_k are the Z -dimensional mean vector and the and $Z \times Z$ covariance matrix of the k^{th} Gaussian, respectively.

At the authentication stage, the assignment of a user's incoming soft biometric feature vector to a cluster occurs according to the maximum likelihood (ML) criterion.

Modelling the Noise

Let us now define the ground truth value $x_{s_n}^g$ as the soft biometric trait n of user ω and \tilde{x}_{s_n} as the l^{th} value measured by the system ($\tilde{X}_{s_n}(\omega) = \{\tilde{x}_{s_n,1}(\omega), \dots, \tilde{x}_{s_n,L}(\omega)\}$), where L is the total number of measurements. For an adequately large number $T = M \times L$ of measurements, the noise distribution that is induced as error in the measurement (i.e. noise) by the system can be estimated as described hereafter.

As long as T is large enough for reliable statistical estimates, the normalized values $e_{s_n,l}(\omega_m) = \tilde{x}_{s_n,l}(\omega_m) - x_{s_n}^g(\omega_m)$ can be produced. Having these data for the whole registered population, it is trivial to fit the normalized values distribution by a $1D$ Gaussian Mixture of the following type:

$$p(e_s | \omega) = \sum_{k=1}^K \pi_k \mathcal{N}_p(e_s | \mu_k, \sigma_k)$$

where $\mathcal{N}_p(e_s | \mu_k, \sigma_k)$ stands for the k^{th} single Gaussian distribution that contributes to the mixture. The values π_k , μ_k and σ_k can be easily computed by utilizing the iterative Expectation-Maximization (EM) algorithm on the data's histogram, until convergence.

The initial parameter regarding the number K of single Gaussian distributions in the $1D$ mixture model is experimentally selected, as the one that produces an acceptable error value in the $\chi^2 - test$:

$$\chi^2 \equiv \sum_{b=1}^B \frac{(O_b - E_b)^2}{E_b}, \quad (5.5)$$

where B is the total number of bins in the histogram, O_b the actual number

of samples in each bin and $E_b \equiv Tp(x_s)$. Once the two parameters, namely the degrees of freedom and the minimum allowed confidence f are set for the test, the value of χ^2 is cross-checked in the corresponding statistical tables. If it is below the corresponding threshold, it can be claimed that the data are compatible with the imposed mixture model with confidence f [225].

Consequently, $p(e_s|\bar{\omega})$ can be calculated as

$$p(e_s|\bar{\omega}) = \frac{p(e_s) - p(\omega)p(e_s|\omega)}{1 - p(\omega)} \quad (5.6)$$

where $p(\omega) = \frac{1}{M}$ and $p(e_s) = \frac{1}{L}$ are priors.

It should be noted that the augmentation process is applied only to these users, whose soft biometric traits resemble the claimed ones. Yet, it is important to highlight that the previous frameworks for augmenting biometric recognition with soft biometrics assumed independence between the soft biometrics in an ad-hoc manner, which does not hold per se. On the contrary, the independence between the inserted systematic error for each soft biometric is guaranteed by definition, since the distribution models refer to the measurement errors (not to the soft biometric traits) that are produced from independent measurement processes.

In this context, the goal herein is to find a generic expression of the conditional probability $p(\omega|x_c, e_{s_1}, \dots, e_{s_N})$ that denotes the final recognition score:

$$p(\bar{\omega}|x_c, e_{s_1}, \dots, e_{s_N}) = 1 - p(\omega|x_c, e_{s_1}, \dots, e_{s_N}) \quad (5.7)$$

while according to Bayes' theorem

$$p(\bar{\omega}|x_c, e_{s_1}, e_{s_2}, \dots, e_{s_N}) = \frac{p(x_c, e_{s_1}, e_{s_2}, \dots, e_{s_N}|\bar{\omega})}{p(x_c, e_{s_1}, e_{s_2}, \dots, e_{s_N})} p(\bar{\omega}) \quad (5.8)$$

The nominator can be analyzed as following:

$$\begin{aligned} p(x_c, e_{s_1}, e_{s_2}, \dots, e_{s_N}|\bar{\omega}) &\propto p(x_c, e_{s_1}, e_{s_2}, \dots, e_{s_{N-1}}|\bar{\omega}, e_{s_N})p(e_{s_N}|\bar{\omega}) \\ &= p(x_c, e_{s_1}, e_{s_2}, \dots, e_{s_{N-1}}|\bar{\omega})p(e_{s_N}|\bar{\omega}) \\ &= p(x_c, e_{s_1}, e_{s_2}, \dots, e_{s_{N-2}}|\bar{\omega}, e_{s_{N-1}})p(e_{s_{N-1}}|\bar{\omega})p(e_{s_N}|\bar{\omega}) \\ &= p(x_c, e_{s_1}, e_{s_2}, \dots, e_{s_{N-2}}|\bar{\omega})p(e_{s_{N-1}}|\bar{\omega})p(e_{s_N}|\bar{\omega}) \\ &= p(x_c|\bar{\omega})p(e_{s_1}|\bar{\omega})p(e_{s_2}|\bar{\omega}) \dots p(e_{s_{N-1}}|\bar{\omega})p(e_{s_N}|\bar{\omega}) \end{aligned} \quad (5.9)$$

While regarding the denominator the following holds:

$$p(x_c, e_{s_1}, e_{s_2}, \dots, e_{s_N}) = p(x_c)p(e_{s_1})p(e_{s_2})\dots p(e_{s_N}) \quad (5.10)$$

since by definition the geometric gait signature is uncorrelated to the soft biometric error measurements

$$p(x_c, e_{s_1}, e_{s_2}, \dots, e_{s_N}) = p(x_c)p(e_{s_1}, e_{s_2}, \dots, e_{s_N}) \quad (5.11)$$

and provided that the latter stem from distinct measurement processes and thus variables e_{s_n} are held as *i.i.d.*

$$p(e_{s_1}, e_{s_2}, \dots, e_{s_N}) = p(e_{s_1})p(e_{s_2})\dots p(e_{s_N}) \quad (5.12)$$

In this context, Equation (5.8) can be expressed as the combination of Equation (5.9) with Equation (5.10), as

$$(5.8) = \frac{(5.9)}{(5.10)}p(\bar{\omega}) \quad (5.13)$$

which results to

$$p(\bar{\omega}|x_c, e_{s_1}, e_{s_2}, \dots, e_{s_N}) = \frac{p(e_{s_1}|\bar{\omega})p(e_{s_2}|\bar{\omega})\dots p(e_{s_N}|\bar{\omega})p(x_c|\bar{\omega})}{p(x_c)p(e_{s_1})p(e_{s_2})\dots p(e_{s_N})}p(\bar{\omega}) \quad (5.14)$$

This way, provided that

$$p(\bar{\omega}|x_c) = \frac{p(x_c|\bar{\omega})}{p(x_c)}p(\bar{\omega}) \quad (5.15)$$

Equation (5.14) can be written as following

$$\begin{aligned} p(\bar{\omega}|x_c, e_{s_1}, e_{s_2}, \dots, e_{s_N}) &= \frac{p(e_{s_1}|\bar{\omega})p(e_{s_2}|\bar{\omega})\dots p(e_{s_N}|\bar{\omega})\frac{p(\bar{\omega}|x_c)}{p(\bar{\omega})}p(x_c)}{p(x_c)p(e_{s_1})p(e_{s_2})\dots p(e_{s_N})}p(\bar{\omega}) \\ &= \frac{p(e_{s_1}|\bar{\omega})p(e_{s_2}|\bar{\omega})\dots p(e_{s_N}|\bar{\omega})p(\bar{\omega}|x_c)}{p(e_{s_1})p(e_{s_2})\dots p(e_{s_N})} \\ &= \frac{1}{p(e_{s_1})p(e_{s_2})\dots p(e_{s_N})}p(e_{s_1}|\bar{\omega})p(e_{s_2}|\bar{\omega})\dots p(e_{s_N}|\bar{\omega})p(\bar{\omega}|x_c) \\ &= \frac{1}{p(e_{s_1})p(e_{s_2})\dots p(e_{s_N})} \prod_{n=1}^N p(e_{s_n}|\bar{\omega})p(\bar{\omega}|x_c) \end{aligned} \quad (5.16)$$

Where the term

$$A = \frac{1}{p(e_{s_1})p(e_{s_2})\dots p(e_{s_N})} \quad (5.17)$$

consists of priors and is constant for any users ω . This way,

$$p(\bar{\omega}|x_c, e_{s_1}, e_{s_2}, \dots, e_{s_N}) = A \prod_{n=1}^N p(e_{s_n}|\bar{\omega})p(\bar{\omega}|x_c) \quad (5.18)$$

where the term $f_b = \prod_{n=1}^N p(e_{s_n}|\bar{\omega})$ is the attenuation factor.

5.3. Summary

The work presented in the current chapter is a significant contribution in the field of multi-biometric systems. In particular, the current chapter aims at highlighting the contribution derived by the combination of soft anthropometric traits, that can be captured unobtrusively. Herein, two alternatives are delivered for the enhancement of existing biometric systems via the utilization of anthropometric or soft biometric traits that can be efficiently captured, without requiring any additional sensors and without imposing any significant processing encumbering on top of the baseline biometric system.

In particular, the chapter was divided in two main sections. Initially, it has been attempted to further evaluate the recognition capacity of the features that can be extracted from the upperbody. In particular, the static lengths of the limbs of the upperbody of each user (i.e. the length of the arm sections, the length of the shoulders and the height of the head) are estimated and their recognition potential is evaluated. In particular, an attributed graph matching based methodology has been proposed for the validation of the skeleton lengths of the person requesting authentication.

Although the aforementioned anthropometric characteristics of the upperbody can be easily represented as a fully connected graph, this is not the general case for other soft biometric traits. Moreover, the graph based approach treats the static biometrics as a separate biometric modality, the result of which should be then fused with the baseline biometric system. However, fusion of biometrics is a computationally expensive process, while given the limited recognition capacity of soft biometrics in general in large dataset [222] [66], the final result may not always be in favour of the recog-

dition performance.

For this reason, and a generic probabilistic framework for boosting the client recognition has been presented, utilizing continuous soft biometric traits. In order, to exhibited the general applicability of the p Indicatively, it can be mentioned that the improvements in the given datasets (see Section 3.2.2) are characterized by an improvement of $\sim 2.5\%$ and $> 20\%$ on average for the two aforementioned approaches, respectively.

6. Experimental Evaluation of Prehension Biometrics

The current chapter deals with the experimental evaluation of the methodologies presented in Chapter 4. As it is described below, by utilizing the appropriate datasets, described in Section 3.2.2, the results of the carried out experiments are presented in Section 6.1 and in Section 6.2 for the unimodal and the multimodal case, respectively.

In particular, Section 6.1 starts with the classification of the extracted activity-related features, in terms of their recognition capacity, via the estimation of their relative entropy and their mutual entropy. Next, the section goes on with the evaluation of the recognition performance when several combinations of these features are utilized. The experiments regard two basic scenarios (i.e. activities) from three distinct, but similar, datasets.

Finally, Section 6.2 presents the potential of the proposed prehension based concept by evaluating it in a multimodal scenario with a slightly enhanced -compared to the current state of the art- gait recognition algorithm, as described in Section B.1 in Section B.

6.1. Experimental Results of the Prehension based User Recognition

The application and the evaluation of the approach described in Chapter 4, regarding the reaching and grasping activities, is presented in the current section.

6.1.1. Experimental Setup

An experiment containing two different kinds of everyday movements/activities has been conducted, in order to evaluate the performance of activity related

traits (i.e. movement of the arm and of the fingers) in practice. Namely, each user was instructed to perform both a simple *Reaching and Grasping* activity (i.e. *Interaction with a Microphone Panel*) and a more complex one, a short *Phone Conversation*, as is also presented in the description of the dataset *DB.P.2* in Section 3.2.2.

The complete framework that is proposed herein includes the tracking of the user’s head, arm’s and fingers’ joints via the described equipment (Section 4.1.1 and Section 4.2.1). In particular, the system that has been set up for the execution of the experiments is a three-layered system (i.e. Tracking, Feature Extraction and Decision Taking) and is described in Figure 6.1, exploiting a series of motion-related features for user authentication.

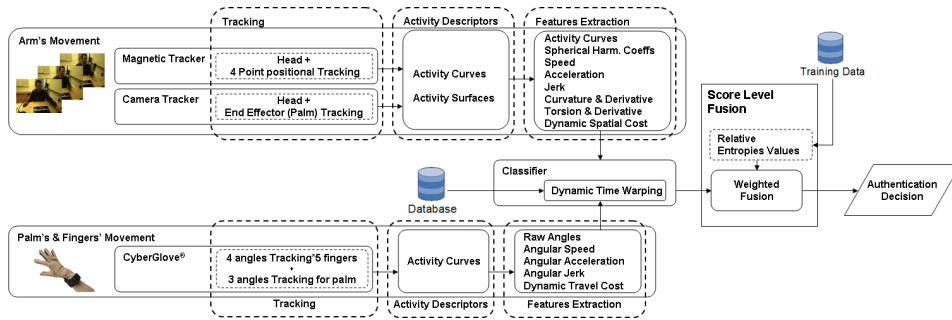


Figure 6.1.: Overview of the proposed system

Following both Hoff’s assumption [189] about the two distinct phases of a prehension movement and the theoretical background provided in Section 4.1 and Section 4.2, a two-fold approach is followed here.

6.1.2. Feature Classification

An analysis was conducted in order to investigate the recognition capacity of the content of the extracted activity-related features (Section 4.1 and Section 4.2), in terms of the evaluating tools presented in Section 4.3.

For the current study, the dataset *DB.P.2* has been utilized and as it is also mentioned in its extended description in Section 3.2.2, the 3 first recorded repetitions of 29 enrolled users performing each movement have been used for the extraction of the features (i.e. in total $3 \times 29 = 87$ signals per movement for each extracted feature) that are classified in the following Section (i.e. Section 6.1.2).

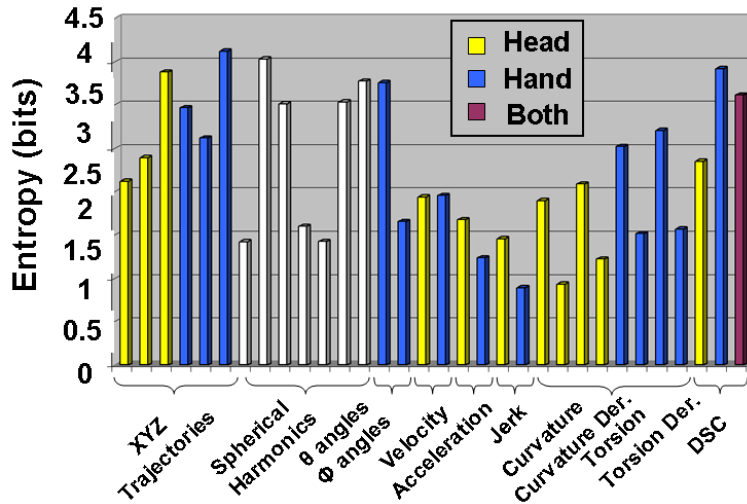


Figure 6.2.: Relative Entropy Values from features extracted by the Camera Tracker for the “Reaching and Grasping” experiment

More specifically, the general intra-individual f_i^{intra} probability distribution was constructed, by using all measurements of feature i from all users. Similarly, the probability distribution f_i^{inter} was constructed, as well, by using all measurements of feature i from the training sessions of a single user, as described in Section 4.3.

The extracted biometric traits were grouped with reference to the tracker used for each experiment. The relative entropy (Section 4.3.1) values are exhibited in Figure 6.2, Figure 6.3, Figure 6.4, Figure 6.5, Figure 6.6 and Figure 6.7, respectively. Regarding what the relative entropy values stand for, it could be mentioned that they refer to the overlapping percentage of the distribution of the values of a subset over the distribution of the values of the whole set, as shown in Eq. (4.27). Thus, the relative entropy values can not only be used as a metric indicating the distinctiveness, but also the recognition capacity of certain biometric variables.

Notable is the high discriminative capacity of both the raw Activities Curves and the most of the 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 4.1.2). Equally interesting is the fact that the spatial cost of the hand is of high

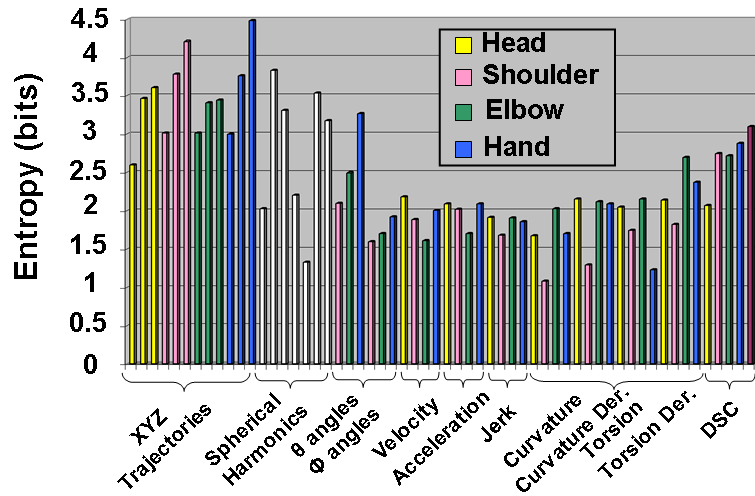


Figure 6.3.: Relative Entropy Values from features extracted by the Magnetic Tracker for the “Reaching and Grasping” experiment

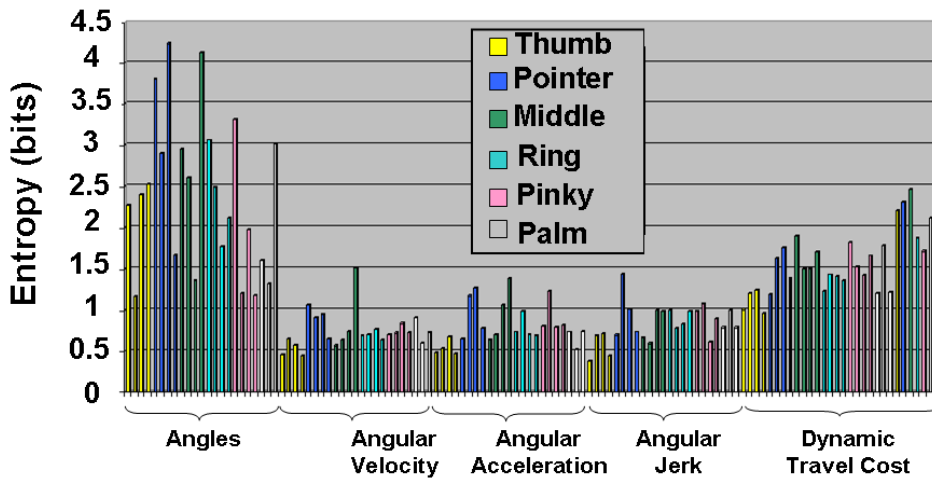


Figure 6.4.: Relative Entropy Values from features extracted by the CyberGlove for the “Reaching and Grasping” experiment

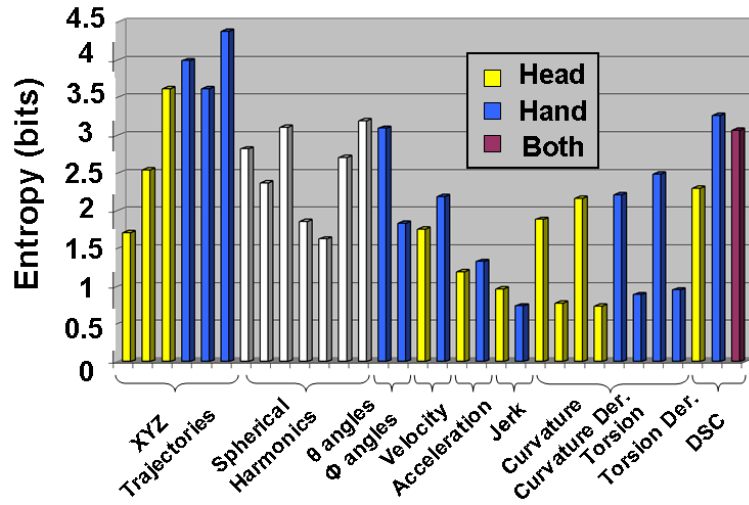


Figure 6.5.: Relative Entropy Values from features extracted by the Camera Tracker for the “Phone Conversation” experiment

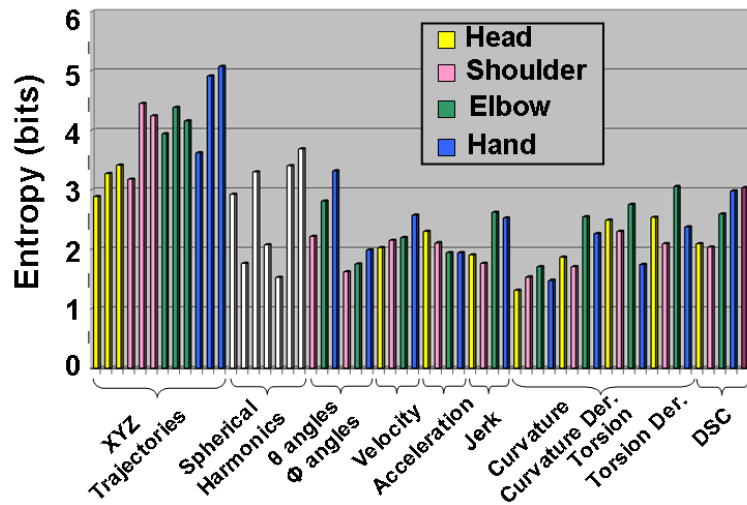


Figure 6.6.: Relative Entropy Values from features extracted by the Magnetic Tracker for the “Phone Conversation” experiment

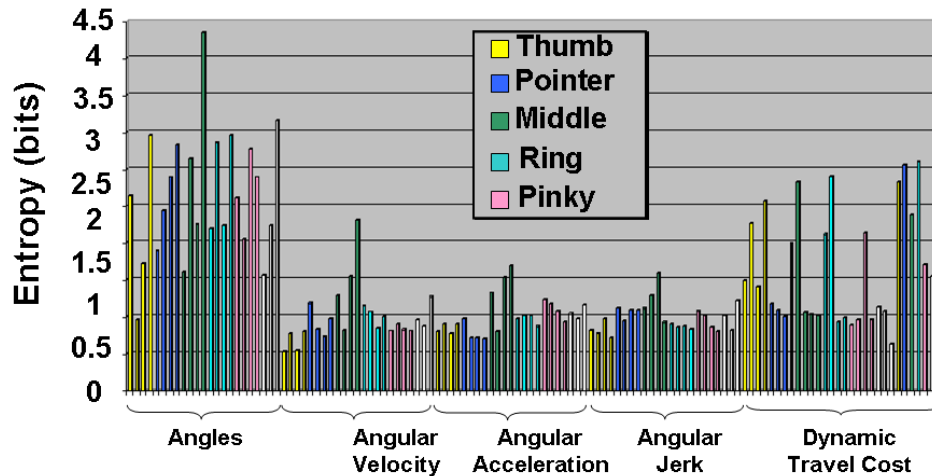


Figure 6.7.: Relative Entropy Values from features extracted by CyberGlove for the “Phone Conversation” experiment

discriminative capacity. Intuitively, it can be claimed that the larger the total spatial cost (i.e. blue DSC feature in Figure 6.2, Figure 6.3, Figure 6.5 and Figure 6.6), the taller (i.e. bigger the arm of the user) 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 phalanxes may provide enhanced distinctiveness among users. The red line stands for the total DTC, summed up over all fingers.

6.1.3. Authentication Capacity of Activity Curves during the Reaching movement

As an initial approach towards the issue of biometric recognition from activity related traits of the movements of the upperbody of the user, the Activity Curves (Section 4.1.2) of the end effector of the arm (i.e. palm), have been utilized (Figure 6.5 and Figure 6.6).

Apart from reasons of simplicity, the selection of this subset of features (i.e. Activity Curves) has been based on their high relative entropy values (Figure 6.2, Figure 6.3, Figure 6.5 and Figure 6.6). It should be noted, that the Activity Curve of the head throughout each movement is only used as reference point (i.e. spatial normalization) for the Activity Curve of the

palm.

Another reason for the utilization of this simplified signature of the user is the evaluation of the enhancement that can be introduced by the incorporation of the static anthropometric model (Section 5.1) and some ergonomics related factors in the recognition process. In particular, the anthropometric model of the user is verified against the one that corresponds to the claimed ID and the results are normalized and fused at a score level, while movements that are characterized by a low ergonomic factor are discarded from the recognition process.

Similarly to the study performed in the previous section (i.e. Section 6.1.2), 3 “gallery”¹ recordings of 29 subjects of the DB.P.1 dataset have been used for the training of an HMM signature template, while 1 recording for each subject as “probe”. The red lines in the ROC curves of Figure 6.8 and Figure 6.9 present the authentication performance, in terms of ROC curves and the corresponding Equal Error Rate (EER) scores, of the “Phone Conversation” and the “Reach & Grasp” (i.e. “Interaction with the Microphone Panel”), respectively.

It should be noted, that given that the proposed approach does not aim at applications for user identification, mainly authentication results will be discussed hereby. In order to compensate for the fact that the ACTIBIO dataset DB.P.1 does not contain anthropometric information, synthetic subjects have been created by merging ACTIBIO dataset DB.P.1 and the 29 subjects from the proprietary Anthropometric dataset DB.A.1.

Following this, the improvements achieved, when incorporating the skeleton-related anthropometric characteristics via the utilization of Attributed Graph Matching (AGM), can be noted in the reduced EER values of the blue and green curves on the same Figures, for the two methods proposed in Section 5.1, respectively. It should be noted that the fusion between the recognition results of the dynamic Activity Curves and the AGM approach has been realized at score level via the utilization of a standard Support Vector Machines (SVM) algorithm, trained on the data acquired from the manual annotation of skeleton on the extra 14 subjects of the Proprietary Anthropometric Dataset DB.A.1.

As expected, when combining both static and dynamic extracted informa-

¹The term “gallery” refer to the set of reference recorded sequences, whereas the term “probe” stands for the test sequences to be verified or identified.

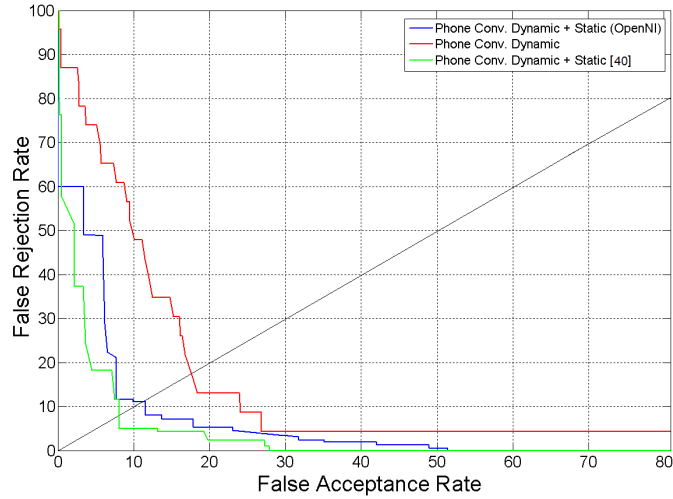


Figure 6.8.: Phone Conversation - ROC Curves for the fused scores.

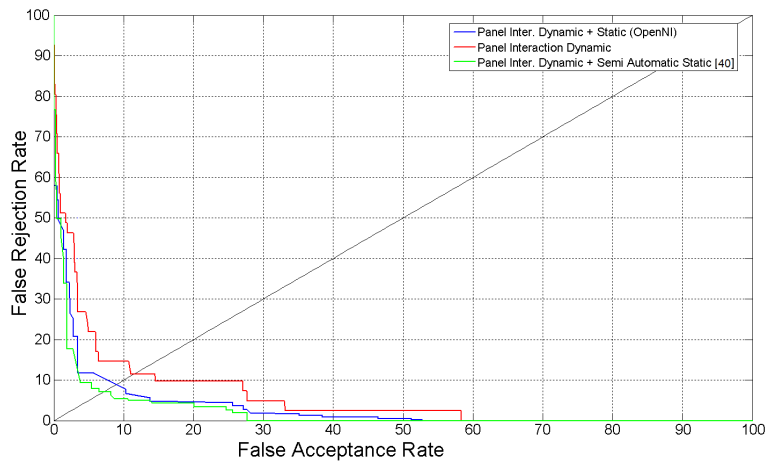


Figure 6.9.: Interaction with the Microphone Panel - ROC Curves for the fused scores.

tion the authentication performance of the system improves further. Specifically, the fusion performed by the SVM achieved an EER score of 8.3% and 7.2%, for the two activities, respectively, when the fully automatic detection was utilized. The EER scores are even lower in the case the semi-automatic particle filtering method has been utilized as shown in Table 6.1.

Moreover, the multicamera environmental setting (i.e. a calibrated frontal stereo-camera with a top monocular camera) of ACTIBIO Dataset DB.P.1 (see Section 3.2.2) allows for the estimation and incorporation of the Er-

gonomic Zones (Section 1.4.1) in the recognition procedure as a quality factor for the recognition capacity of the ongoing movement. In order to calculate the distance $d_{torso,object}$, both the torso and each object have to be first detected on the recording setting. Given that the head position is detected as described in Section 4.1.1, the underlying body part refers to the user's torso. On the other hand, each object can be detected by the top camera as shown in Figure 6.10. Generally, objects are coarsely described in a rotation-invariant way based on their contours (Figure 6.10b). Specifically, each object is described by its aspect ratio, the area it occupies and its colour.

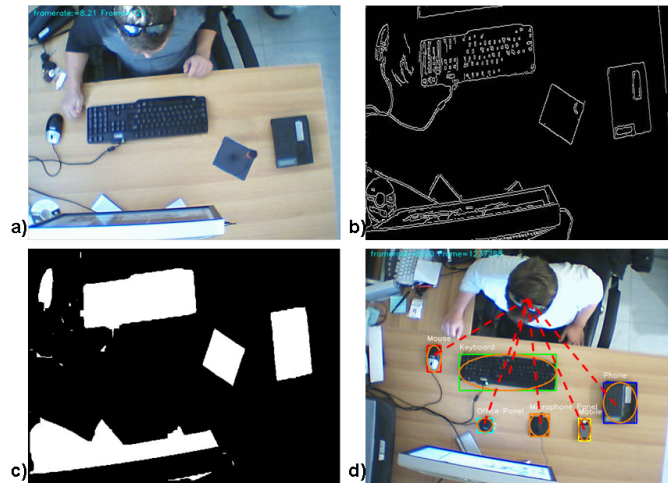


Figure 6.10.: Object detection: a)Top camera view, b)Contour extraction c)Objects' area detection, d)Tagging of objects.

Since the two cameras are calibrated with each other, the distance $d_{torso,object}$ can be easily calculated as illustrated by the red dotted lines shown in Figure 6.10(d).

In this respect, an important metric about the quality and the evaluation of the extracted signature is proposed and is defined in equation (6.1) as the product of the tracking quality factor f_q (equation (6.2)), enhanced by a user-object distance factor b ($0 \leq b \leq 1$), which changes over the human ergonomic spheres (equation (6.3)).

$$f_{q,final} = b \cdot f_q \quad (6.1)$$

$$f_q = 1 - \frac{N_{missHead} + N_{missRHand} + N_{missLHand}}{3N_{frames}} \quad (6.2)$$

where $N_{missHead}$, $N_{missRHand}$, $N_{missLHand}$ are the amount of frames in which the Head, the right and the left Hand were not detected, respectively. N_{frames} is the total number of frames of the sequence.

$$b = \begin{cases} 0.1 \cdot d_{torso,object} + 0.5, & \text{if } d_{torso,object} < 5cm \\ 1, & \text{if } 5cm \leq d_{torso,object} \leq 35cm \\ -0.02 \cdot d_{torso,object} + 1.7, & \text{if } d_{torso,object} > 35cm \end{cases} \quad (6.3)$$

It should be noted that the abstract values presented in this equation refer to a male user of average height (i.e. $1.75m$) and have been adjusted via several trials among a set 17 subjects.

The lower the quality factor the less probable the extracted dynamic features to contain valuable biometric information for authentication. Accordingly, if $f_{q,final} \leq 0.5$ the extracted features are discarded and no authentication process takes place.

The quality factor can be used so as to augment the of the authentication performance of the system in the following manner: “Forced” movements² that include the stretching of the user have been noticed to exhibit inherent deviations from regular ones (i.e. within the convenient zone of the user). Thus, no authentication potential is expected to be found in such movements, given that in the current study the user is expected to act under regular, relaxed conditions, similar to the ones during the enrollment session. In this respect, the quality factor described above contributes to the implicit detection of such movements, in order to be excluded from classification.

In the table below, the reader can notice the improvements in the authentication potential when also the ergonomy-based quality factor is enabled.

The EERs are summarized in Table 6.1, whereby the improvements of the proposed ergonomy-based quality factor are included. The “Dynamic” column refers to the recognition performance of the proposed technique when only dynamic information is used, while the “Static” column refers to the recognition performance exclusively based on the static anthropometric in-

²As “forced” are defined these movements that involve an interaction with an environmental object outside the convenient zone of the user

Table 6.1.: Authentication Performance (EER) of Activity Curves

	Dynamic	Static		Fusion		Fus. & Ergon.	
		[42]	[160]	[42]	[160]	[42]	[160]
Phone Conversation	16.7%	11.3%	13.23%	8.3%	10.8%	7.9%	10.1%
Reach and Grasp (i.e. Interaction with Mic. Panel)	10.32%	11.3%	13.23%	7.2%	9.12%	6.7%	8.4%

formation of the user. The score-level fused results of these aforementioned techniques are presented under the “Fusion” column, while under the last column exhibits the authentication performance when ergonomics restrictions come into play. As it can be noticed by the reader, the ergonomics restrictions ignite a fall in the EER score of 0.6% in both experimental scenarios, as it can be seen from the comparison between the last four columns of the table. Given the size of the testing dataset, such a decrease corresponds to the correct classification of two falsely classified subjects. This improvement stems from the fact that specific repetitions have been excluded from evaluation in the authentication step, since they exhibited low ergonomic confidence. Thus, a reduced false rejection rate has been achieved.

6.1.4. Authentication Capacity of Activity Curves for multiple Authentication

Next, provided the nature of the proposed biometric module, experiments have been carried out regarding the recognition potential of this method in multiple authentication scenarios.

Although the Proprietary Continuous Reaching Dataset (DB.P.4) (see Section 3.2.2) has been originally recorded for experimenting on the affective state of the users, towards the investigation of the effect of stress induction in activity-related user recognition cases, in the current experiment, only the vision-based recordings will be used, so as to evaluate the improvements of multiple authentication when compared to instantaneous one.

In this context, provided that during the game session, 25 individuals (i.e. players) are indirectly forced to perform the same movement (i.e. via pressing the same button on the screen), this information can be used as a

repeating signature. Game related movements, such as pointing gestures, that are drawn along the full area of the screen, cover most of the cases that may come up in game. The movements that are performed in the ergonomic 3D space in the transverse, sagittal and coronal human anatomy planes have been selected, which force him/her to extend to both the corners and the middle of the screen.

Similarly to Section 6.1.3 and without loss of generality, only the Activity Curves (Section 4.1.2) of the head and the right or left arm are taken into consideration. However, extending the experiment in Section 6.1.3, whereby only the position of the palm was taken into account, hereby all the joints of the arm have been utilized (i.e. head, shoulder, elbow, hand), so as to better capture the individual differences in movement (Figure 6.11)

In particular, given the vast majority of extracted movements that are available in this dataset, 5 enrollment sets of trajectories for each of the described movements (see Section 3.2.2) are used for the generation of a signature for each player (i.e. user), while the rest (typically > 10 sets of trajectories) is used as probe, contributing to the multiple authentication scenario, as describe below.

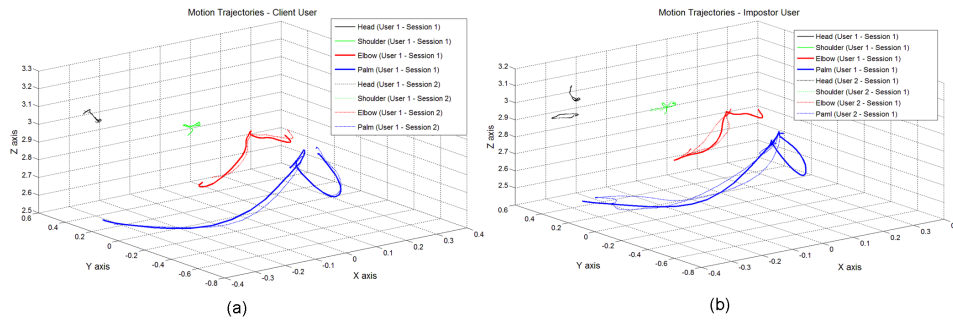


Figure 6.11.: Extracted motion trajectories from the user’s head (black), shoulder (green), elbow (red) and palm (blue) during a specific movement: (a) The intra-similarities in the motion trajectories between different repetitions of the same user are obvious. - (b) The inter-variances in the motion trajectories between the same movement performed by different users are obvious.

At this point, it should be noted that each movement is normalized with reference to the position of the head of the user at the first frame of the movement, so as to retrieve position invariant biometric traits. This way, it

becomes evident that \mathbf{S}_c is a set of 4 normalized state vectors. Given that \mathbf{S}_c exhibits a strong dependence on temporal relations and ordering, it is essential that its processing is performed via the appropriate spatiotemporal means. For this reason, a robust classification tool that is able to efficiently cope with the spatiotemporal information of the signatures can be provided by the Hidden Markov Model (HMM) algorithm.

The authentication performance of the biometric system, when based on the Activity Curves of the players are illustrated in Figure 6.12 for all movements, in terms of ROC curves along with the corresponding EER values (i.e. as indicated by the cross-section of the curves with the diagonal line).

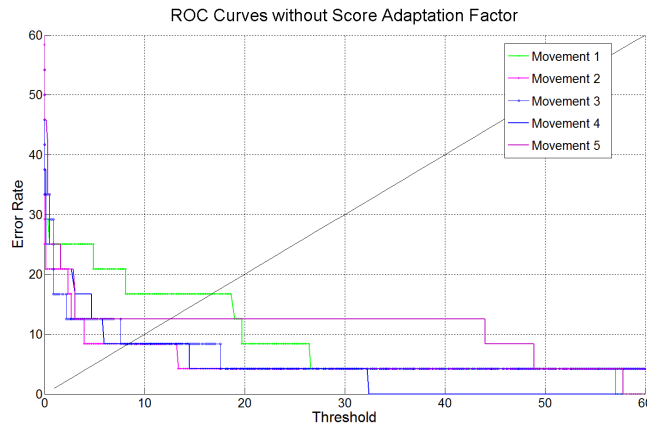


Figure 6.12.: The authentication results for each movement when each user is at the normal stress level.

These EER values are also summed up in Table 6.2 along with the corresponding overall authentication EER score, when the authentication capacity of all movements is taken into account in a weighted average scheme. The weighting factor assigned to the authentication score of each movement is estimated as follows

$$w_i = \frac{EER_i}{\sum_{k=1}^5 EER_k} \quad (6.4)$$

whereby i stands for the label of each movement.

Table 6.2.: EER Scores for each single movement (see Figure 3.4).

Movement ID	1	2	3	4	5	Overall (Eq. 6.4)
Normal State	16.6%	8.6%	8.6%	8.6%	12.5%	7.8%

6.1.5. Authentication Capacity of Spherical Harmonics during the Reaching Movement

The authentication capacity of only the features based on Spherical Harmonics analysis is evaluated. Specifically, provided the high relative entropy values of the group of the Spherical Harmonics as activity-related features (Figure 6.5 and Figure 6.6), extracted from the Activity hyper-Surface (Section 4.1.2), the evaluation of the recognition performance of a system that is explicitly based on them (i.e. Spherical Harmonics) is presented hereafter. For this reason, the “Phone Conversation” of the ACTIBIO dataset DB.P.1 has been utilized. Similarly to the experiment described in Section 6.1.3, herein the 29 subjects of DB.P.1 are utilized. In particular, the training set is formed by 4 gallery sessions, while 1 movement performed by each subject out of the remaining ones is used as the testing data.

In order to provide an overall matching score result between the user requesting access and the corresponding claimed ID, based on the authentication performance of each unity surface, a fusion of these partial matching distances for each RP has to be performed. However, given the general structure of the extracted simplified *Activity Surface* (Section 4.1.2), it is expected that the authentication capacity is higher for some Reference Points (RPs) than for some others (see Table 4.2 in Section 4.1.2 for correspondence) . Thus, the optimal fusion score, that would combine unequally amounts of information from each RP is defined as follows

$$S_{tot} = \sum_{j=1}^N w_j S_j = w_1 S_1 + w_2 S_2 + \dots + w_N S_N \quad (6.5)$$

whereby w_j is the weight coefficient for each of the N RPs and S_j the corresponding partial matching distance.

For the current biometric system, $N = 7$ and the values for each w_j are

defined according to the following equation:

$$w_j = 1 - \frac{EER_j}{\sum_{j=1}^N EER_j} \quad (6.6)$$

where EER_j stands for the Equal Error Rate score for the j^{th} *RP* Spherical Harmonics Coefficients.

Prior to any processing towards fusion, all scores have been normalized according to their corresponding z-score:

$$c_i^z = \frac{c_i^* - \mu_y}{\sigma_y} \quad (6.7)$$

whereby the mean value μ_y and the variance σ_y for each *RP* have been calculated separately.

At this point, it should be noted that the estimation of the weights w_j has been performed, based on different sessions of the ACTIBIO Reaching Dataset (DB.P.1) (see Section 3.2.2), than the ones used for the evaluation of the authentication performance.

As it has already been mentioned, the performance of the proposed system improves the authentication capacity of the system presented in Section 6.1.3. Specifically, the results, acquired by utilizing a Hidden Markov Model (HMM) as classifier (Section 4.4.2), exhibited an Equal Error Rate (EER) score of 15%. Within the current work, this outcome is further verified, by evaluating the authentication capacity of the raw trajectories by a different classifier, based on dynamic programming. Namely, the Dynamic Time Warping (DTW) algorithm (Section 4.4.3) has been implemented and utilized for classifying the probe motion trajectories with respect to the gallery ones. In this case, the EER score has been found at 17.24%.

All reported results are summed in Table 6.3, along with the authentication capacity for each of the proposed Reference Points (RPs), that form the axis-origins for Spherical Harmonics Analysis. As it has been explained, the authentication performance of a single surface mapping with respect to its RP provides moderate authentication performance. However, the fusion of all these results is capable of authenticating a user with much higher robustness and confidence. The interpretation of this can be stated with the fact that although activity surfaces from different users may resemble from one point of view (RP), they will exhibit significant differentiations from

Table 6.3.: Authentication results based on Spherical Harmonics.

	HMM	DTW	SH
RP_1	N/A	N/A	23.07%
RP_2	N/A	N/A	19.56%
RP_3	N/A	N/A	19.17%
RP_4	N/A	N/A	15.38%
RP_5	N/A	N/A	17.24%
RP_6	N/A	N/A	16.51%
RP_7	N/A	N/A	15.64%
Overall/Fusion Score	15%	17.24 %	11.76%

almost any other RP. On the contrary, activity surfaces derived from same users would exhibit increased average similarity for all RPs.

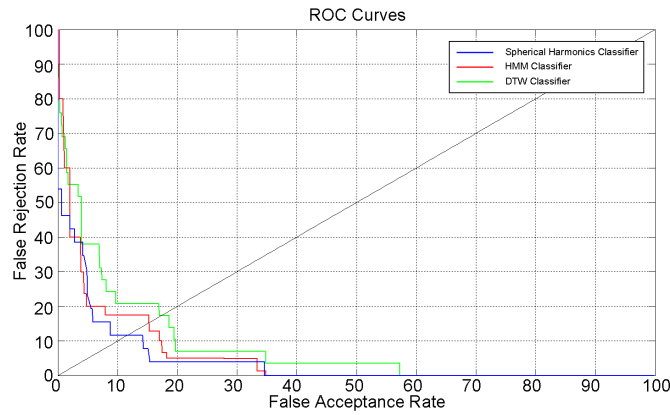


Figure 6.13.: ROC Curves comparing the authentication performance of the proposed SH method and SoA.

The reader can notice that the proposed surface descriptor is superior to the previously proposed methods, both in terms of authentication performance (Figure 6.13), but also in terms of invariance, given that the spherical harmonics analysis is by definition rotationally invariant. This would mean that the authentication performance remains unaffected by the angle the camera is turned with respect to the user. The provided view-invariance stems directly from the inherent property of Spherical Harmonics Analysis for rotational invariance and is the main reason for improving the authentication performance.

6.1.6. Authentication Capacity of the full Prehension movement

The full signature, as formed after the application of the proposed feature selection method (Section 4.3), is herein evaluated in terms of recognition performance. Similarly to Section 6.1.2, the dataset, on which the current study has been based, contains the 29 subjects of the DB.P.2. Thereby, 3 sessions have been used as training data, while the 4th and the 5th recording are used as the testing data in the present and at a later time session (see the full description of the current dataset in Section 3.2.2), respectively, so as to provide an estimation of the performance potential of the proposed approach over time.

Initially, the findings of the results of algorithm are presented. Moreover, the authentication capacity of the features extracted from the movement of the fingers are also exhibited, following a similar approach for feature selection.

Thus, following the findings of Section 6.1.2, it should be noted that among the most indicative features, there may still be redundancy, given that it is very likely that some feature 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, a representation via Confusion Matrices would be meaningless. However, 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 [191] 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) [193], etc.) is utilized herein, taking into account both the Kullback-Leibner 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 & 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 (Figure 6.14, Figure 6.15 and Figure 6.16), starting from 1 to the total number of extracted features $N_{movement,tracker}$, with respect to the tracker and the movement studied. Based on a confusion matrix including the mutual entropies, the $n_{movement,tracker}^i$ features are preserved 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. $n_{movement,tracker}^{i+1} = n_{movement,tracker}^i + 1$). 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 4.4). This way an EER score is estimated in the testing dataset and noted down, while the algorithm proceeds to the next iteration.

It has been noted that after a certain number $n_{movement,tracker}$ of utilized features, the authentication performance decreases (i.e. EER score increases accordingly in Figure 6.14, Figure 6.15 and Figure 6.16), since the use of less distinctive features has a negative effect to the performance of the system. Thus, the features, that are indicated as unimportant/redundant, will be discarded from the authentication procedure, since their utilization has a negative contribution. This way, features with high mutual information values with others can be discarded, without serious loss in the overall discrimination capacity of the system.

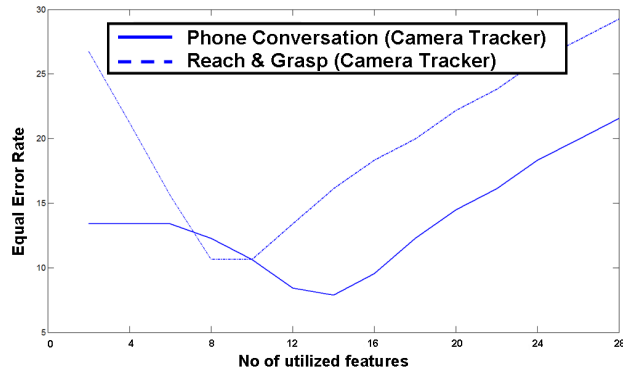


Figure 6.14.: The EER value as a function of the utilized features applied in decreasing order of relative entropy for the Camera Tracker.

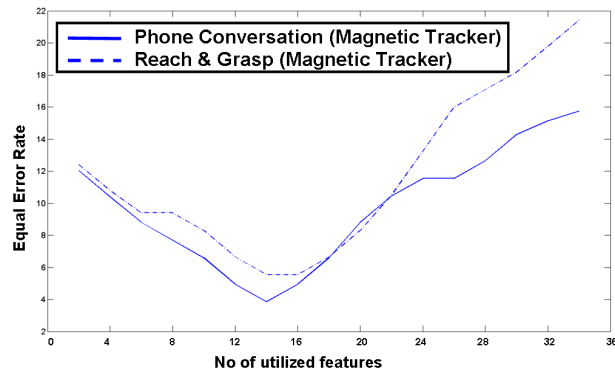


Figure 6.15.: The EER value as a function of the utilized features, applied in decreasing order of relative entropy for the Magnetic Tracker.

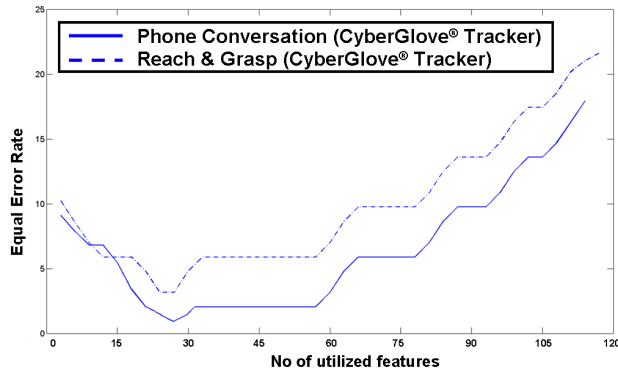


Figure 6.16.: The EER value as a function of the utilized features applied in decreasing order of relative entropy for the CyberGlove.

In this respect, Table 6.5 on page 161 includes the features per activity that are maintained, as the most valuable ones. The feature names mentioned in this table correspond to Figure 6.2, Figure 6.3, Figure 6.4, Figure 6.5, Figure 6.6 and Figure 6.7, while the notations for fingers of the human are analytically illustrated in 4.13(b) and shortly presented in Figure 4.4.

Additionally in Figure 6.14, 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.

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. The aforementioned findings regarding the optimal number of preserved, most discriminative features, were utilized herein and the results are shown with the help of ROC Curves (Figure 6.17, Figure 6.18 and Figure 6.19). 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.

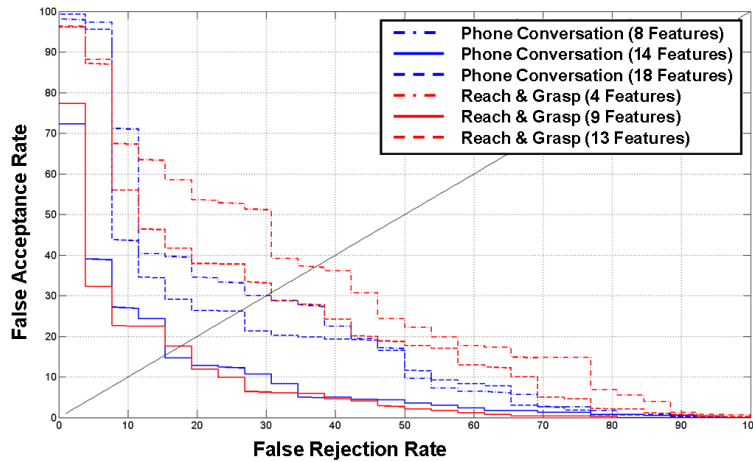


Figure 6.17.: ROC Curves for the performed activities in Session 5 as recorded by the Camera Tracker.

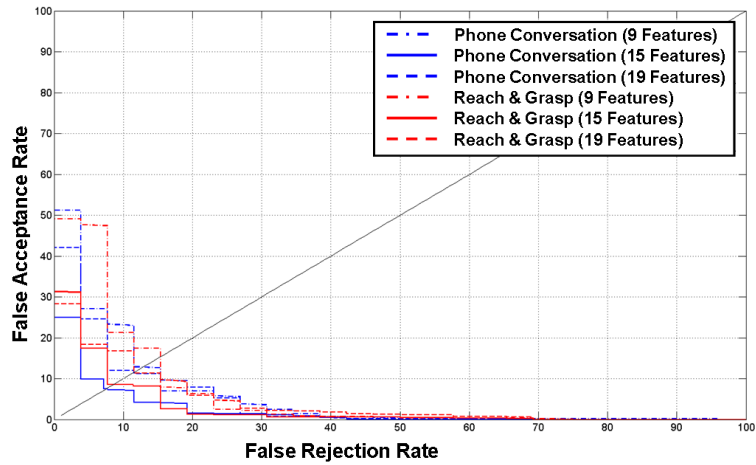


Figure 6.18.: ROC Curves for the performed activities in Session 5 as recorded by the Magnetic Tracker.

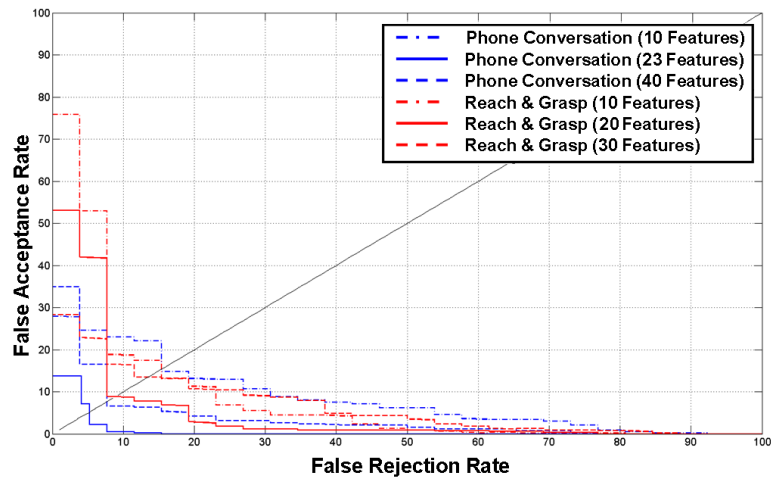


Figure 6.19.: ROC Curves for the performed activities in Session 5 as recorded by the CyberGlove Tracker.

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 6.4), as expected.

6.1.7. Experimental results in a realistic environment with the ACTIBIO Database

The findings regarding the most indicative features for authentication purposes (Section 6.1.6) were applied to the real environment of the ACTIBIO database *DB.P.1*. Although a detailed description of this dataset is included in Section 3.2.2, it should be noted at this point, that the dataset contains 29 subjects, whereby 3 repetitions of the movements have been utilized as training data for each user, while a 4th one was used as probe data. Unfortunately, however, the ACTIBIO database does not include measurements with Cyberglove. Thus, the evaluation of our framework was based only on the proposed camera tracker (Section 4.1.1).

Following this, only the most significant features, (see Table 6.5 in Section 6.1.2), were extracted. Figure 6.20 illustrates the variations of the system’s performance in terms of EER scores for different numbers of the most indicative features (Table 6.5).

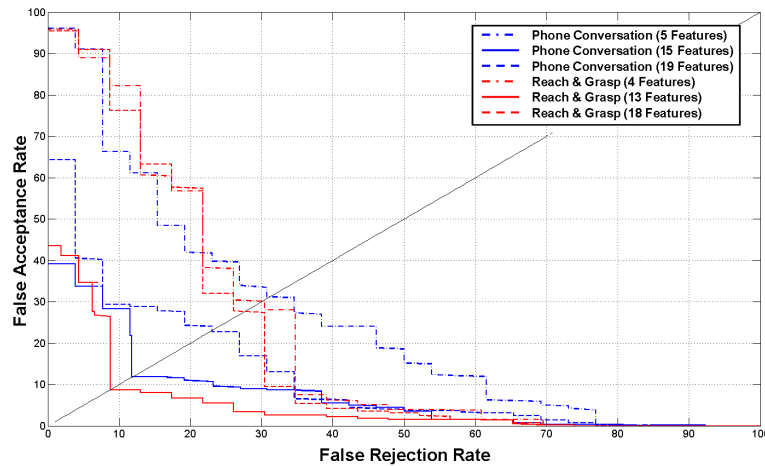


Figure 6.20.: ROC Curves for the performed activities in the ACTIBIO Database, as they were recorded by the Camera Tracker

6.1.8. Experimental results in a large synthetic Database

Finally, the proposed approach (Section 6.1.6) was evaluated in the large Virtual Prehension Dataset (DB.P.3), so as to verify the validity of the proposed method in a large scale dataset.

Following a similar setup as the experiments performed in the previous two dataset (i.e. ACTIBIO database *DB.P.1* and Proprietary Prehension Dataset (*DB.P.2*)), the training data set included 100 virtual users (see Section 3.2.2 for a detailed description regarding the generation of the traits of these users), whereby 4 virtual repetitions have been used for training the virtual biometric signatures, while a extra repetition has been used as the testing/probe set.

In this respect, the changes of the EER scores for different number of utilized features are presented in Figure 6.21 and Figure 6.22 for the two proposed activities.

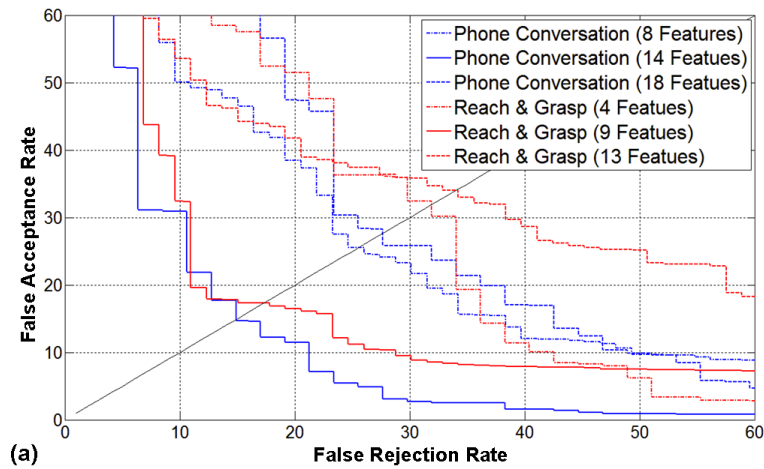


Figure 6.21.: ROC Curves for the performed activities in the synthetic Database as recorded by the Camera tracking device.

It should be noted that the results from the synthetic database should be compared with the ones of Session 2 (Figure 6.17, Figure 6.18 and Figure 6.19), given that the generated trajectories for the virtual subjects were based on statistics from all sessions.

Finally, the fused results of the proposed framework are presented in Table 6.4. It should be noted that the fusion was performed as described at the end of Section 4.4.

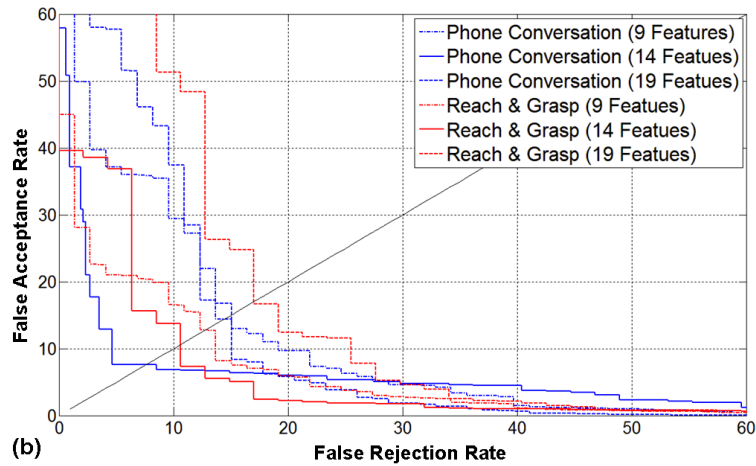


Figure 6.22.: ROC Curves for the performed activities in the synthetic Database as recorded by the Magnetic tracking device.

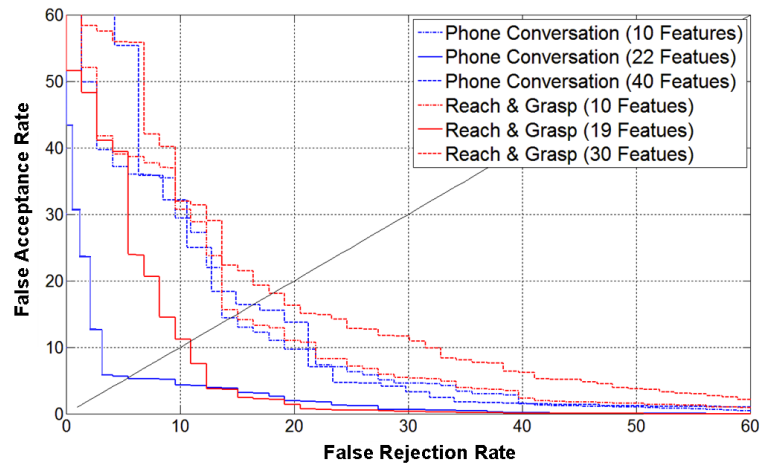


Figure 6.23.: ROC Curves for the performed activities in the synthetic Database as recorded by the CyberGlove tracking device.

Table 6.4.: Overall Authentication Errors after final Fusion

	Phone Conversation			Reaching & Grasping		
	<i>Session 1</i>	<i>Session 2</i>	<i>Virtual Subjects</i>	<i>Session 1</i>	<i>Session 2</i>	<i>Virtual Subjects</i>
Camera Tracker & Cyberglove	1.1%	3.9%	4.7%	6.8%	7.7%	8.2%
Magnetic Tracker & Cyberglove	0.8%	3.3%	3.6%	2.6%	6.2%	6.4%

Table 6.5.: Remaining - Most valuable Features per Tracking Device

	Camera Tracker	Magnetic Tracker	Cyberglove
Initially extracted features	<p>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</p> <p>Phone Conversation Y,Z Trajectories for Head; X,Y,Z Trajectories for Hand; 5 Spherical Harmonics; θ angle for orientation; DSC for Hand; Hand velocity; Hand torsion.</p> <p>Reach & Grasp 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.</p>	<p>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</p> <p>Phone Conversation 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.</p> <p>Reach & Grasp 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; 1^{st} Derivative of Palm torsion.</p>	<p>23 for the phalanxes angles, 69 the angular velocity, angular acceleration and angular jerk and 23+6+1 for the dynamic travel</p> <p>Phone Conversation Finger a: W_0, W_2 and W_3 angles; W_2 total travel cost. Finger b: W_4, W_5 and W_6 for angles, W_5 jerk, total travel cost; Finger c: W_9 and W_{10} for angles, W_{10} for velocity, total travel cost; Finger d: W_{12}, W_{13}, W_{14} for angles, W_{13} and total travel cost; Finger e: -; Palm: W_{19} for angles (i.e. transverse palm movement), W_{19} travel cost.</p> <p>Reach & Grasp 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 travel cost; Finger e: -; Palm: W_{19} and W_{20} for angles (i.e. transverse and sagittal palm movement).</p>
Finally preserved features	<p>Phone Conversation Y,Z Trajectories for Head; X,Y,Z Trajectories for Hand; 5 Spherical Harmonics; θ angle for orientation; DSC for Hand; Hand velocity; Hand torsion.</p> <p>Reach & Grasp 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.</p>	<p>Phone Conversation 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.</p> <p>Reach & Grasp 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; 1^{st} Derivative of Palm torsion.</p>	<p>Phone Conversation Finger a: W_0, W_2 and W_3 angles; W_2 total travel cost. Finger b: W_4, W_5 and W_6 for angles, W_5 jerk, total travel cost; Finger c: W_9, W_{10} and W_{11} for angles, W_{10} for velocity, total travel cost; Finger d: W_{12}, W_{13}, W_{14} for angles, W_{13} and total travel cost; Finger e: -; Palm: W_{19} for angles (i.e. transverse palm movement), W_{19} travel cost.</p> <p>Reach & Grasp 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 travel cost; Finger e: -; Palm: W_{19} and W_{20} for angles (i.e. transverse and sagittal palm movement).</p>

6.2. Experimental results of Prehension based biometrics in a multimodal approach

In order to evaluate the effectiveness of the proposed Prehension biometric feature (Section 4.1) in multimodal systems, an experiment has been set up, utilizing the concept of the so-called “on-the-move” biometry [90], which sets as the final objective the non-stop authentication in a very unobtrusive and transparent manner, where the user is not requested to perform any special action. In particular, a novel scheme for the integration of two activity-related traits in a multimodal biometric recognition system (i.e. prehension and gait based recognition) is presented, using a score level fusion of the individual modalities. The selected modalities are chosen so as to satisfy the unobtrusiveness constraints of the framework.

Regarding the activity related biometric module, the activities of the user are captured and the corresponding Activity Curves, as described in Section 4.1.2 are extracted (Figure 6.24). Finally, they are used as input to a Hidden Markov Model (HMM) algorithm, both for training and for classification (Section 4.4.2).

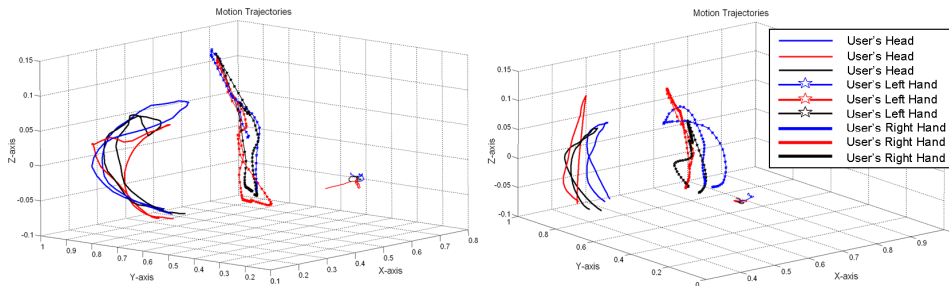


Figure 6.24.: Extracted motion trajectories from (a) User 1 and (b) User 2 during the combined movement of *Inserting a Card* and *Typing a pinword*.

In parallel, a feature-based gait recognition system is proposed that can handle realistic events, such as user stops (Section B.3 and random walking paths (Section B.4. Thus, the gait system can be adopted for environments, where the user can freely move within the working space and perform everyday activities.

A detailed description of the utilized gait recognition algorithm is provided in Section B. However, for reasons of consistency and for the con-

venience of the reader, a short description is included hereafter. Following the architecture illustrated in the upper part of the building blocks diagram in Figure 6.25, the reader can notice that initially the user’s silhouette is extracted from each frame (i.e. every colour frame is accompanied by the corresponding depth information of the recorded scene) individually. Once a gait cycle is detected, all silhouettes within it are collected and processed, so as to form the so-called Gait Energy Image (GEI) [246]. It should be noted that any silhouette depicting a non-walking user (i.e. stop detection) has been removed from the frame sequence of the gait cycle, as proposed in Section B.3, while any silhouette that deviates from the fronto-parallel walking direction, with respect to the camera, as proposed in Section B.4. Finally, three well known 2D transforms (i.e. the Radial Integration, the Circular Integration and the Krawtchouk Moments transforms) are applied on the GEI, as described in Section B.1 and their outcome forms the user signature.

The fusion between the two modalities is performed at the score level optimally and is parameterized via a Genetic Algorithm (see Section B.5 of Section B). The architecture of the proposed biometric recognition framework is illustrated in Figure (6.25).

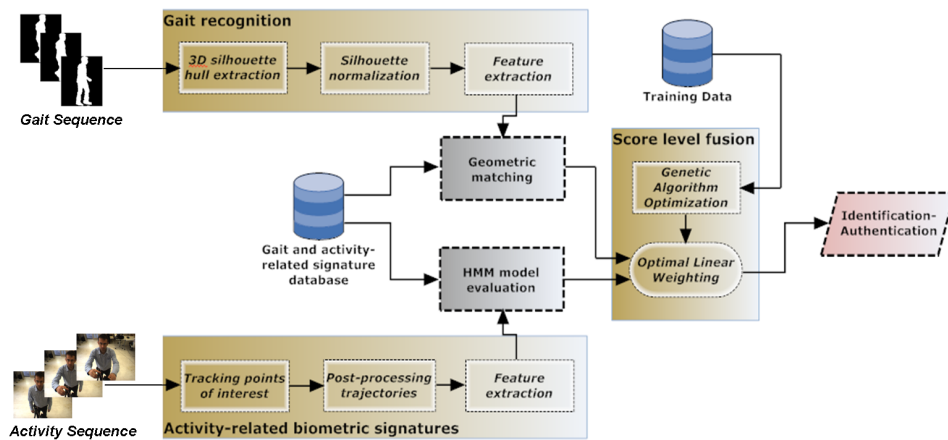


Figure 6.25.: Architecture of the proposed gait recognition framework.

The application scenario expects that the user walks along a corridor in arbitrary walking paths. In the middle of the path, there exists a control panel, where the user is supposed to stop, in order to insert his authorization card and to type his personal pin. Then, the user continues his way to the

door at the other end of the corridor. The whole scene is constantly recorded by two stereoscopic cameras.

6.2.1. Gait Recognition results

The aforementioned scenario has been conducted, based on the combination of the datasets ACTIBIO Gait Dataset (DB.G.2), including 29 subjects, and the Proprietary Gait Dataset (DB.P.G.3), including 14 subjects, regarding the gait recordings and the ACTIBIO Reaching Dataset (DB.P.1) and Proprietary Activity & Gait Dataset (DB.P.G.3) again with 29 and 14 subjects, respectively, regarding the Activity-related recordings. Despite the high amount of recordings included in both gait datasets (see detailed description in Section 3.2.2), only two recordings of the “normal” walking repetitions are used for the construction of the training set. Following this, the probe set contains 1 recording from each of the “normal” and the “view-stop” scenario.

The pixel-wise differences in the extracted Gait Energy Images (*GEI*), as described in Section B.1 of Section B, between the non-stop-detection approach and the proposed framework are demonstrated in Figure 6.26. The reader can notice the significant denaturation of the *GEI* image in the absence of the stop detection, due to the contribution of those frames, whereby the user has been standing still.



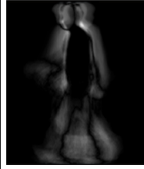


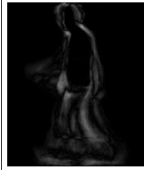
	Gallery GEI	Probe GEI	Absolute Difference	Difference Percentage
Without Stop Detection				6.50%
With Stop Detection				3.59%

Figure 6.26.: 1st row: Great variations between the gallery and the probe even between a client user, when stop detection is disabled. - 2nd row: Low denaturation rate of the probe GEI, when stop detection is enabled at the probe sample.

The improvements of the proposed gait recognition modality (i.e. stop detection & silhouette rotation) when the *RIT* and *KRM* algorithms are utilized as classifiers can be seen in Figure 6.27.

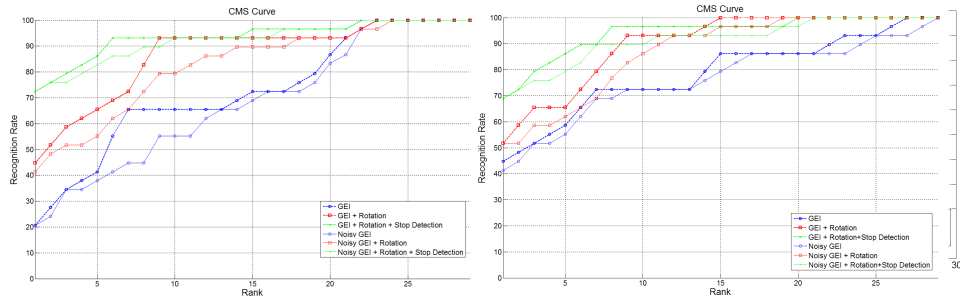


Figure 6.27.: Improvements in Gait due to silhouette rotation & stop detection algorithm (29 Subjects) - Left: (*RIT* classifier) / Right: (*KRM* classifier).

Specifically, the reader can notice the significant contribution of the rotation algorithm to the method proposed in [178]. In particular, the identification rates (red line in Figure 6.27) are increased by a mean ratio of 20% (peek ratio improvement 35%) in the case of the *RIT*-classifier. Similarly, as far as *KRM* features are concerned, an improvement of a mean ratio of 10% (peek ratio improvement 23%) can be observed. In addition, when stop detection algorithm was enabled, the identification rates increased even more by a mean ratio value of 25% and 20%, in both the *RIT* and *KRM* classifier cases, respectively.

The proposed algorithms have been also tested in terms of their resistance against noise. For this reason, Additional White Gaussian Noise (*AWGN*) with a Peak Signal-to-Noise Ratio (*PSNR*) of 24.1237dB has been added to the extracted gait silhouettes, by the successive down-scaling to 25% of the original size of their resolution and up-scaling them back [226], prior to the generation of each *GEI* (see Figure 6.28). The derived results (Figure 6.27) caused only a small degradation in the module's recognition performance, which proved the robustness of the proposed approach under noisy environments.

In the same respect, the proposed enhancements exhibit significant improvements regarding the authentication performance of the gait module, as indicated by the EER results in Table 6.6. Similarly, the degradation

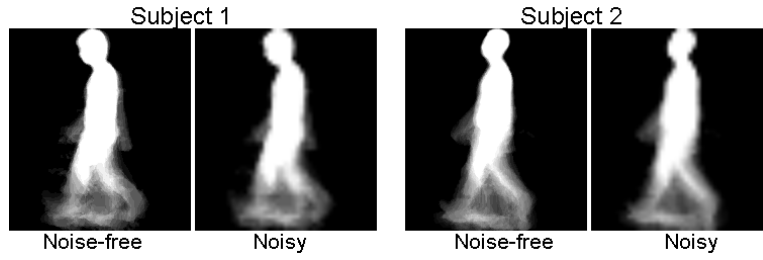


Figure 6.28.: a) Noise free vS. Noisy ($PSNR = 24.1237dB$) Silhouettes.

caused by noise insertion can be considered rather low.

Table 6.6.: Activity (ACTIBIO Dataset) - Equal Error Rates

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

6.2.2. Activity-Related recognition results

For the Activity-related part of the experiment, the 29-subject ACTIBIO Reaching Dataset (DB.P.1) and the 14-subject Proprietary Activity & Gait Dataset (DB.P.G.3) have been utilized. In particular, for the evaluation of the recognition performance on both datasets, 3 recordings have been used for the construction of the signature for each user, while 1 - different from the one utilized for the training of the genetic algorithm - was used as probe.

In this respect, the recognition performance, as well as in the Equal Error Rate (EER) score are exhibited in Table 6.7.

Table 6.7.: Activity (ACTIBIO Dataset) - Recognition Performance

	Identification (CMS)		Authentication
	Rank-1	Rank-5	Equal Error Rate
DB.P.1 - 29 Subjects	67%	96%	14.7%

6.2.3. Fusion results

The score level fusion between the three classification approaches (i.e. RIT (Section B.1), KRM (Section B.1), and HMM (Section 4.4.2)) is performed, as described in Section 6.2.3. The recognition and verification performance of the final improved multimodal system, as they have been derived from tests carried out on the 29-subject ACTIBIO database, can be seen in Figure 6.29a and in Table 6.8, respectively.

Specifically, the significantly increased Rank-1 identification rate of the multimodal system has reached a score of 83%, while at Rank-5 the identification rate has correctly recognized all the users. Additionally, the score-level combination of the two activity related traits (i.e. trajectory based activity recognition & gait recognition) has managed to decrease the overall EER score of the system to 9% as indicated in the last column of the Table 6.8. In the same respect, the system exhibited strong resilience in both authentication and identification performance, even during the “noisy” experiment, as shown in the Figure 6.29 and in Table 6.8.

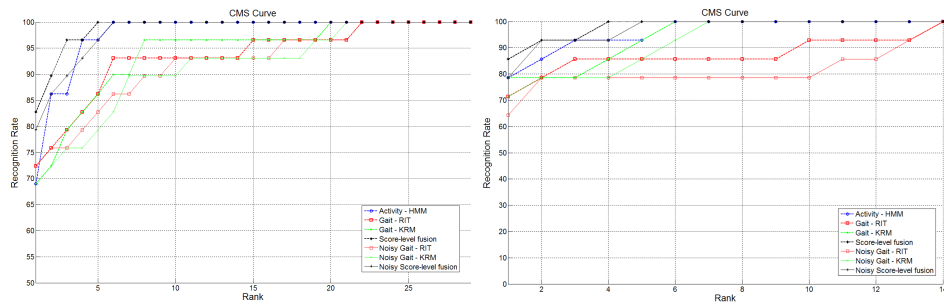


Figure 6.29.: CMS Diagram of the final multimodal system - Left: (29-Subjects dataset) / Right: (14-Subjects dataset).

Similarly, the corresponding CMS curves and the EER scores for the custom dataset including 14-subject are depicted in Figure 6.29b, and along the second row of Table 6.8 respectively.

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

Table 6.8.: Multimodal System (ACTIBIO Dataset) - Equal Error Rates

	RIT	KRM	HMM	Fusion	Fusion (noise)
EER (29 Subjects)	15.9%	16.5%	14.7%	10.4%	11.3%
EER (14 Subjects)	14.3%	17.2%	12.1%	8.9%	9.7%

Fusion results

The fusion methodology that is proposed herein has its basis on the Genetic Algorithm (GA) initially presented in [245] and requires the normalization of the derived scores to a common basis. For this reason the following normalization formula has been utilized:

$$d^{norm} = \left(\frac{0.5}{T_L}\right)e^{-\frac{d}{d^{max}}} \quad (6.8)$$

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

In general, a GA is selected to do the fusion between the proposed biometric modalities in the cases that there is absence of a priori knowledge regarding the distribution of the estimated similarity scores. GAs are very efficient optimization methods, since they are capable of detecting near global optimum solutions without the need of a priori knowledge of the premise space and of any non-convexities within it. Thus, in order to optimize the performance of the multimodal gait biometric system and supplementary fuse the activity-related biometric traits, the genetic algorithm described in [245] is estimating the optimal weights for the three biometric descriptors.

At this point, it should be noted that the utilized genetic algorithm has been trained via the the 14-subject DB.P.G.3 dataset, but using different repetitions than the ones used for the aforementioned authentication/identification experiments.

In particular, the optimal weights used for score fusion based on a simple weighted averaging scheme are estimated using the genetic algorithm described in Section B.5 of Section B. For the training of the fusion algorithm, the recording of the 14 subjects contained in the Proprietary Gait Dataset

(DB.G.3) (see Section 3.2.2) has been utilized. Specifically, the used gallery and the corresponding probe sequences stem from different repetitions than the ones that have later been used for the actual recognition purposes.

The experimental tests with the aforementioned genetic algorithm (Section B.5) resulted in the following optimal weighted values:

$$w_{RIT} = 0.34075, w_{KRM} = 0.21425, w_{HMM} = 0.445 \quad (6.9)$$

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

$$Sim(x, y) = \sum_{n \in T} \frac{w_n}{D_n} = \frac{w_{RIT}}{D_{RIT}(x, y)} + \frac{w_{KRM}}{D_{KRM}(x, y)} + \frac{w_{HMM}}{D_{HMM}(x, y)} \quad (6.10)$$

whereby x ranges from 1 to N_P number of probes to identify, y denotes all the subjects in the training database $y = \{1, \dots, N_G\}$ and $D_n(x, y) = 1/Sim_n(x, y)$ denotes the total dissimilarity, between the probe subject x and the gallery subject y given the feature set $n \in E^{full}$, where $E^{full} = \{RIT, KRM, HMM\}$.

The proposed fusion method is only used to estimate the optimal weights once and then the trained algorithm is applied as is, for the online identification of individuals with no further training or altering of the weights. The final results are presented in Table 6.8, whereby a noticeable improvement in the

7. Evaluation Static

Anthropometric Trait related Enhancements

In this chapter follows the experimental evaluation of the methodologies presented in Chapter 5. By utilizing the appropriate datasets of Section 3.2.2, the results of the carried out experiments per case follow in Section 7.1 for the enhancement in performance accuracy via anthropometric and soft biometric traits, respectively. In particular, the algorithms presented in Section 5.2 are significantly augmenting biometric systems (i.e. gait and face recognition systems) of the current state of the art, as shown in Section 7.1.

7.1. Experimental results of the Soft Biometrics based enhancements

The application and the evaluation of the probabilistic framework described in Section 5.2, regarding the enhancement of biometric systems via the incorporation of continuous soft biometric traits, is presented hereafter. In particular, in order to verify the generic application of the proposed approach, two experiments have been conducted. The first one refers to the enhancement of a well known gait related biometric approach that has been initially presented in [178], while the second one regards a state of the art method for pose invariant *3D* face recognition, proposed by Beretti et al. in [251].

7.1.1. Enhancing Gait Recognition with Soft Biometrics

In order to carry out the evaluation of the probabilistic framework, proposed in Section 5.2, on top of a known gait recognition system, two similar

gait related datasets have been utilized, i.e. the HUMABIO Gait Dataset (DB.G.1) and the ACTIBIO Gait Dataset (DB.G.2). In particular, the first one includes the recordings of 75 subjects walking in several scenarios and in different time sessions, as described in Section 3.2.2. Similarly, the second dataset includes 29 subjects walking in front of the camera under an equivalent variety of scenarios. In any case, however, the training set is formed by 2 recordings for each subject, when he/she is performing the “normal” fronto-parallel walking scenario, as it is described in Section 3.2.2, while for the testing set 1 different recording is used per case, as shown later in the current Section. Similarly, the time-related experiment regards the comparison of an incoming gait signature, as recorded at a later time than the registration period, with the user’s template in the gallery dataset.

The application of the framework of Section 5.2 in a gait recognition scenario requires first of all the development of a geometric gait recognition algorithm. Moreover, the height and stride length soft biometric features should be extracted. Finally, the feature space has to be partitioned and the probabilities $p(e_s|\omega)$ (Equation (5.5)) have to be modelled, as indicated in Section 5.2.1 and Section 5.2.1, respectively, for each soft biometric trait s .

Herein, the feature vector x_c (see Section 5.2.1), that refers to the dynamic gait features (i.e. hard biometric), is extracted using two gait recognition algorithms. The first algorithm is presented in [178] and is based on the two well know Radon Transforms that are applied to gait sequence silhouettes (i.e. *BS-RIT* and *BS-CIT*). The second algorithm is based on matching spatiotemporal descriptors of the human gait, the so-called Gait Energy Images [246] and is based on matching spatiotemporal images of human gait (i.e. *GEI-RIT* and *GEI-CIT*). In order to make the current thesis self-contained, a short description of the aforementioned algorithms are described in Section B.1.

Moreover, the “height” and “stride length” soft biometric features should be extracted. This is trivially achieved from the stereoscopic gait sequences, as the highest-lowest part of the subject and to the largest distance between the legs within a gait cycle, respectively. It becomes evident that the process followed for the estimation of the stride length is prone to bad illumination and the corresponding shadows that are created on the walking floor, which may have as result the occlusion of the edges of the feet. On the other

hand, since the height is estimated as the mean height value of all recorded frames is robust to illumination changes along the walking path. Thus, the errors in measurements mainly stem from possible variations in the types of shoes/hills worn by the users and from the natural hopping of humans during walking. A more detailed analysis for the estimation of “height” and “stride” can also be found in Section B.2.

The claim that the error measurements of the utilized soft biometrics are i.i.d. with each other is based on the fact that they stem from independent measurement processes, as described in [178].

Thereby, the process followed for the estimation of the stride length is prone to bad illumination and the corresponding shadows that are created on the walking floor, which may have as result the occlusion of the edges of the feet, as shown in Figure 7.1.

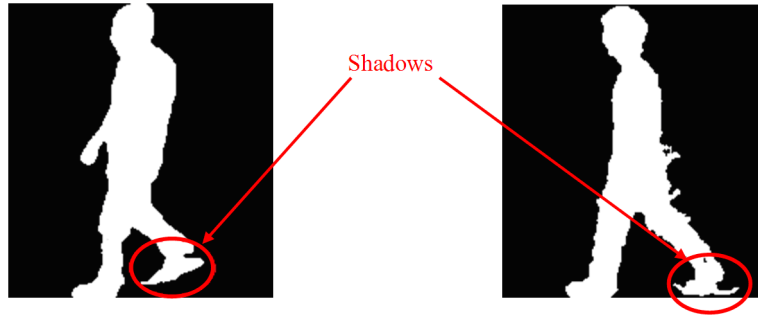


Figure 7.1.: The estimation of the stride length is very sensitive to the shadows that are created on the floor walking level of the user’s walking path.

On the other hand, since the height is estimated as the mean height value of all recorded frames and is thus, robust to illumination changes along the walking path. In this case, the measurement error mainly stems from possible variations in the types of shoes/hills worn by the users, as well as due to the natural hopping of humans during walking.

The aforementioned claim can be easily verified by the following diagrams, which prove that in our experiments the measurement errors between the utilized soft biometrics can be held as i.i.d. variables. In particular, the correlation factor between the calculated error values (see Figure 7.2) was found 0.0755.

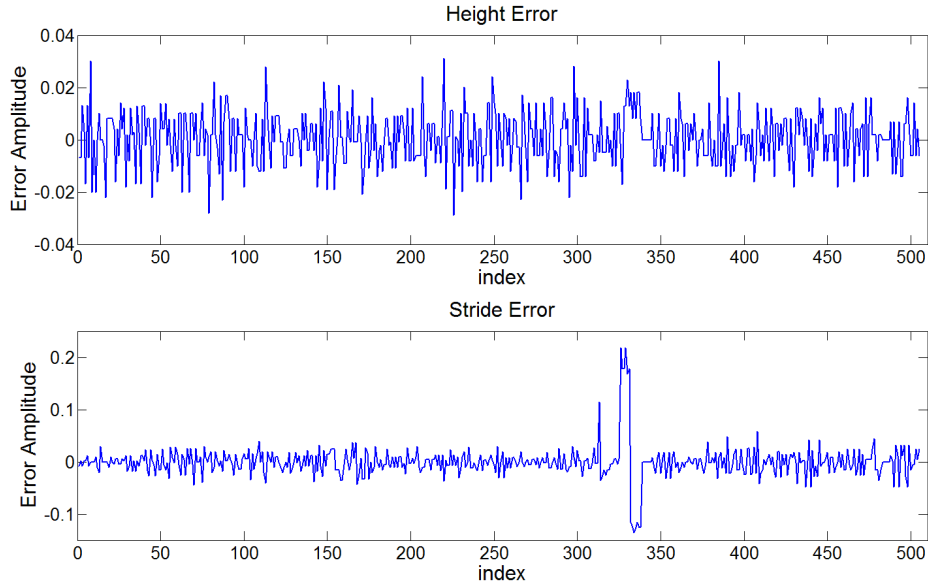


Figure 7.2.: The correlation factor between the error measurements of the utilized soft biometrics is found 0.0755. Thus, these error measurements can be characterized as i.i.d. variables, without loss of generality. The x-axes of all presented diagrams, so as to correspond to the soft biometric values of the same user (6 soft biometric values per user).

Modeling minor Clusters and a-priori Probabilities

A significant task towards applying the proposed boosting framework is to model the probability density function (*pdf*) of the noise induced by the system when measuring the soft biometric traits.

In this respect, following the steps in Section 5.2.1, the parameters of the Gaussian Mixture that best fits the normalized data are presented in Table 7.1.

Alternatively, a visual illustration of the aforementioned fitting can be found in Figure 7.3(a).

In the same context, the *pdf* that best fits the stride measurement's error is a single Gaussian distribution with a mean value $\mu = 1.64228$ and a standard deviation $\sigma = 0.09276$, as shown in Figure 7.3(b).

Next, following the methodologies presented in paragraphs 5.2.1, 5.2.1 and 5.2.1, the partitioning of the feature space is shown in Figure 7.4.

Table 7.1.: Gaussian Mixture fitting the Height Distribution

Cluster No. k	π_k	Mean μ_k	Standard Deviation σ_k
1	0.06746	1.45382	0.04130
2	0.01190	1.22333	0.01599
3	0.18452	1.57194	0.02717
4	0.29960	1.65682	0.02291
5	0.13095	1.85379	0.02816
6	0.30556	1.74110	0.02387

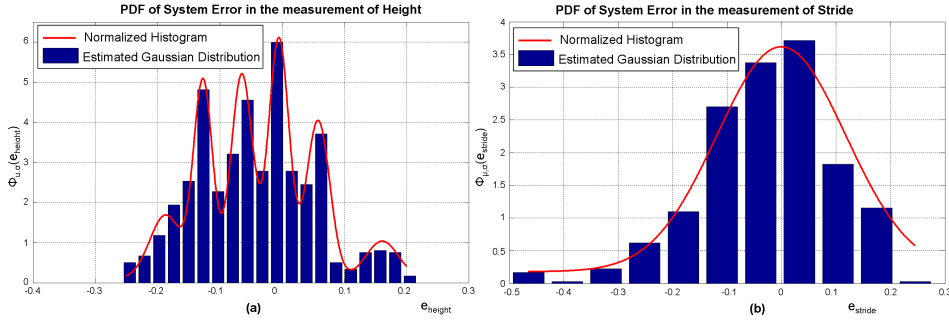


Figure 7.3.: Distribution of the Systematic Error in (a) Height and (b) Stride Measurements and the corresponding fitting curve.

Experimental results of the proposed approach

The proposed framework was tested both in terms of state of the art curves (i.e. ROC, CMS and score distributions) and experimental evaluation on well known datasets, whereby sequences from different recording sessions are used for enrolment (“*gallery*”) and identification/authentication (“*probe*”).

The proposed algorithms have been tested in both the HUMABIO Gait Dataset (DB.G.1) and ACTIBIO Gait Dataset (DB.G.2) databases (Section 3.2.2) that include gait sequences captured with stereoscopic cameras. Herein, only the gallery measurements of the HUMABIO Gait Dataset (DB.G.1) database (Section 3.2.2) have been used as the reference for both error modelling (Figure 7.3) and feature space partitioning (Figure 7.4). The performance of the system was evaluated via the probe recordings. Similarly, the ground truth values for the Soft Biometric data of each user have been measured by a manual annotator on the recorded 3D data. Moreover, these measurements have been verified via actual (i.e. real world)

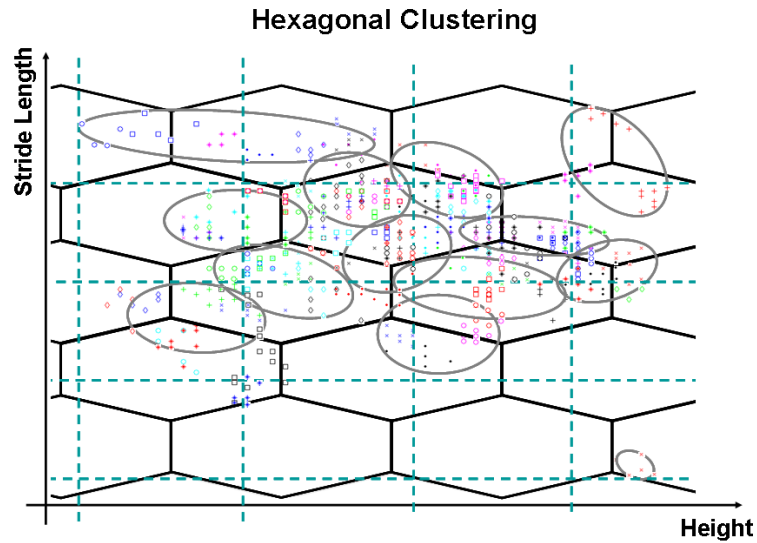


Figure 7.4.: Three alternatives for partitioning the feature space are studied: (a) UOG, (b) UHG and (c) non-linear 2D GG.

measurements and questionnaires during the capturing of the databases.

Identification Results: Figures 7.5, 7.6 and 7.7 present the comparative identification results on both datasets for the GEI-RIT, BS-RIT and BS-CIT experiments, respectively, as they are described in Section B.1 of Section B. All aforementioned diagrams illustrate the efficiency of the proposed approach using different algorithms for gait feature extraction and different databases. Four curves are displayed in each figure that correspond to the Cumulative Matching Scores (*CMS*) using solely the geometric gait features described in Section B, and different clustering techniques that boost the geometric gait feature with soft biometrics (i.e. gait, height and stride). As expected, the more non-linear the partitioning of the soft biometrics feature space is, the more significant is the increase in gait recognition efficiency. It should be also emphasized that from a theoretical point of view, the proposed framework is expected to advance the recognition rate for incorrect identification cases for subjects that exhibit soft biometric features of substantial discrimination power, i.e. for subjects that lie within the minor clusters.

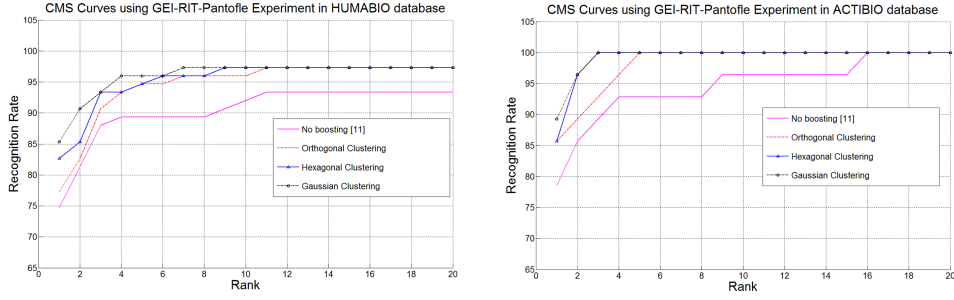


Figure 7.5.: Cumulative Matching Scores (CMS) for the $GEI - RIT$ algorithm described in Section B.1 as evaluated on both databases.

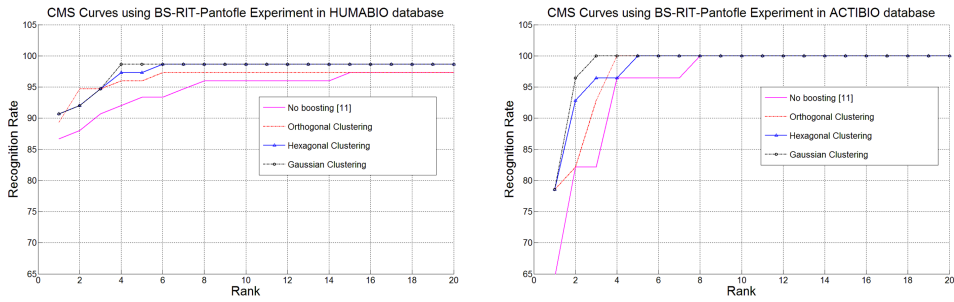


Figure 7.6.: Cumulative Matching Scores (CMS) for the Baseline RIT algorithm described in Section B.1 as evaluated on both databases.

Authentication Results: Concerning the authentication performance of the proposed approach, the False Acceptance (FAR) and False Rejection Rates (FRR) are extracted and illustrated in Figures 7.8, 7.9 and 7.10 for the GEI-RIT, BS-RIT and BS-CIT experiments on both databases, respectively. It should be mentioned that the proposed framework manages to decrease the FAR and FRR in the equal error rate EER point from 12.16% to 2.7% in the (BS-RIT Experiment) HUMABIO database in the Gaussian clustering case, while slightly lower improvements can be noticed in the orthogonal and the hexagonal partitioning cases. Similar improvements can be noticed in the ACTIBIO database (GEI-RIT experiment), where the EER falls from 15.28% to 3.57%. It should be noted that the increment in performance becomes more notable for difficult application scenarios, where the state-of-the-art gait recognition and authentication scenarios cannot achieve very high recognition and authentication rates (e.g. the Time-Scenarios [178]).

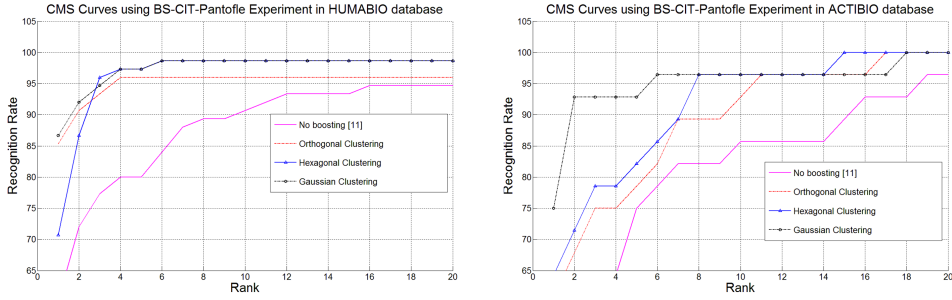


Figure 7.7.: Cumulative Matching Scores (*CMS*) for the Baseline *CIT* algorithm described in Section B.1 as evaluated on both databases.

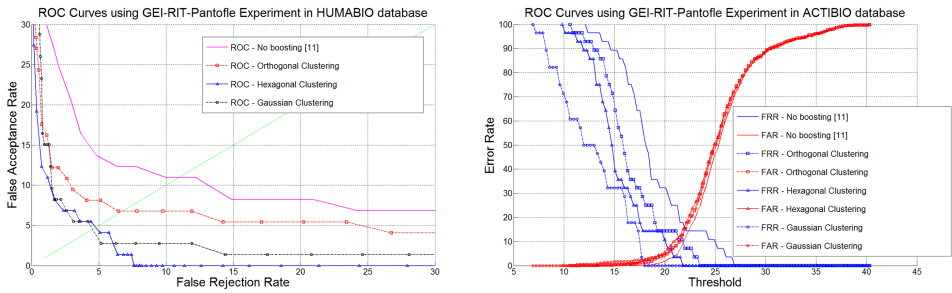


Figure 7.8.: Receiver Operating Characteristics (ROC) for the GEI-RIT algorithm described in Section B.1 as evaluated on both databases.

A direct, quantitative comparison with the boosting framework that has been proposed by Moustakas et al. in [6] can be found in Table 7.2 and Table 7.3 for the authentication and the identification performance, respectively.

Thereby, the superiority of the currently proposed scheme can be easily concluded by the corresponding advances of Equal Error Rates (EER) and the identification scores within first three ranks (i.e. Rank-1 and Rank-3).

Last but not least, it can be easily derived that the current framework provides significantly improved performance, given the fact the experimental results in [6] prove the superiority - in terms of recognition performance - of that approach over the one presented in [222].

Score Distributions: Following the aforementioned advancements in both identification and authentication performance of the proposed framework,

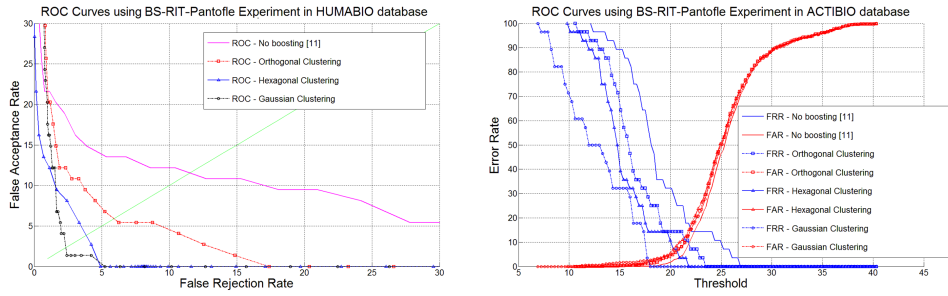


Figure 7.9.: Receiver Operating Characteristics (ROC) for the GEI-RIT algorithm described in Section B.1 as evaluated on both databases.

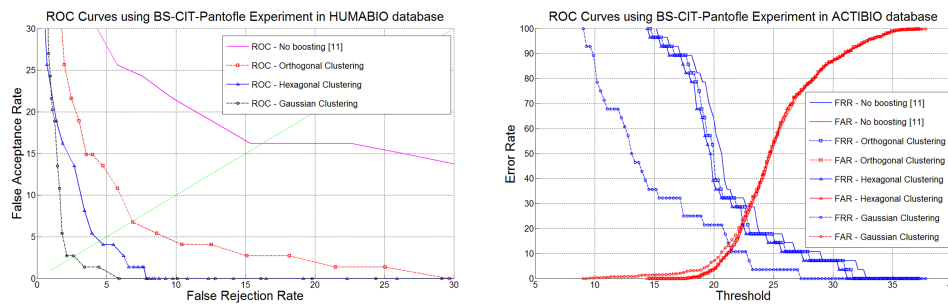


Figure 7.10.: Receiver Operating Characteristics (ROC) for the GEI-RIT algorithm described in Section B.1 as evaluated on both databases.

it is worth highlighting the improvements that are introduced in the distribution of scores between clients and impostors. A more difficult experiment than that has been herein utilized concerns the human recognition on a different date from the one he/she has been registered to the system. In particular, the recognition process in the current experiment has been attempted 6 months after the enrollment day for both databases (for further details see [178]).

As it can be seen in Figure 7.11a and 7.12a, the genuine scores are completely mixed with the ones of the impostors. However, after the application of the proposed framework, a significant separation of them can be noticed in Figures 7.11b and 7.12b.

Table 7.2.: EER Scores Comparison between the proposed method and [6]

Experiment	Initial	[6]	OG	HCG	GG
ACTIBIO - BS/RIT	28%	16%	11%	8.5%	4.3%
HUMABIO - BS/RIT	19%	16%	13.5%	9%	3.25%
ACTIBIO - GEI/RIT-Time	25%	15%	18.5%	14.8%	11.1%
ACTIBIO - BS/RIT-Time	28%	15.2%	17.3%	14.8%	11.5%
HUMABIO - BS/CIT-Time	17.5%	15%	8.07%	6.9%	5.2%

Table 7.3.: Identification Performance Comparison between the proposed method and [6]

Experiment	Initial		[6]		OG		HCG		GG	
	R-1	R-3	R-1	R-3	R-1	R-3	R-1	R-3	R-1	R-3
ACTIBIO - BS/RIT	58%	82%	70%	94%	70%	90%	79%	96%	79%	100%
HUMABIO - BS/RIT	87%	91%	91%	93%	89%	94%	91%	94%	91%	94%
ACTIBIO - GEI/RIT-Time	68%	83%	75%	91%	76%	79%	76%	86%	79%	96%
ACTIBIO - BS/RIT-Time	50%	72%	63%	79%	69%	82%	68%	85%	69%	89%
HUMABIO - BS/CIT-Time	45%	67%	65%	77%	75%	84%	61%	87%	76%	86%

7.1.2. Enhancing Face Recognition with Soft Biometrics

In order to exhibit the generic nature and applicability of the framework presented in Section 5.2, in the current Section the state of the art 3D face recognition system presented by Berretti et al. in [251] is herein utilized, as the baseline algorithm.

The proposed approach has been evaluated on the BIOTAFTOTITA 3D Face Dataset (DB.F.1) (see Section 3.2.2), that contains facial recordings of 54 subjects under different scenarios. Although the utilized 3D face matching algorithm exhibits high robustness in difficult environmental conditions and strange poses, within the current work, frames from the simple case of neutral pose in 0° have been selected pose has selected for both gallery and probe.

In particular, only the first 5 most discriminative frames of each person have been selected to form the gallery signatures. The distinctiveness is

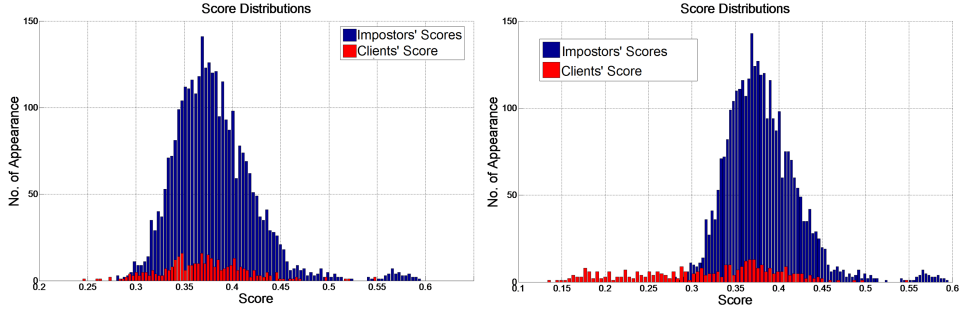


Figure 7.11.: Scores Distribution before and after the application of the proposed framework (Gaussian Clusters) in the HUMABIO RIT-Time experiment.

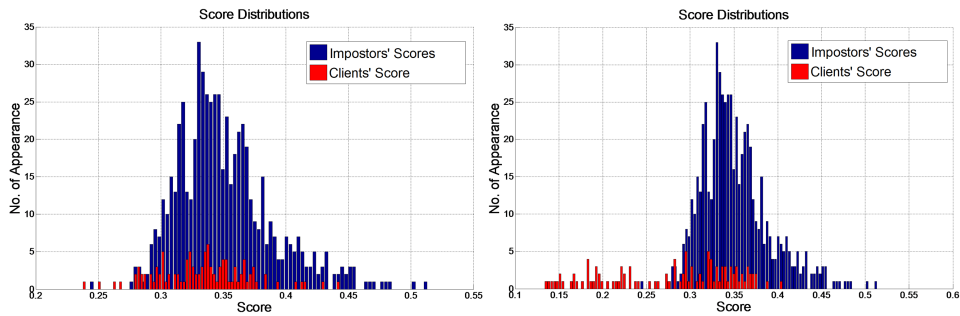


Figure 7.12.: Scores Distribution before and after the application of the proposed framework (Gaussian Clusters) in the ACTIBIO RIT-Time experiment.

evaluated by creating a confusion matrix with the similarity measure, described below, between all frames of the recorded (gallery) session. This way, 5 frames of this session are selected to be included in the biometric signature, while another session regarding a neutral pose in 0° for each subject has been used only for testing.

Similarly to the approach followed in Section 7.1.1, the proposed probabilistic framework is applied, while the performance of the combined system is evaluated and compared to similar state-of-the-art augmenting frameworks, that use soft biometric traits.

In order to remove the noisy information from the facial images, such as areas with hairs or background areas, the work of Beretti et al. has been enriched with a preprocessing step for drawing face-specific ellipses (first

row in Figure 7.14). In particular, by using as the center of the ellipse the exact location of the nosetip, the axes of the ellipse are calculated as a function of the distance between the eyes [339].

The step for the detection of the nose tip precedes the background segmentation and is initially based on an initial detection of the location of the nose tip (N_t^{kinect}) from the coloured image, as delivered by the Kinect SDK toolkit. However, since the detection accuracy of this algorithm is not sufficient (red spots in Figure 3) a post processing algorithm has been initiated. In particular, all points within a sphere of $4cm$ around N_t^{kinect} undergo PCA transformation and the new nose tip location is calculated as the median value of the closest points to the origin of the depth axis. Then, by mapping this depth value on the initial surface, one can easily estimate a good approximation of the actual nose tip location ($N_t(x_0, y_0, z_0)$), as indicated by the blue spots in Figure 7.13.

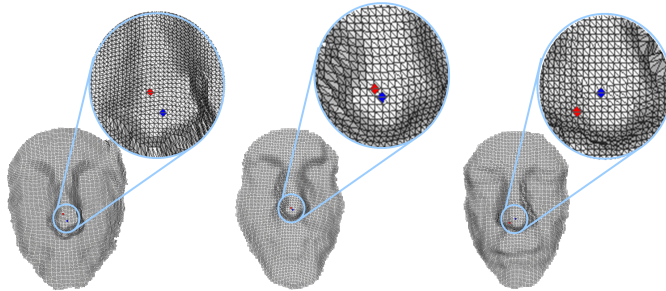


Figure 7.13.: The red spots indicate the location of the nose tip as detected by processing the colour related information of the face, while the blue ones point the location estimated by processing the depth information.

Then, the Dijkstra algorithm is applied on all facial points within the ellipse, all geodesic distances are normalized to the eyes-to-nose distance and this way, isogeodesic stripes of equal width (i.e. $1cm$) can be estimated, concentric and centered on the nose tip (see third row in Figure 7.14). Thereafter, the so-called $3D$ weighted walkthroughs ($3DWWs$) are computed between pairs of isogeodesic stripes (interstripe $3DWW$) and between each stripe and itself (intrastripe $3DWW$), as described in [251]. In particular, the $3DWWs$ are computed aggregate measures (i.e. directional indices) that provide a representation for the mutual displacement between

the set of points of two spatial entities (i.e. isogeodesic stripes). Finally, the computed $3DWWs$ are cast to a graph representation where stripes are used to label the graph nodes and $3DWWs$ to label the graph edges.

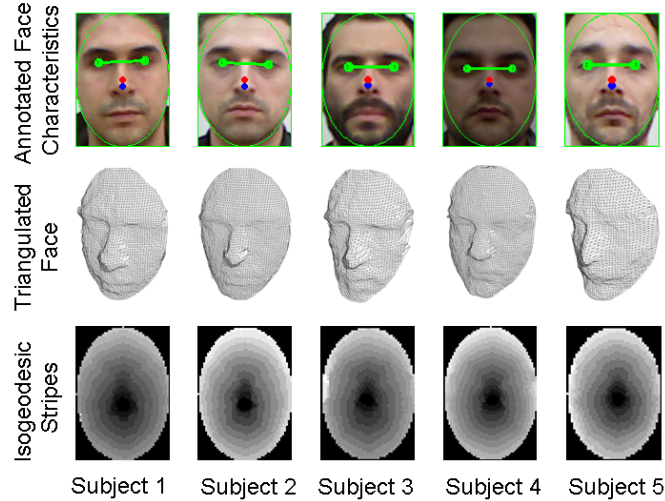


Figure 7.14.: First Row: Face-specific ellipses are drawn on the facial images so as to exclude any noisy, non-facial. points from processing; Second Row: $3D$ reconstruction of the extracted facial image; Third Row: The extracted isogeodesic stripes are drawn within the boundaries defined by each face-specific ellipse.

This way, the face recognition problem is reduced to an efficient graph matching issue that is suited to being employed in very large data sets with the support of appropriate index structures. Thus, the measure of the similarity $p(\omega|x_c)$ between two face models represented through their corresponding graphs is a combination of both the intrastripe and interstripe $3DWWs$ similarity measure, and the second summation.

All geodesic distances are normalized with respect to the Euclidean eye-to-nose distance in [251], so as to maintain a common reference for all comparisons. This way, significant information regarding the actual size of the face is discarded in favour of efficiency in processing. However, providing the distances between the so-called “nodal points” (i.e. eyes, nose and mouth), and the structural proportions of the face of the average user, the aforementioned lost information can be retrieved and significantly contribute to the recognition performance. In particular, herein the a) eye-to-eye x_{s_1} , b)

eyes-to-nose x_{s_2} and c) nose-to-mouth x_{s_3} distances for the set X of the available soft biometrics $X = \{x_{s_1}, x_{s_2}, x_{s_3}\}$.

In the current work the detection of these nodal points is performed via the utilization of the Face Tracking algorithm included in the “Kinect for Windows SDK” that is able to locate and track the movement and the orientation of the face and all its corresponding nodal points (i.e. eyes, eyebrows, mouth) with sufficient accuracy. However, small fluctuations in the distances between the measured aforementioned points are very likely to occur. This can be easily interpreted as the systematic error introduced in each measurement. Similarly to the analysis performed in Section 7.1.1, the independence between the errors of the utilized soft biometric traits can be easily proven by a similar analysis with the one presented in Section 7.1.1.

Based on the methodology described in Section 5.2 and as it was proven by the analysis presented in Section 7.1.1, the Gaussian Clustering performs better than the rest. This way, provided that the ISODATA algorithm indicated a number of 13 clusters, the soft biometric feature space is described as a Gaussian mixture (Figure 7.15) via the expectation-maximization (EM) algorithm.

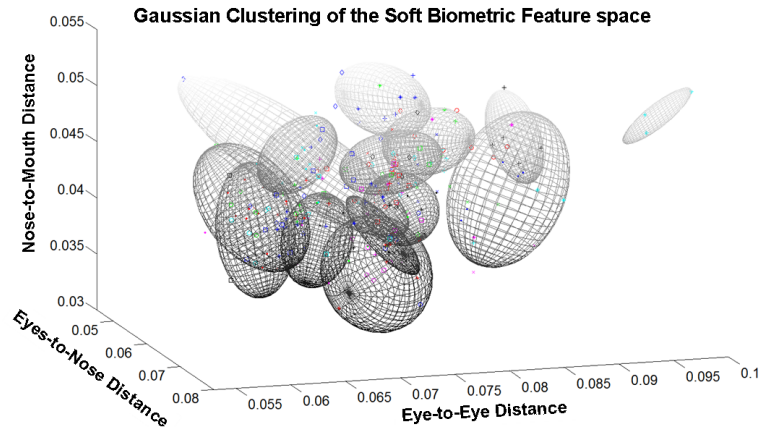


Figure 7.15.: The 3D soft biometric feature space has been partitioned in 13 Gaussian Clusters, according to the spatial proximity of the features.

Thereby each 3D Gaussian is described as

$$\mathcal{N}(\mathbf{f}_\omega | \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{Z/2} |\Sigma_k|^{1/2}} e^{-\frac{1}{2}(\mathbf{f}_\omega - \mu_k)^T \Sigma_k^{-1} (\mathbf{f}_\omega - \mu_k)} \quad (7.1)$$

where μ_k and Σ_k are the 3D mean vector and the 3×3 covariance matrix of the k^{th} Gaussian, respectively, and Vector \mathbf{f}_ω includes all utilized soft biometric trait values, $\mathbf{f}_\omega = \{x_{s_n,1}(\omega), \dots, x_{s_n,Z}(\omega)\}$. At the authentication stage, the assignment of a user to a cluster is performed via the maximum likelihood (*ML*) criterion.

In the same context, the modelling of the noise induced for each utilized soft biometric feature can be seen in Figure 7.16.

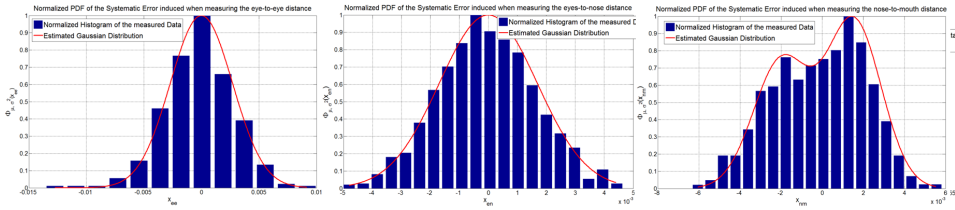


Figure 7.16.: Distribution of the induced noise during in (a) Eye-to-Eye distance (b) Nose-to-Eye distance (c) Nose-to-Mouth distance and the corresponding fitting curves.

Experimental results of the proposed approach

Concerning the **authentication** and **identification** performance of the proposed approach, the Receiver Operating Characteristics (ROC) curves and the corresponding Cumulative Matching Scores (CMS) curves are illustrated in Figure 7.17(a) and Figure 7.17(b), respectively.

The reader can easily notice a significant fall in the equal error rate EER point from 6% to 1.9% in the demanding BIOTAFTOTITA database, after the application of the proposed algorithm, while the approach proposed by Moustakas et al. [6] and Marcialis et al. [220] achieves only a 4.8% of EER. Similar improvements are concluded by the experimental results regarding the identification performance of the system, where the proposed algorithm converges to 100% of correct identifications at Rank-2, while the approach suggested in citeMoustakas10 and in [220] starts with a lower identification rate (i.e. 89% at Rank-1) and converges to 100% only at Rank-3.

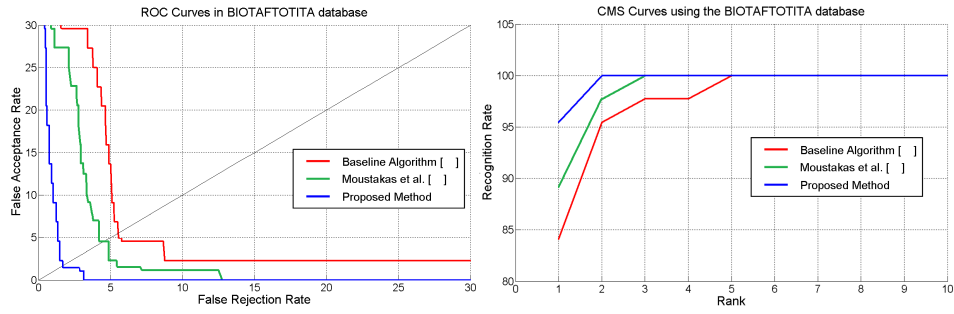


Figure 7.17.: (a) Cumulative Matching Scores (*CMS*) and (b) Receiver Operating Characteristics (*ROC*) for the 3*D* Face Recognition algorithm with and without counting in the contribution of the Soft Biometric Traits.

As it can be seen in Figure 7.18(a), the **genuine scores** are completely mixed with the ones of the impostors. However, once applying the proposed enhanced matching probability (Equation (5.18)), a significant separation of them can be noticed in Figures 7.18(b).

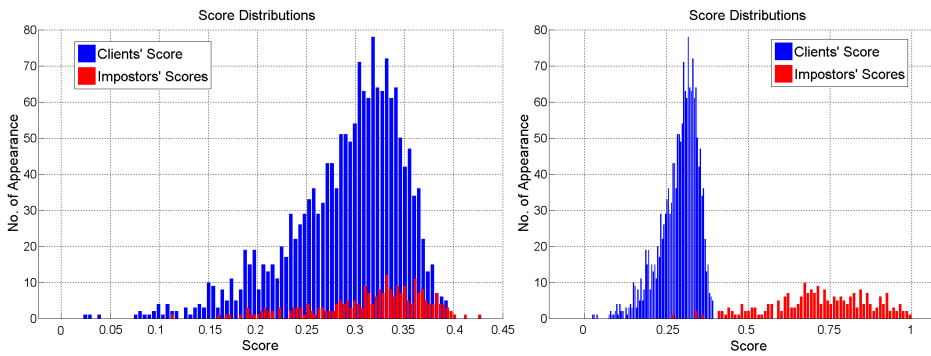


Figure 7.18.: Scores Distribution with and without counting in the contribution of the Soft Biometric Traits.

8. Conclusions

The presented thesis dealt with the development and the investigation of novel concepts regarding the challenging field of emerging behavioural biometrics (i.e. unobtrusive on-the-move-biometry), including the introduction of a completely new biometric trait, based on prehension activities, as well as the efficient exploitation of anthropometric and soft biometric traits for the enhancement of the performance in user recognition. In the next few sections, a summary of the current thesis follows in Section 8.1, along with a critical discussion (Section 8.2) and a few ideas for future work (Section 8.3).

8.1. Summary of the Thesis

The thesis started with a thorough analysis and reference to psychologically related studies in order to form a solid motivation as for the reason that the prehension related biometrics can form a trustful biometric trait. Namely, the different outcomes of the motor behaviour in respect with inherent individual differences and the indirect coupling between the perception and the reproduction of actions are referenced, in order to exhibit the mapping of behaviour onto movements serves complementarily to the psychologically proven ability of humans of differentiating behaviours and identities via subtle visual changes in movements.

Following this, an introduction in the state of the art (SoA) approaches in biometric recognition, attempting, this way, to clarify the need for and the advantages of transiting from the traditional biometric recognition methods to the so-called activity-related (i.e. behavioural) ones. After presenting and briefly analyzing the most significant works regarding existing behavioural biometrics, the extension of current SoA and the motivation towards the introduction of a new activity related trait (i.e. Prehension biometrics) extracted during everyday movements (e.g. Phone Conversation, etc.) is in

depth analyzed thoroughly justified (see Section 2.2.2 and Section 1.3).

In particular, an extension in the activity-related, unobtrusive authentication framework has been presented, that is related to prehension biometrics and includes the dynamic information of the movement of the head, the arm, the palm and the fingers derived when performing short everyday activities, without any special predefined scenario. The proposed approach refers to various activities, which include the reaching, grasping or interacting with an object in the vicinity of a user. The quality of the tracking of the movement is verified with respect to both the accuracy of the tracking algorithm and the ergonomic zones of the user.

In this context, 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 between the features.

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 depended on the quality of tracking, as 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 modalities (i.e. Reaching related and Grasping related modalities) have been evaluated via their corresponding extracted features on 4 datasets (Section 3.2.2), explicitly designed for the needs of prehension based recognition systems. For this reason, 4 different scenarios have been designed, including unimodal recognition and multimodal fusion based on-the-move recognition cases, as well as the exploitation of multiple authentication experiments. All experimental results exhibited very promising recognition rates in real time, achieving high authentication and identification rates even under difficult environmental conditions.

Extending the aforementioned uni-biometric recognition concept, the importance of the utilization of multi-biometric recognition becomes evident, while the advantages using soft-biometric traits instead of other physical or

dynamic traits are illustrated and supported by existing state-of-art works.

In this respect, of high significance is the contribution of the current thesis in the incorporation of static anthropometric characteristics in the recognition process of existing biometric systems towards the improvements in performance and the secure storage of the biometric template. The proposed static anthropometric profile is seen to have a significant contribution to the overall authentication capacity, when small datasets are regarded.

A more generic and novel probabilistic framework for augmenting biometric recognition algorithms via soft biometrics was also proposed. Hereby, the soft biometrics related partitioning of the feature space and the probabilistic modelling of the independent systematic error during soft biometric measurements are seamlessly combined with gait biometrics, so that fusion at score level is avoided.

Experimental validation in biometric recognition regarding the contribution of soft biometric systems proved significant improvements in efficiency, authentication and identification potential. Two experiments have been conducted, so as to prove the general applicability of the proposed approach and its advantages over other SoA relevant works. Namely, one using 3 known gait related dataset and one proprietary (*3D*) face related one, validating the initial assumptions regarding the expected impact. In particular, the prehension based biometric system has exhibited significant improvements when anthropometric information was taken into account, while SoA face and gait recognition systems were significantly augmented via the integration of the soft biometric based probabilistic framework. This way, it can be claimed that the latter has the ability to be directly applied to any biometric system detecting at least one soft biometrics trait.

Remaining in the context of multi-biometric approach, a multi-biometric approach based on two behavioural biometrics (i.e. prehension and gait) has been attempted. Although some improvements in gait recognition have been suggested, including the integration of a stop detection algorithm and a silhouette rotation algorithm for compensating for small divergence of the user from a straight path, the main scope was the investigation of such a purely behavioural biometric system.

Although the prehension based biometric trait may not be that accurate from itself to form a stand alone biometric system yet, it 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, fingerprint scan, etc.), as well as an integral part of an “on-the-go” authentication system. In any case, prehension biometrics are recommended for transparent, multiple authentication, so as to renew the validity of the claimed ID of the user transparently, while soft biometrics have been proven to form a significant boosting factor for any biometric system.

8.2. Critical Discussion

The current one starts with a list summarizing the most significant achievements of the current thesis, while a more detailed discussion follows thereafter.

- The intuitive assumption that prehension biometrics can form a solid biometric trait has been stated, based on a series of psychological researches.
- The high potential of prehension biometrics has been proved via experimental validation.
- Application of the proposed prehension biometric trait in multiple authentication scenarios.
- The feature classification showed that different features are suitable for different movements.
- Prehension biometrics are currently applicable only for verification purposes or as a complementary modality.
- High impact of anthropometric soft biometric characteristics in multi-biometric applications regarding accuracy and recognition performance.

As it has already been claimed, emerging biometrics, in general, cannot compete existing and widely adopted static biometric systems, such as fingerprint or iris, in terms of recognition performance. In this context, prehension biometrics (or other existing activity-related biometrics) do not aim at replacing these modalities. More preferably, they are intended to act in a

complementary and unobtrusive way to existing approaches, when no other biometric can be collected, when an additional level of security is demanded or when multiple authentication is supported by the environmental setting and the scenario.

Contrary to traditional biometrics, it is in the nature of behavioural ones (e.g. prehension, gait, etc), not only to be highly dependent on the context, but also to be able to exploit contextual information via the captured data. This can be easily perceived, by confronting the “isolated” data capturing environment of static biometrics (e.g. fingerprint/iris scanner, etc.) to the open environment for data capturing of activity related biometrics (e.g. office, workplace, airport, etc.). This way, it becomes evident that methodologies, similar to the one presented regarding the ergonomic zones, have to be developed, addressing further environmental influences on the biometric procedure.

Time persistence and lasting robustness are further critical issues of biometric traits. In this respect, it can be claimed that biometric signatures stemming from behavioural biometric modalities, in general, are more robust and persistent in time than simple static biometrics. For instance, although static biometrics, such as fingerprint, may easily be duplicated or degraded over time due to several reasons (e.g. friction, scratches, etc.), behavioural and habitual reactions and patterns, expressed by personalized behaviour traits are difficult to be changed, imitated or forgotten.

Last but not least, a significant advantage of the vision-based approaches proposed in the current thesis is the unobtrusive way they exploit for acquiring the required data. However, this seemingly positive feature may turn in a serious drawback when referring to privacy, ethical and legal issues. In particular, in traditional biometric systems, where the users have to undergo a specific procedure, so as to provide their biometric data (i.e. in fingerprint recognition the user has to place his finger in a special scanner), they are aware of being recognized and they are indirectly submitting their consent to this procedure. On the contrary, privacy issues and issues of annoyance may arise when the user could find himself under a recognition procedure anytime, despite not being always in line with this.

8.3. Future Work

A possible plan for future work based on the outcomes and the achievements of the current thesis regards the investigation of the items presented in the following list:

- Context-aware biometrics: influence of the spatiotemporal setting and the environmental conditions.
- Affective-aware behavioural biometrics.
- Activity hyper-Surface for activity recognition.
- Soft biometric keys replacing the PINs for secure template storage, revocability and cancelability.

More analytically, the current thesis delivers an extensive study regarding the recognition potential of a highly non-intrusive novel biometric modality, i.e. the Prehension biometrics. The demonstrated results indicated a notable recognition capacity of the extracted features, while significant improvements via multiple authentication methods and multimodal approaches ignite further research in the field of unobtrusive behavioural biometrics. In particular, future work in this domain could include the involvement and the corresponding study of more modalities, with high authentication capacity, enhanced environmental invariance and low correlation with each other, so as to minimize possible redundancy issues. A possible extension can also regard the recognition capacity of further daily activities. For instance, being inspired from the periodical nature of gait or the default way a short phone conversation is executed in an office environment, further patternized activities can be detected for various environments of use and accordingly investigated.

Moreover, it is strongly believed that the development of a more robust body tracker that will be able to cover the movements of the full body, both in terms of a skeleton model but also as shape, will significantly contribute to the amount of extracted features and thus, to the overall recognition performance, since it would take advantage of enhanced dynamic and static anthropometric information about the movements of the user. In general, future work in the topics addressed in the current thesis also includes the

benchmarking of the system in larger databases, so as to verify the of the proposed approaches robustness for real-world applications.

Regarding the enhancements suggested by the utilization of soft biometric traits in existing biometric systems of both dynamic and static recognition nature, an expansion to a higher dimensionality of the soft biometrics feature space is expected to further improve performance, while the proposed approach has the potential to be integrated in a series of other biometric systems. In the same respect, an issue that could also be of scientific interest is the lifting of the assumption for independency in the error measurements of the biometric traits, providing thus, a framework able to exploit any combination of soft biometric characteristics. Furthermore, the introduction of biometric keys, replacing of traditional PINs seems to gain wide acceptance towards secure template storage, since it not only offers improvements in performance, but it also expedites the users from the inconvenient noting and remembering of passwords. Of course, hereby issues regarding revocability and cancellability have to be thoroughly studied.

Last but not least, and always with respect to the emerging behavioural biometric technologies, the influence of the environmental context on differentiations on behavioural patterns should be highlighted. In this respect, being inspired by the effect of ergonomic factors on the recognition performance of the proposed system, it is strongly believed that significant efforts should be laid towards the investigation of context-aware biometrics. This way, the time, the location, the level of noise, the history of recent activities, the weather, the illumination level, etc. can be considered as contextual conditions, able to provide valuable assisting quality- or significance-related information about the recorded traits. Similar information can be also exploited by advanced biometric fusion modules, so as to improve the overall efficiency.

Bibliography

- [1] P. Shelke and A. Badiye, “Biometrics for Enhancement of Security Standards”, *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 9, pp. 97–102, 2012.
- [2] National Science and Technology Council Subcommittee on Biometrics and Identity Management, “The National Biometrics Challenge”, September 2011.
- [3] N.V. Nedap, “Selecting Identification Technology in [http : //www.nedap – securitymanagement.com/ru/component/docman/doc_download/35-selecting – identification – technology.html](http://www.nedap-securitymanagement.com/ru/component/docman/doc_download/35-selecting-identification-technology.html), last accessed in Nov. 2012.”
- [4] H. Sakoe and S. Chiba, “Dynamic programming algorithm optimization for spoken word recognition”, in *Readings in speech recognition*, vol. 26, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990.
- [5] Microsoft Corporation, “Kinect for Windows”, 2012.
- [6] K. Moustakas, D. Tzovaras, and G. Stavropoulos, “Gait Recognition Using Geometric Features and Soft Biometrics”, *IEEE Signal Processing Letters*, vol. 17, no. 4, pp. 367–370, 2010.
- [7] F. A. Qazi, “A survey of biometric authentication systems”, In proc. of *International Conference on Security and Management (SAM '04)*, pp. 61–67, 2004.
- [8] Q. Xiao, “Security issues in biometric authentication”, *Information Assurance Workshop, IAW 2005*, pp. 8–13, 2005.
- [9] B. H and C. G, “Audio-visual speech synchrony measure for talking-face identity verification”, in *IEEE ICASSP*, Honolulu, HI, pp. II.233–II.236, 2007.

- [10] G. Potamianos, C. Neti, J. Luetten, and I. Matthews, *Audiovisual Automatic Speech Recognition: An Overview*. Cambridge, MA: MIT Press, 2004.
- [11] R. Landais, H. Bredin, L. Zouari, and G. Chollet, “Vérification audiovisuelle de l’identité”, in *in Proc. Traitement et Analyse de l’Information: Méthodes et Applications (TAIMA)*, 2007.
- [12] A. Drosou, G. Stavropoulos, D. Ioannidis, K. Moustakas, and D. Tzovaras, “Unobtrusive multi-modal Biometric Recognition Approach using Activity-related Signatures”, *Computer Vision, IET*, vol. 5, no. 6, pp. 367 – 379, 2010.
- [13] A. K. Jain, S. C. Dass, and K. Nandakumar, “Soft Biometric Traits for Personal Recognition Systems”, in *Proc. of International Conference on Biometric Authentication (ICBA)*, pp. 731–738, 2004.
- [14] A. Ross, A. Nandakumar, A. Jain, “Handbook of Multibiometrics”, *International Series on Biometrics*, vol. 6, Springer, 2006.
- [15] Y. Wang and Z. Liu, “A Survey on Multimodal Biometrics”, *Advances in Automation and Robotics*, G. Lee (Ed.), vol. 2, LNEE 123, pp. 387396, 2011.
- [16] L. Hong, A. Jain and S. Pankanti, “Can Multibiometrics Improve Performance?”, in *Proc. of IEEE workshop on Automatic Identification Advanced Technologies*, New Jersey, USA, pp. 59–64, 1999.
- [17] A. Jain and A. Kumar, “Biometric Recognition: An Overview, in *Second Generation Biometrics: The Ethical, Legal and Social Context*, vol. 11, Mordini, Emilio; Tzovaras, Dimitros (Eds.), 2012.
- [18] A. K. Jain, K. Nandakumar, X. Lu, and U. Park, “Integrating faces, fingerprints, and soft biometric traits for user recognition”, in *In Proc. of the Biometric Authentication Workshop*, pp. 259–269, 2004.
- [19] A. Ross and A.K. Jain, “Information Fusion in Biometrics”, *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115–2125, 2003.
- [20] A. Lanitis and N. Tsapatsoulis, ”On the Quantification of Aging Effects on Biometric Features”, *Artificial Intelligence Applications and*

Innovations IFIP Advances in Information and Communication Technology, vol. 339, pp. 360-367, 2010

- [21] T. Zhang, X. Li, D. Tao, and J. Yang, “Multimodal biometrics using geometry preserving projections”, *Pattern Recognition*, vol. 41, no. 3, pp. 805–813, 2008.
- [22] P. P. Boda and E. Filisko, “Virtual Modality: a Framework for Testing and Building Multimodal Applications”, *Workshop on Spoken Language Understanding for Conversational Systems (HLTNAACL '04)*, 2004.
- [23] B. Bhanur and J. Han, “Model-based human recognition 2D and 3D gait”, *Advances in Pattern Recognition: Human Recognition at a Distance in Video*, Springer, pp. 65-94, 2011.
- [24] G. Junxia, D. Xiaoqing, W. Shengjin, and W. Youshou, “Action and gait recognition from recovered 3-D human joints”, *IEEE Trans. Syst., Man, Cybern., Part B: Cybernetics*, vol. 40, no. 4, pp. 1021-1033, 2010.
- [25] C. Chen, J. Liang, H. Zhao, H. Hu and J. Tian, “Frame difference energy image for gait recognition with incomplete silhouettes”, *Pattern Recognition Lett.*, vol. 30, no. 11, pp. 977-984, 2009.
- [26] D.K.Wagg and M.S.Nixon, “On automated model-based extraction and analysis of gait”, in *Proc. IEEE. Int. Conf. Automatic Face and Gesture Recognition*, pp. 11–16, 2004.
- [27] L. Wang, H. Ning, T. Tan and W. Hu, “Fusion of static and dynamic body biometrics for gait recognition”, in *proc. IEEE Int. Conf. Comput. Vis.*, pp. 1449-1454, 2003.
- [28] L. Wang, H. Ning, T. Tan, and W. Hu, “Fusion of static and dynamic body biometrics for gait recognition”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 2, pp. 149–158, 2003.
- [29] D. Cunado, M. Nixon, and J. N. Carter, “Automatic extraction and description of human gait models for recognition purposes”, *Computer Vision and Image Understanding*, vol. 90, pp. 1–41, 2003.

- [30] A. Gritai, Y. Sheikh, and M. Shah, “On the use of anthropometry in the invariant analysis of human actions”, in *Proc. Int. Conf. Pattern Recog.*, pp. 923926, 2004.
- [31] F. Lv and R. Nevatia, “Recognition and segmentation of 3-D human action using HMM and multi-class AdaBoost”, in *Proc. Eur. Conf. Comput. Vis.*, pp. 359372, 2006.
- [32] T. Moeslund, A. Hilton, and V. Kruger, “A survey of advances in vision-based human motion capture and analysis”, *Comput. Vis. Image Underst.*, vol. 104, no. 2/3, pp. 90-126, 2006.
- [33] J. Deutscher, A. Blake, and I. Reid, “Articulated body motion capture by annealed particle filtering”, in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 126133, 2000.
- [34] R. Kehl, M. Bray, and L. Van Gool, “Full body tracking from multiple views using stochastic sampling”, in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 129136, 2005.
- [35] V. Parameswaran and R. Chellappa, “View invariance for human action recognition”, *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 83-101, 2006.
- [36] I. Bouchrika and M. Nixon, “Model-based feature extraction for gait analysis and recognition”, in *proc. Int. Conf. Comput. Vis./Comput. Graph. Collaboration Techn. Appl.*, pp. 150-160, 2007.
- [37] N. Boulgouris, K. Plataniotis, and D. Hatzinakos, “Gait Recognition Using Linear Time Normalization”, *Pattern Recognition*, vol. 39, pp. 969–979, 2006.
- [38] S. Sarkar, P. J. Phillips, Z. Liu, I. Robledo-Vega, P. Grother, and K. W. Bowyer, “The Human ID Gait Challenge Problem: Data Sets, Performance, and Analysis”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162–177, 2005.
- [39] N. Boulgouris, K. Plataniotis, and D. Hatzinakos, “Gait recognition using dynamic time warping”, in *IEEE 6th Workshop on Multimedia Signal Processing*, (Siena), pp. 263–266, 2004.

- [40] C. Changhong, L. Jimin, Z. Heng, H. Haihong, and T. Jie, “Factorial HMM and Parallel HMM for Gait Recognition”, *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 39, no. 1, pp. 114–123, 2009.
- [41] A. Barreto, J. Zhai, and M. Adjouadi, “Non-intrusive physiological monitoring for automated stress detection in human-computer interaction”, in *Proceedings of the IEEE international conference on Human-computer interaction (HCI’07)*, Heidelberg: Springer-Verlag, pp. 29–38, 2007.
- [42] M. Alcoverro, J. R. Casas, M. Pardas, Skeleton and Shape Adjustment and Tracking in Multicamera Environments, vol. 6169, Mallorca, pp. 88–97, 2010.
- [43] M. Pantic, A. Nijholt, A. Pentland and T. Huanag “Human-Centred Intelligent HumanComputer Interaction (HCI): how far are we from attaining it?”, *Human-Centred Intelligent HumanComputer Interaction* vol. 1, no. 2, pp. 168–187, 2008.
- [44] G. Sandbach, S. Zafeiriou, M. Pantic and L. Yin, “Static and dynamic 3D facial expression recognition: A comprehensive survey”, *Image and Vision Computing*, vol. 30,m pp. 683697, 2012
- [45] Z. Liu and S. Sarkar, “Improved Gait Recognition by Gait Dynamics Normalization”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 6, pp. 863–876, 2006.
- [46] M. Goffredo, J. N. Carter, and M. Nixon, “Front-view Gait Recognition”, in *IEEE Second International Conference on Biometrics: Theory, Applications and Systems (BTAS 08)*, pp. 1-6, 2008.
- [47] M. Goffredo, R. D. Seely, J. N. Carter, and M. S. Nixon, “Markerless View Independent Gait Analysis with Self-camera Calibration”, in *IEEE International Conference on Automatic Face and Gesture Recognition*.
- [48] A. Kale, A. Chowdhury, and R. Chellappa, “Towards a view invariant gait recognition algorithm”, in *Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 143–150, 2003.

- [49] F. Jean, A. B. Albu, and R. Bergevin, “Towards view-invariant gait modeling: Computing view-normalized body part trajectories”, *Pattern Recogn.*, vol. 42, no. 11, pp. 2936–2949, 2009.
- [50] N. Spencer and J. Carter, “Towards pose invariant gait reconstruction”, in *Proceedings of the IEEE International Conference on Image Processing, ICIP 2005*, pp. 261–264, 2005.
- [51] H. Hu, “Enhanced Gabor Feature Based Classification Using a Regularized Locally Tensor Discriminant Model for Multiview Gait Recognition”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 7, 2013.
- [52] N. Boulgouris and X. Huang, “Gait Recognition Using HMMs and Dual Discriminative Observations for Sub-Dynamics Analysis”, *IEEE Trans. Image Process.*, vol. 22, no. 9, 2013
- [53] W. Kusakunniran, Q. Wu, J. Zhang, Y. Ma and H. Li, “A New View-Invariant Feature for Cross-View Gait Recognition”, *IEEE Trans. Inf. Forensics Security* vol. 8, no. 10, 2013
- [54] D. Matovski, M. Nixon, S. Mahmoodi and J. Carter, “The Effect of Time on Gait Recognition Performance”, *IEEE Trans. Inf. Forensics Security* vol. 7, no. 2, 2012
- [55] D. Xu, Y. Huang, Z. Zeng and X. Xu, “Human Gait Recognition Using Patch Distribution Feature and Locality-Constrained Group Sparse Representation”, *IEEE Trans. Image Process.*, vol. 21, no. 1, 2012
- [56] W. Kusakunniran, Q. Wu, J. Zhang and H. Li, “Gait Recognition Under Various Viewing Angles Based on Correlated Motion Regression”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 6, 2012.
- [57] X. Huang and N. Boulgouris, “Gait Recognition With Shifted Energy Image and Structural Feature Extraction”, *IEEE Trans. Image Process.*, vol. 21, no. 4, 2012
- [58] A. Veeraraghavan, A. R. Chowdhury, and R. Chellappa, “Role of Shape and Kinematics in Human Movement Analysis”, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR’04)*, vol. 1, pp. 730–737, 2004.

- [59] M. Fairhurst, F. Deravi, N. Mavity, J. George, and K. Sirlantzis, *Intelligent Management of Multimodal Biometric Transactions*. Springer Berlin-Heidelberg, 2003.
- [60] R. Seely, S. Samangoeei, M. Lee, J. Carter, and M. Nixon, “The University of Southampton Multi-Biometric Tunnel and introducing a novel 3D gait dataset”, in *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 1–6, 2008.
- [61] N. Boulgouris and Z. Chi, “Gait recognition using radon transform and linear discriminant analysis”, *IEEE Trans. Image Process*, vol. 16, no. 3, pp. 731–740, 2007.
- [62] M. Karg, K. Khlentz and M. Buss, “Recognition of Affect Based on Gait Patterns”, *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 4, pp. 1050-1061, 2010.
- [63] K. Naugle, C. Hass, J. Joyner, S. Coombes and C. Janelle, “Emotional state affects the initiation of forward gait”, , vol. 11, no. 2, pp. 267–277, 2011.
- [64] J. A.K. and P. U., *Facial marks: Soft biometric for face recognition*,. PhD thesis, MSU (Michigan State University), 2009.
- [65] L. H. L. Hong and A. J. A. Jain, “Integrating faces and fingerprints for personal identification”, 1998.
- [66] A. K. Jain, A. Ross, and S. Prabhakar, “An Introduction to Biometric Recognition”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, 2004.
- [67] J. Candamo, M. Shreve, D. Goldgof, D. Sapper and R. Kasturi, “Understanding Transit Scenes: A Survey on Human Behavior-Recognition Algorithms”, *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, 2010
- [68] W. Bowyer, K. Chang and P. Flynn, “A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition”, *Computer Vision and Image Understanding*, vol. 101, pp. 115, 2006.

- [69] C. Kotropoulos and I. Pitas, “Rule-based face detection in frontal views”, *Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, pp. 2537–2540, 1997.
- [70] E. Osuna, R. Freund, and F. Girosi, “Training Support Vector Machines: An Application to Face Detection”, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 130–136, 1997.
- [71] H. Schneiderman and T. Kanade, “A Statistical Method for 3D Object Detection Applied to Faces and Cars,”, *IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 746–751, 2000.
- [72] L. Goldmann, U. J. Mönich, and T. Sikora, “Components and Their Topology for Robust Face Detection in the Presence of Partial Occlusions”, *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 559–569, 2007.
- [73] P. Viola and M. J. Jones, “Robust Real-Time Face Detection”, *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [74] R. Xiao, M.-J. Li, and H.-J. Zhang, “Robust multipose face detection in images”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 31–41, 2004.
- [75] T. Kailath, “The Divergence and Bhattacharyya Distance Measures in Signal Selection”, *IEEE Trans. Commun. Tech.*, vol. 15, pp. 52–60, 1967.
- [76] D. Comaniciu, V. Ramesh, and P. Meer, “Real-time tracking of non-rigid objects using mean shift”, in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, vol. 2, pp. 142–149, 2000.
- [77] A. Cheddad, J. Condell, K. Curran, and P. McKeivitt, “A skin tone detection algorithm for an adaptive approach to steganography”, *Signal Processing archive*, vol. 89, no. 12, pp. 2465–2478, 2009.
- [78] M. Corey, F. Farzam, and J. H. Chong, “The effect of linearization of range in skin detection”, in *IEEE International Conference on Information, Communications and Signal Processing*, pp. 1–5, 2007.

- [79] U. A. Khan, M. I. Cheema, and N. M. Sheikh, “Adaptive video encoding based on skin tone region detection”, in *IEEE Students Conference*, pp. 129–134, 2002.
- [80] V. Vezhnevets, V. Sazonov, and A. Andreeva, “A survey on pixel-based skin color detection techniques”, in *Graphicon*, pp. 85–92, 2003.
- [81] Y. Liu and C. X. Wang, “An improved algorithm of human skin detection in video image based on linear combination of 2nd order Markov and Wiener predictor”, in *International Symposium on Computer Science and Computational Technology*, pp. 665–668, 2008.
- [82] J.-C. Terrillon, M. Shirazi, H. Fukamachi, and S. Akamatsu, “Comparative Performance of Different Skin Chrominance Models and Chrominance Spaces for the Automatic Detection of Human Faces in Color Images”, in *4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 54–61, 2000.
- [83] A. Diplaros, T. Gevers, and N. Vlassis, “Skin detection using the EM algorithm with spatial constraints”, in *IEEE International Conference on Systems, Man and Cybernetics*, pp. 3071 – 3075, 2004.
- [84] X. Zhao, F. Boussaid, and A. Bermak, “Characterization of a 0.18mm CMOS color processing scheme for skin detection”, *IEEE Sensors Journal*, vol. 7, no. 11, pp. 1471–1474, 2007.
- [85] W. Skarbek and A. Koschan, “Colour image segmentation-a survey”, Technische Universitaet Berlin, Oct. 1994.
- [86] B.Wu, H. Ai, C. Huang, and S. Lao, “Fast rotation invariant multi-view face detection based on real Adaboost”, in *6th Int. Conf. Autom. Face Gesture Recogn.*, pp. 79–84, 2004.
- [87] C. Huang, H. Ai, Y. Li, and S. Lao, “High-Performance Rotation Invariant Multiview Face Detection”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 671–686, 2007.
- [88] T. Mita, T. Kaneko, and O. Hori, “Joint Haar-Like Features for Face Detection”, in *10th IEEE Intl Conf. Computer Vision*, vol. 2, pp. 1619–1626, 2005.

- [89] S. Baluja, M. Sahami, and H. Rowley, “Efficient Face Orientation Discrimination”, in *IEEE Intl Conf. Image Processing*, pp. 589–592, 2004.
- [90] J. R. Matey, O. Naroditsky; K. Hanna, R. Kolczynski, D. J.LoIacono; S. Mangru, M. Tinker, T. M. Zappia, W. Y. Zhao, “Iris on the Move: Acquisition of Images for Iris Recognition in Less Constrained Environments”, *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1936–1947, 2006.
- [91] R. Cappelli, D. Maio, D. Maltoni, J. L.Wayman, A. K. Jain, “Performance Evaluation of Fingerprint Verification Systems,” *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 28, no. 1, pp. 3–18, 2006.
- [92] A. Hicklin, H. Korves, B. Ulery, M. Zoepfl, M. Bone, P. Grother, R. Michaels, S. Otto, C. Watson, “Fingerprint vendor Technology Evaluation,” *Summary of results and analysis Report*, 2004.
- [93] A. K. Jain, K. Nandakumar and A. Ross, “Score normalization in multimodal biometric systems”, *Elsevier Pattern Recognition*, vol. 38, pp. 2270–2285, 2005.
- [94] N. Poh, T. Bourlai, J. Kittler, L. Allano, F. Alonso-Fernandez, O. Ambekar, J. Baker, B. Dorizzi, O. Fatukasi, J. Fierrez, H. Ganster, J. Ortega-Garcia, D. Maurer, A.A. Salah, T. Scheidat and C. Vielhauer, “Benchmarking Quality-Dependent and Cost-Sensitive Score-Level Multimodal Biometric Fusion Algorithms”, *IEEE Trans.Inf. Forensics Security*, vol. 4, no. 4, 2009.
- [95] N. Poh, J. Kittler and T. Bourlai, “Quality-Based Score Normalization With Device Qualitative Information for Multimodal Biometric Fusion”, *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 3, 2010.
- [96] N. Poh, D. Windridge, V. Mottl, A. Tatarchuk and A. Elisseyev, “Addressing Missing Values in Kernel-Based Multimodal Biometric Fusion Using Neutral Point Substitution”, *IEEE Trans.Inf. Forensics Security*, vol. 5, no. 3, 2010.

- [97] N. Poh and J. Kittler, “A Unified Framework for Biometric Expert Fusion Incorporating Quality Measures”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, 2012.
- [98] M. Monwar and M. Gavrilova, “Multimodal Biometric System Using Rank-Level Fusion Approach”, *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 4, 2009.
- [99] M.C.C. Abreu and M. Fairhurst, “Analyzing the Benefits of a Novel Multiagent Approach in a Multimodal Biometrics Identification Task”, *IEEE Systems Journal* vol. 3, no. 4, 2009.
- [100] A. Kumar, V. Kanhangad, and D. Zhang, “A New Framework for Adaptive Multimodal Biometrics Management”, *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 1, 2010.
- [101] F. Alonso-Fernandez, J. Fierrez, D. Ramos and J. Gonzalez-Rodriguez, “Quality-Based Conditional Processing in Multi-Biometrics: Application to Sensor Interoperability”, *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 5, 2010.
- [102] R. Seely, S. Samangoei, L. Middleton, J. Carter and M. Nixon, “The University of Southampton Multi-Biometric Tunnel and introducing a novel 3D gait dataset”, in Proc. 2nd *IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2008.
- [103] M. Fairhurst, R. Guest, F. Deravi and J. George, “Using Biometrics as an enabling technology in balancing universality and selectivity for management of information access,”, *Universal Access: Theoretical Perspectives, Practice and Experience: 7th ERCIM International Workshop on User Interface for All*, Springer-Verlag Lecture Notes in CS 2615, pp. 249–259, 2002.
- [104] M. C. Fairhurst, F. Deravi, N. Mavity, J. George, K. Sirlantzis, “Intelligent Management of Multimodal Biometric Transactions,” *LECTURE NOTES IN COMPUTER SCIENCE*, no. 2774, pp. 1254–1260, 2003.
- [105] M. Garris, E. Tabassi, and C. Wilson, “NIST Fingerprint Evaluations

- and Developments”, *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1915–1926, 2006.
- [106] A. Dantcheva, C. Velardo, A. D’Angelo, and J.-L. Dugelay, “Bag of soft biometrics for person identification”, *Multimedia Tools and Applications*, vol. 51, pp. 739–777, 2010.
- [107] A. Dantcheva, J. Dugelay and P. Elia, “Soft biometrics systems: Reliability and asymptotic bounds”, in *proc. 4th IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, pp. 1–6, 2010
- [108] S. Chowhan and G. N. Shinde, “Iris Biometrics Recognition Application in Security Management”, in *Congress on Image and Signal Processing (CISP)*, vol. 1, pp. 661–665, 2008.
- [109] K. I. Chang, K. W. Bowyer, and P. J. Flynn, “An evaluation of multimodal 2D+3D face biometrics” *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 4, pp. 619–24, 2005.
- [110] R. Jayadevan, S. R. Kolhe, and P. M. Patil, “Dynamic Time Warping based Static Hand Printed Signature Verification”, *Journal of Pattern Recognition Research*, vol. 4, no. 1, pp. 52–65, 2009.
- [111] P. Campisi, E. Maiorana, M. L. Bosco, and A. Neri, “User authentication using keystroke dynamics for cellular phones”, *Signal Processing (IET)*, vol. 3, no. 4, pp. 333, 2009.
- [112] R.V. Yampolskiy, “Behavioural biometrics: a survey and classification” ,, *Int. J. Biometrics*, vol. 1, no. 1, pp. 81–113, 2008
- [113] L. Benedikt, D. Cosker, P.L. Rosin and D. Marshall, “Assessing the Uniqueness and Permanence of Facial Actions for Use in Biometric Applications”, *IEEE Trans. Syst., Man, Cybern. A, Cybern.*, vol. 40, no. 3 , pp. 449–460, 2010.
- [114] J. L. Wayman (Ed.), “Large-scale Civilian Biometric Systems - Issues and Feasibility”, in *Proc. of Card Tech / Secur Tech ID*, 1997, available online at <http://www.engr.sjsu.edu/biometrics/nbtccw.pdf>

- [115] A. Dantcheva, N. Erdogmus, and J.-L. Dugelay, “On the reliability of eye color as a soft biometric trait”, in *IEEE Workshop on Applications of Computer Vision (WACV)*, pp. 227–231, 2011.
- [116] K. Ricanek and B. Barbour, “What Are Soft Biometrics and How Can They Be Used?” *Computer*, vol. 44, no. 9, pp. 106 – 108, 2011.
- [117] A. K. Jain, S. C. Dass, and K. Nandakumar, “Can soft biometric traits assist user recognition?” in *In Proceedings of SPIE*, pp. 561–572, 2004.
- [118] V. Bettadapura, “Face Expression Recognition and Analysis: The State of the Art”, *Computer Vision and Pattern Recognition*, pp. 1–27, 2012.
- [119] A. Abate, M. Nappi, D. Riccio, G. Sabatino, “2D and 3D face recognition: A survey”, *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885-1906, 2007.
- [120] X. Li, A. Li and X. Bai, “3D face detection and face recognition: state of the art and trends”, in proc , *International Conference on Image Processing and Pattern Recognition*, 2010.
- [121] M. Pantic, “Machine Analysis of Facial Behaviour: Naturalistic and Dynamic Behaviour”, *Philos Trans R Soc Lond B Biol Sci*, vol. 364, pp. 3505-3513, 2009.
- [122] F. Matta and J. Dugelay, “Person recognition using facial video information: A state of the art”, *Image and Vision Computing*, vol. 20, no. 3, pp. 180-187, 2009.
- [123] A. Hadid, J. Dugelay and M. Pietikainen, “On the use of dynamic features in face biometrics: recent advances and challenges”, *Signal, Image and Video Processing*, vol. 5, no. 4, pp. 495-506, 2011.
- [124] B. Heisele, P. Ho, J. Wu, T. Poggio, “Face recognition: Component based versus global approaches”, *Computer Vision and Image Understanding*, vol. 91 no. 1-2, pp. 621, 2003.
- [125] S. More and P. Deore, “A survey on gait biometrics”, *World Journal of Science and Technology*, vol. 2, no. 4, pp. 146-151, 2012.

- [126] K. Luu, K. Ricanek, T. Bui, and C. Suen, “Age Estimation Using Active Appearance Models and Support Vector Machine Regression”, in *Proc. 3rd Intl Conf. Biometrics: Theory, Applications, and Systems (BTAS)*, pp. 314–318, 2009.
- [127] W. Yang, C. Chen, K. Ricanek, and C. Sun, “No Title”, in *Proc. 8th Intl Conf. Advances in Neural Net-works (ISNN), LNCS 6676, Springer*, 2011, pp. 251–257.
- [128] P. Phillips, “Improving Face Recognition Technology”, *Computer*, pp. pp. 84–86, 2011.
- [129] S. Baluja and H. Rowley, “Boosting Sex Identification Performance”, *Intl J. Computer Vision*, pp. 111–119, 2007.
- [130] Y. W.;, K. Ricanek, C. C.;, and Y. C.;, “Gender classification from infants to seniors”, in *Biometrics: 4th IEEE International Conference on Theory Applications and Systems (BTAS)*, pp. 1 – 6, 2010.
- [131] X. Q., Z. S., and T. T., “Learning appearance primitives of iris images for ethnic classification”, in *IEEE International Conference on Image Processing (ICIP)*, pp. II.405 – II.408, 2007.
- [132] L. Liu, W. Jia and Y. Zhu “Survey of Gait Recognition”, *ICIC 2009*, D.-S. Huang et al. (Eds.), Springer-Verlag Berlin, pp. 652-659, 2009
- [133] D. Reid, S. Samangooei, C. Chen, M. Nixon and A. Ross, “Soft Biometrics for Surveillance: An Overview”, *Handbook of Statistics, Elsevier*, vol. 31, 2013.
- [134] *Biometric Recognition*, Z. Sun, J. Lai, X. Chen, T. Tan (Eds.), Springer, 2011.
- [135] X. Q., Z. S., and T. T., “Global texture analysis of iris images for ethnic classification”, *LNCS - Advances in Biometrics (ICB)*, vol. 3832, pp. pp. 411–418, 2005.
- [136] “ACTIBIO ICT STREP”, 2008. [Online]. Available: <http://www.actibio.eu:8080/actibio>

- [137] S. Lagree and B. K. W., “Predicting ethnicity and gender from iris texture”, in *IEEE International Conference on Technologies for Homeland Security (HST)*, pp. 440 – 445, 2011.
- [138] E. Jeges, I. Kispal, and Z. Hornak, “Measuring human height using calibrated cameras”, in *Conference on Human System Interactions*, pp. 755–760, 2008.
- [139] Y. R., G. Rosenbush, and Q. Z., “Computational Approaches for Real-time Extraction of Soft Biometrics”, in *19th International Conference on Pattern Recognition (ICPR)*, pp. 1–4, 2008.
- [140] R. Zewail, A. Elsafi, M. Saeb, and N. Hamdy, “Soft and Hard Biometrics Fusion for Improved Identity Verification”, in *The 47th Mid-west Symposium on Circuits and Systems (MWSCAS)*, pp. I225–8, 2004.
- [141] K. Niinuma, U. Park, and A. Jain, “Soft Biometric Traits for Continuous User Authentication”, *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 771 – 780, 2010.
- [142] V. S. Meenakshi and G. Padmavathi, “Securing Iris Templates using Combined User and Soft Biometric based Password Hardened Fuzzy Vault”, *Computer Science*, vol. 7, no. 2, pp. 1–8, 2010.
- [143] D. Maltoni, A. Jain, and S. Prabhakar, *Handbook of fingerprint recognition*. Springer Professional Computing, 2nd ed., 2009.
- [144] C. Poynton, “Frequently Asked Questions about Color”, 1997.
- [145] R. R. Porle, A. Chekima, F. Wong, and G. Sainarayanan, “Wavelet-based skin segmentation for detecting occluded arms in human body pose modelling system”, in *IEEE Proc. International Conference on Intelligent and Advanced Systems 2007*, pp. 764–769, IEEE, 2007.
- [146] R. Eldridge, H. Rudolph, J. Duckworth, and P. Wilson, “Improved Background Removal through Polarisation in Vision-Based Tabletop Interface”, in *6th IEEE/ACIS International Conference on Computer and Information Science (ICIS)*., pp. 326 – 331, 2007.

- [147] L. Lu and G. Hager, “Dynamic foreground-background extraction from images and videos using random patches”, in *20th Annual Conference on Neural Information Processing Systems*, pp. 929–936, 2006.
- [148] M. J. Jones and J. M. Rehg, “Statistical Color Models with Application to Skin Detection”, *Int. Journal of Computer Vision*, vol. 46, no. 1, pp. 81–96, 2002.
- [149] A. Colmenarez, B. Frey and T.S. Huang, “A probabilistic framework for embedded face and facial expression recognition”, in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 592–597, 2003.
- [150] G. Aggarwal, C. Park, and A. K. R. Chowdhury, “A System Identification Approach for Video-based Face Recognition”, *Pattern Recognition*, vol. 4, no. C, pp. 2–5, 2004.
- [151] R. Chellappa, V. Kruger, and S. Z. S. Zhou, “Probabilistic recognition of human faces from video” in *Computer Vision and Image Understanding*, vol. 91, pp. 214245, 2003.
- [152] L. F. Chen, H. Y. M. Liao, and J. C. Lin, “Person Identification using facial motion”, in *IEEE Proc. International Conference of Image Proceeding*, vol. 2, pp. 677–680, 2001.
- [153] K. S. Huang and M. M. Trivedi, “Streaming face recognition using multicamera video arrays”, in *Proc. of 16th International Conference on Pattern Recognition*, vol. 4, pp. 213–216, 2002.
- [154] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, “Visual tracking and recognition using probabilistic appearance manifolds”, *Computer Vision and Image Understanding*, vol. 99, no. 3, pp. 303–331, 2005.
- [155] X. Liu and T. Chen, “Video-based face recognition using adaptive hidden markov models”, *IEEE Proc. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 340–345, 2003.
- [156] F. Matta and J. L. Dugelay, “A Behavioural Approach to Person Recognition”, in *Proc. of IEEE International Conference on Multimedia and Expo*, pp. 1461 - 1464, 2006.

- [157] F. Matta and J.-L. Dugelay, “Person recognition using human head motion information”, in *International Conference on Articulated Motion and Deformable Objects (AMDO2006)*, vol. 4069, 2006.
- [158] U. Saeed, F. Matta, and J. L. Dugelay, “Person Recognition based on Head and Mouth Dynamics”, in *Proc. of IEEE 8th Workshop on Multimedia Signal Processing*, pp. 29–32, 2006.
- [159] Scharstein, D. and Szeliski, R., “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms”, *International Journal of Computer Vision*, vol. 47, no. 1, pp. 7–42, 2002.
- [160] L. PrimeSense, W. Garage, S. Production, <http://www.openni.org/> (2010).
- [161] G. Welch and G. Bishop, “The Kalman Filter,” last accessed in April 2012 in www.cs.unc.edu/~kalman, 2012.
- [162] J. Bassili, “Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face”, *Journal of Personality and Social Psychology*, vol. 37, pp. 20492059, 1979.
- [163] B. Knight and A. Johnston, “The role of movement in face recognition”, *Visual Cognition*, vol. 4, pp. 265274, 1997.
- [164] H. Hill and A. Johnston, “Categorizing sex and identity from the biological motion of faces”, *Current Biology*, vol. 11, no. 11, pp. 880885, 2001.
- [165] A. OToole, D. Roark and H. Abdi, “Recognizing moving faces: A psychological and neural synthesis”, *Trends in Cognitive Science*, vol. 6, pp. 261266, 2002.
- [166] L. Ballard, D. Lopresti and F. Monroe, “Forgery Quality and Its Implications for Behavioral Biometric Security”, *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 5 , pp. 1107–1118 , 2007.
- [167] B. L. B. Li and R. Chellappa, “A generic approach to simultaneous tracking and verification in video”, *IEEE Transactions on Image Processing*, vol. 11, no. 5, pp. 530–544, 2002.

- [168] N. Fox, R. Gross, J. Cohn, and R. Reilly, “Robust Biometric Person Identification Using Automatic Classifier Fusion of Speech, Mouth, and Face Experts”, *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 701 – 714, 2007.
- [169] Z. Sun and T. Tan, “Ordinal Measures for Iris Recognition”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2211, 2009.
- [170] N. Schmid, M. Ketkar, H. Singh, and B. Cukic, “Performance Analysis of Iris-Based Identification System at the Matching Score Level”, *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 1, pp. 154 – 168, 2006.
- [171] G. Zheng, C.-J. Wang, and T. Boulton, “Application of Projective Invariants in Hand Geometry Biometrics”, *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 4, pp. 758 – 768, 2007.
- [172] A. Jain and J. Feng, “Latent Palmprint Matching”, *IEEE Pattern Analysis and Machine Intelligence*, vol. 31, pp. 1032 – 1047, 2009.
- [173] K. Delac and M. Grgic, “A survey of biometric recognition methods”, in *Proceedings Elmar 2004. 46th International Symposium Electronics in Marine*, pp. 184–193, 2004.
- [174] H. Junker, J. Ward, P. Lukowicz, and G. Tröster, “User Activity Related Data Sets for Context Recognition”, in *Proc. Workshop on ‘Benchmarks and a Database for Context Recognition’*, pp. 1-6, 2004.
- [175] I. Bouchrika and M. Nixon, “Exploratory factor analysis of gait recognition, in proc. *IEEE Int. Conf. Autom. Face Gesture Recog.*, pp. 1-6, 2008.
- [176] M. Goffredo, I. Bouchrika, J. N. Carter, and M. S. Nixon, “Self-Calibrating View-Invariant Gait Biometrics”, *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 4 pp. 997–1008, 2010.
- [177] A. Hadid, M. Pietikäinen, and S. Z. Li, “Learning Personal Specific Facial Dynamics for Face Recognition from Videos”, in *Analysis and Modeling of Faces and Gestures*, pp. 1–15, Springer Berlin / Heidelberg, 2007.

- [178] D. Ioannidis, D. Tzovaras, I. G. Damousis, S. Argyropoulos, and K. Moustakas, “Gait Recognition Using Compact Feature Extraction Transforms and Depth Information”, *IEEE Trans. Inf. Forensics Security.*, vol. 2, no. 3, pp. 623–630, 2007.
- [179] M. Goffredo, I. Bouchrika, J. N. Carter, and M. S. Nixon, “Performance analysis for automated gait extraction and recognition in multi-camera surveillance”, *Multimedia Tools and Applications*, vol. 50, pp. 75–94, Oct. 2009.
- [180] A. Kale, N. Cuntoor, and R. Chellappa, “A framework for activity-specific human identification”, in Proc. of *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, pp. 3660–3663, 2002.
- [181] A. Kale, A. Sundaresan, A. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Kruger, and R. Chellappa, “Identification of Humans Using Gait”, *IEEE Trans. Image Processing*, vol. 13, pp. 1163–1173, 2004.
- [182] K. Moustakas, M. Strintzis, D. Tzovaras, S. Carbini, O. Bernier, J. Viallet, S. Raidt, M. Mancas, M. Dimiccoli, E. Yagci, S. Balci, and E. Leon, “Masterpiece: physical interaction and 3D content-based search in VR applications”, *IEEE Multimedia*, vol. 13, no. 3, pp. 92, 2006.
- [183] D. Tzovaras, G. Nikolakis, G. Fergadis, S. Malasiotis, and M. Stavrakis, “Design and Implementation of Haptic Virtual Environments for the Training of the Visually Impaired Engineering”, *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, no. 2, pp. 266–278, 2004.
- [184] V. Schoenefeld, “Spherical Harmonics”, 2005.
- [185] T. Flash and N. Hogan, “The coordination of arm movements: an experimentally confirmed mathematical model”, *Journal of neuroscience*, vol. 5, no. 7, pp. 1688–703, 1985.
- [186] Y. K. Uno and R. M. Suzuki, “Formation and Control of Optimal Trajectory in Human multijoint Arm Movement - minimum Torque-Change model”, *Biol. Cybern.*, vol. 61, no. 2, pp. 89–101, 1989.

- [187] M. Turvey, *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*. New Jersey: Lawrence Erlbaum, 1977.
- [188] S. R. Goodman and G. Gottlieb, “Analysis of kinematic invariances of multijoint reaching movement”, *Biol. Cybern.*, vol. 73, pp. 311–322, 1995.
- [189] B. Hoff and M. A. Arbib, “Models of Trajectory Formation and Temporal Interaction of Reach and Grasp”, *Motor Behaviour*, vol. 25, no. 3, pp. 175–192, 1993.
- [190] D. A. Rosenbaum, R. J. Meulenbroek, J. Vaughan, and C. Jansen, “Posture-based motion planning: Applications to grasping.”, *Psychological Review*, vol. 108, no. 4, pp. 709–734, 2001.
- [191] F. Lacquaniti and J. F. Soechting, “Coordination of arm and wrist motion during a reaching task.”, *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 2, pp. 399–408, Apr. 1982.
- [192] S. Jaric, D. Corcos, G. Gottlieb, D. Ilic, and M. Latash, “The effects of practice on movement distance and final position reproduction: implications for the equilibrium-point control of movements”, *Experimental Brain Research*, vol. 100, no. 2, pp. 353–359, 1994.
- [193] H. Peng, F. Long and C. Ding, “Feature selection based on mutual information: criteria of max-dependency, maxrelevance, and min-redundancy”, *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 27, no. 8, pp. 1226–1238, 2005
- [194] M. Wiesendanger, O. Kazennikov, S. Perrig, and P. Kaluzny, “Two handsone action: The problem of bimanual coordination”, in *Hand and Brain: Neurophysiology and psychology of hand movement* (W. A., H. P., and F. R., eds.), pp. 283–300, 1996.
- [195] G. Gomez and E. F. Morales, “Automatic feature construction and a simple rule induction algorithm for skin detection”, in *Proc. of the ICML Workshop on Machine Learning in Computer Vision (MLCV)*, pp. 31–38, 2002.

- [196] A. Bobick and J. Davis, “The recognition of human movement using temporal templates”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 3, pp. 257–267, 2001.
- [197] A. F. Bobick and A. Y. Johnson, “Gait recognition using static, activity-specific parameters”, in *IEEE Proc. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 423–430, 2001.
- [198] J. Jung, Z. Bien, S. Lee, and T. Sato, “Dynamic-footprint based person identification using mat-type pressure sensor”, *Engineering in Medicine and Biology Society*, vol. 3, pp. 2937– 2940, 2003.
- [199] M. G. Ryan, Visual Target Tracking, Patent US20100197399.
- [200] B. V. Wyk, M. V. Wyk, “Kronecker product graph matching”, *Pattern Recognition* vol. 36, no. 9, pp. 2019–2030, 2003.
- [201] P. Viola and M. Jones, “Rapid Object Detection using a Boosted Cascade of Simple”, in *IEEE Proc. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 511–518, 2001.
- [202] A. Hadid and M. Pietikainen, “Combining appearance and motion for face and gender recognition from videos”, *Pattern Recognition*, vol. 42, no. 11, pp. 28182827, 2009.
- [203] M. Tistarelli, M. Bicego and E. Grosso, “Dynamic face recognition: From human to machine vision”, *Image Vision Comput*, vol. 27, no. 3, pp. 222232, 2009.
- [204] J. Choi, K. Plataniotis, Fellow, Y. Ro, “Face Feature Weighted Fusion Based on Fuzzy Membership Degree for Video Face Recognition”, *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 4, pp. 1270–1282, 2012
- [205] M. Brand, N. Oliver, and A. Pentland, “Coupled hidden Markov models for complex action recognition”, in *Computer Vision and Pattern Recognition (CVPR)*, pp. 994–994, 1997.

- [206] L. Xie, S. F. Chang, A. Divakaran, and H. Sun, “Structure analysis of soccer video with hidden Markov models”, in *IEEE Inter. Conf. Acoustic, Speech, Signal Processing*, pp. 767–775, 2002.
- [207] T. Chen, C. Huang, C. Chang, and J. Wang, “On the use of Gaussian mixture model for speaker variability analysis”, in *Int. Conf. SLP*, 2002.
- [208] C. M. Bishop, *Pattern recognition and machine learning*. Springer Verlag, ed. 1, 2006.
- [209] O. Miguel-Hurtado, L. Mengibar-Pozo and A. Pacut., “A new algorithm for signature verification system based on DTW and GMM.”, in *IEEE International Carnahan Conference on Security Technology (ICCST)*, pp. 206 – 213, 2008.
- [210] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm”, *Journal of Royal Statistical Society*, vol. 39, pp. 1–38, 1977.
- [211] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition”, *Proceedings of the IEEE*, vol. 53, no. 3, pp. 257–286, 1989.
- [212] D. C. V. Ramesh and P. Meer, “Real-Time Tracking of Non-Rigid Objects Using Mean Shift”, in *IEEE Proc. Computer Vision and Pattern Recognition 2007 (CVPR)*, vol. 2, pp. 142–149, 2000.
- [213] D. Mahajan, R. Ramamoorthi, and B. Curless, “A Theory Of Frequency Domain Invariants: Spherical Harmonic Identities for BRDF/Lighting Transfer and Image Consistency”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 197 – 213, 2008.
- [214] V. Schönfeld, “Spherical Harmonics”, tech. rep., 2005.
- [215] S. Wu and Y. F. Li, “Flexible signature descriptions for adaptive motion trajectory representation, perception and recognition”, *Pattern Recognition*, vol. 42, pp. 194–214, Jan. 2009.

- [216] A. Vogianou, K. Moustakas, D. Tzouvaras, and M. Strintzis, “Advances in Biometrics”, *Lecture Notes in Computer Science* vol. 5558, Springer Berlin Heidelberg, 2009.
- [217] J. Aleotti and S. Caselli, “Grasp Recognition in Virtual Reality for Robot Pregrasp Planning by Demonstration”, in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2801–2806, 2006.
- [218] E. Kukula and S. Elliott, “Implementation of hand geometry: an analysis of user perspectives and system performance”, in *IEEE Aerospace and Electronic Systems Magazine*, pp. 3–9, 2006.
- [219] M. Lexa, “Useful Facts About the Kullback-Leibler Discrimination Distance”, tech. rep., Department of Electrical and Computer Engineering, Rice University, Houston.
- [220] G. Marcialis, F. Roli, and D. Muntoni, “Group-specific face verification using soft biometrics”, *Journal of Visual Languages and Computing*, vol. 20, pp. 101–109, 2009.
- [221] A. Drosou, K. Moustakas, D. Ioannidis, and D. Tzouvaras, “Activity Related Biometrics based on Motion Trajectories”, in *BIOSIG 2010: Biometrics and Electronic Signatures* (A. Broemme and C. Busch, eds.), pp. 127–132, GI Edition, 2010.
- [222] A. Jain, S. Dass and K. Nandakumar, “Soft Biometric Traits for Personal Recognition Systems”, in *International Conference on Biometric Authentication, (LNCS)*, (Hong Kong), pp. 731–738, 2004.
- [223] G. H. Ball and D. J. Hall, “ISODATA, A novel method of Data Analysis and Pattern Classification”, tech. rep., Menlo Park, California, Stanford Research Institute, 1965.
- [224] G. McLachlan and T. Krishnan, “The EM Algorithm and Extensions”, *Journal of Classification*, vol. 15, no. 1, pp. 154–156, 1997.
- [225] M. Petrou and C. Petrou, *Image Processing: The Fundamentals, 2nd Edition*. 2010.
- [226] M. S. Nixon, T. Tan, and R. Chellappa, *Human Identification Based on Gait*, 2006.

- [227] T. T. Ltd., “Procomp5 Infinity”, 2011.
- [228] Microsoft Corporation, “Kinect for Windows”, 2012.
- [229] G. Chanel, J. J. Kierkels, M. Soleymani, and T. Pun, “Short-term emotion assessment in a recall paradigm”, *International Journal of Human-Computer Studies*, vol. 67, pp. 607–627, Aug. 2009.
- [230] C. Setz, B. Arnrich, J. Schumm, R. L. Marca, G. Tr, and U. Ehlert, “Using a Wearable EDA Device”, *Technology*, vol. 14, no. 2, pp. 410–417, 2010.
- [231] K. Kim, S. Bang, and S. Kim, “Emotion Recognition System Using Short-Term Monitoring of Physiological Signals”, *Medical and Biological Eng. and Computing*, vol. 42, pp. 419–427., 2004.
- [232] R. L. Mandryk and M. S. Atkins, “A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies”, *International Journal of Human-Computer Studies*, vol. 65, pp. 329–347, Apr. 2007.
- [233] J. A. Healey and R. W. Picard, “Detecting Stress During Real-World Driving Tasks Using Physiological Sensors”, *Transportation*, vol. 6, no. 2, pp. 156–166, 2005.
- [234] S. Ruan, L. Chen, J. Sun, and G. Chen, “Study on the Change of Physiological Signals during Playing Body-Controlled Games”, in *Human Factors*, pp. 349–352, 2009.
- [235] T. Kiryu, I. Sasaki, K. Shibai, and K. Tanaka, “Providing appropriate exercise levels for the elderly.”, *IEEE engineering in medicine and biology magazine : the quarterly magazine of the Engineering in Medicine & Biology Society*, vol. 20, no. 6, pp. 116–24, 2001.
- [236] J. T. K. Jr, K. Pollock and A. Safonova, “A Data-driven Appearance Model for Human Fatigue”, in *Eurographics/ ACM Symposium on Computer Animation (SIGGRAPH)*, 2011.
- [237] O. Miguel-Hurtado, L. Mengibar-Pozo and A. Pacut “A new algorithm for signature verification system based on DTW and GMM”,

- in *IEEE International Carnahan Conference on Security Technology (ICCST'08)* vol. 42, pp. 206–213, 2008.
- [238] J. Daemen and V. Rijmen, “The Rijndael Block Cipher”, 1999.
- [239] W. Boucsein, *Electrodermal Activity*. Springer, 2012.
- [240] E. Mordini, “Biometrics, Human Body, and Medicine: A Controversial History”, *Biometrics, Human Body, and Medicine: A Controversial History*, P. Duquenoy, C. George, K. Kimppa (Eds.), IGO Global, vol.11, 2008.
- [241] J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, eds., *Handbook of Psychophysiology*. Cambridge University Press, 2000.
- [242] K. Xi and J. Hu, “Biometric Mobile Template Protection: A Composite Feature Based Fingerprint Fuzzy Vault”, in *IEEE International Conference on Communications (ICC '09)*, pp.1-5, 2009.
- [243] A. R. Jensen and R. W. D., “The Stroop colorword test: A review.”, *Acta Psychologica*, vol. 25, pp. 36–93, 1966.
- [244] P. Karthikeyan, M. Murugappan, and S. Yaacob, “A Review on Stress Inducement Stimuli for Assessing Human Stress Using Physiological Signals”, in *Blood Pressure*, pp. 420–425, 2011.
- [245] I. Damousis, A. Bakirtzis, and P. Dokopoulos, “Network-constrained economic dispatch using real-coded genetic algorithm”, *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 198 – 205, 2003.
- [246] J. Han and B. Bhanu, “Individual recognition using gait energy image”, *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, pp. 316–22, Feb. 2006.
- [247] P. Salembier, A. Oliveras, and L. Garrido, “Anti-extensive connected operators for image and sequence processing”, *IEEE Transactions on Image Processing*, vol. 7, no. 4, pp. 555–570.
- [248] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, “Improving shadow suppression in moving object detection with HSV color information”, *Intelligent Transportation Systems*, pp. 334–339, 2001.

- [249] X. Han, J. Liu, L. Li, and Z. Wang, “Gait recognition considering directions of walking”, in *Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems*, pp. 1–5, 2006.
- [250] C. Yu, H. Cheng, C. Cheng, and K.-C. Fan, “Efficient Human Action and Gait Analysis Using Multiresolution Motion Energy Histogram”, *Journal on Advances in Signal Processing (EURASIP)*, pp. 13, 2010.
- [251] S. Berretti, A. D. Bimbo, I. C. Society, and P. Pala, “3D Face Recognition Using Isogeodesic Stripes”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2162–2177, 2010.
- [252] C. S. LLC., “Immersion Web Page in <http://www.vrlogic.com/html/immersion/cyberglove.html>, last accessed in April 2012.”
- [253] Y. Z. Y. Zhang and D. W. D. Wang, “Research on object-storage-based intrusion detection”, in *Proc. of 12th International Conference on Parallel and Distributed Systems (ICPADS '06)*, vol. 1, pp. 68-78, 2006.
- [254] R. V. Yampolskiy and V. Govindaraju, “Strategy-based behavioural biometrics: a novel approach to automated identification”, *IJCAT*, vol. 35, no. 1, pp. 29–41, 2009.
- [255] R. N. J. Veldhuis, A. M. Bazen, J. A. Kauffman, and P. H. Hartel, “Biometric verification based on grip-pattern recognition (invited paper)”, in *ISTSPIE 16th Annual Symp on Electronic Imaging Security Steganography and Watermarking of Multimedia Contents* (E. J. D. III and P. W. Wong, eds.), vol. 5306, pp. 634–641, SPIE – The Int. Society for Optical Engineering, Washington, 2004.
- [256] M. Orozco, Y. Asfaw, A. Adler, S. Shirmohammadi, and A. E. Sadiq, “Automatic Identification of Participants in Haptic Systems”, in *Proc. of IEEE Instrumentation and Measurement Technology Conference*, vol. 2, pp. 888 - 892, 2005.
- [257] D. Novikov, R. V. Yampolskiy, and L. Reznik, “Artificial Intelligence Approaches for Intrusion Detection”, in *Proc. of IEEE Long Island*

- Systems, Applications and Technology Conference (LISAT'06)*, pp. 1 - 8, 2006.
- [258] C. C. Michael, “Finding the vocabulary of program behavior data for anomaly detection”, in Proc. of *DARPA Information Survivability Conference and Exposition*, vol. 1, pp. 152–163 2003.
- [259] H. L. H. Lei, S. Palla, and V. Govindaraju, “ER2: an intuitive similarity measure for on-line signature verification”, in Proc. of *9th International Workshop on Frontiers in Handwriting Recognition (IWFHR-9'04)*, pp. 191 - 195, 2004.
- [260] N. Henderson, N. White, P. Hartel, R. Veldhuis, and K. Slump, “Sensing pressure for authentication”, *3rd IEEE Benelux Signal Processing Symp SPS*, pp. 241–244, 2002.
- [261] J. T. Giffin, S. Jha, and B. P. Miller, “Efficient Context-Sensitive Intrusion Detection”, *Construction*, 2004.
- [262] A. Garg, R. Rahalkar, S. Upadhyaya, and K. Kwiat, “Profiling Users in GUI Based Systems for Masquerade Detection” in Proc. of *IEEE Information Assurance Workshop*, pp. 48 - 54, 2006.
- [263] G. Frantzeskou, S. Gritzalis, and S. G. Macdonell, “Source code authorship analysis for supporting the cybercrime investigation process”, in Proc. of *1st International Conference on e-business and Telecommunications Networks (ICETE04)*, pp. 85–92, 2004.
- [264] H. H. Feng, O. M. Kolesnikov, P. Fogla, W. Lee, and W. G. W. Gong, “Anomaly detection using call stack information”, in Proc. of *Symposium on Security and Privacy*, pp. 62–75, 2003.
- [265] E. Erzin, Y. Yemez, A. M. Tekalp, A. Ercil, H. Erdogan, and H. Abut, “Multimodal person recognition for human-vehicle interaction”, in *IEEE Trans. Multimedia*, vol. 13, no. 2 pp. 18 - 31, 2006.
- [266] H. Erdo Ugan, A. Erçil, H. K. Ekenel, S. Y. Bilgin, I. Eden, M. Kirişçi, and H. Abut, “Multi-modal person recognition for vehicular applications”, *Lecture Notes in Computer Science*, vol. 3541, pp. 366–375, 2005.

- [267] R. Brause, T. Langsdorf, and M. Hepp, “Neural data mining for credit card fraud detection”, 1999.
- [268] A. A. E. Ahmed and I. Traore, “Anomaly intrusion detection based on biometrics”, 2005.
- [269] A. A. E. Ahmed and I. Traore, “Detecting computer intrusions using behavioral biometrics”, *Security*, 2005.
- [270] Y. Li, N. Wu, S. Jajodia and X.S. Wang, “Enhancing profiles for anomaly detection using time granularities,”, *Journal of Computer Security*, 2002.
- [271] C. Varenhorst, “Passdoodles: a lightweight authentication method”, retrieved from <http://people.csail.mit.edu/emax/papers/varenhorst.pdf>, 2004.
- [272] T. Westeyn, P. Pesti, K. Park and T. Starner, “Biometric identification using song-based eye blink patterns”, *Human Computer Interaction International (HCII)*, 2005.
- [273] C. Hilaris and J. Sahalos, “User profiling for fraud detection in telecommunication networks,” 5th *International Conference on Technology and Automation (ICTA 2005)*, Greece, 2005.
- [274] T. Goldring, “User profiling for intrusion detection in Windows NT,”, *Computing Science and Statistics*, vol. 35, 2003.
- [275] A. Toney, B. H. Thomas, “Considering Reach in Tangible and Table Top Design,” 1st *IEEE international workshop on horizontal interactive human-computer systems*, pp. 2, 2006.
- [276] S. Jahanbin, C. Hyohoon and A.C. Bovik, “Passive Multimodal 2-D+3-D Face Recognition Using Gabor Features and Landmark Distances,”, *IEEE Trans.Inf. Forensics Security*, vol. 6, no. 4, pp.1287–1304, 2011.
- [277] M.D. Cordea, E.M. Petriu and D.C. Petriu, “Three-Dimensional Head Tracking and Facial Expression Recovery Using an Anthropometric Muscle-Based Active Appearance Model,” in *IEEE Trans. Instrum. Meas.*, vol. 57, no. 8, pp. 1578–1588, 2008.

- [278] K. Kaufman, G. Cervone and R.S. Michalski, “An application of symbolic learning to intrusion detection: preliminary results from the LUS methodology”, *Reports of the Machine Learning and Inference Laboratory*, MLI 03-2, George Mason University, 2003.
- [279] J-F. Mainguet, “Biometrics”, retrieved July 28, 2006 from <http://perso.orange.fr/fingerchip/biometrics/biometrics.htm>, 2006.
- [280] S. Pamudurthy, E. Guan, K. Mueller and M. Rafailovich, “Dynamic approach for face recognition using digital image skin correlation”, *Audio and Video-based Biometric Person Authentication (AVBPA)*, New York, 2005.
- [281] S.J. Stolfo, S. Hershkop, K. Wang, O. Nimeskern and C-W. Hu, “A behaviour-based approach to securing e-mail systems”, *Mathematical Methods, Models and Architectures for Computer Networks Security*, Springer Verlag, 2003.
- [282] S.J. Stolfo, C-W. Hu, W-J. Li, S. Hershkop, K. Wang and O. Nimeskern, “Combining behaviour models to secure e-mail systems”, *CU Tech Report*, retrieved from <http://www1.cs.columbia.edu/ids/publications/EMT-weijen.pdf>, 2003.
- [283] R.V. Yampolskiy and V. Govindaraju, “Use of behavioural biometrics in intrusion detection and online gaming”, *Biometric technology for human identification III*, SPIE Defense and Security Symposium, Orlando, 2006.
- [284] R.V. Yampolskiy and V. Govindaraju, “Dissimilarity functions for behaviour-based biometrics,”, *Biometric Technology for Human Identification IV*, SPIE Defense and Security Symposium, Orlando, 2007.
- [285] M. Orozco, Y. Asfaw, S. Shirmohammadi, A. Adler and A.E. Saddik, “Haptic-based biometrics: a feasibility study,”, *IEEE Virtual Reality Conference*, Alexandria, Virginia, 2006.
- [286] J. Ilonen, “Keystroke dynamics,”, retrieved July 12, 2006) from www.it.lut.fi/kurssit/03-04/010970000/seminars/Ilonen.pdf, 2006.
- [287] S.D. Bella and C. Palmer, “Personal identifiers in musicians”, *Finger movement dynamics*, *Journal of Cognitive Neuroscience*, vol. 18, 2006.

- [288] O. Shipilova, “Person recognition based on lip movements”, retrieved July 15, from <http://www.it.lut.fi/kurssit/03-04/010970000/seminars/Shipilova.pdf>, 2006.
- [289] S. Lyu, D. Rockmore and H. Farid, “A digital technique for art authentication”, *Proceedings of the National Academy of Sciences*, 2004.
- [290] F. Apap, A. Honig, S. Hershkop, E. Eskin and S. Stolfo, “Detecting malicious software by monitoring anomalous windows registry accesses”, *Technical Report*, CUCS Technical Report, 2001.
- [291] S. Kalyanaraman, “Biometric authentication systems: a report”, retrieved July 26, 2006 from <http://netlab.cs.iitm.ernet.in/cs650/2006/TermPapers/sriramk.pdf>, 2006.
- [292] Z. Ciota, “Speaker verification for multimedia application”, *IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2752–2756, 2004.
- [293] K. Wheaton, J. Thompson, A. Syngeniotes, D. Abbott and A. Puce, “Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex”, *Neuroimage* vol. 22, no. 1 pp. 277–288, 2004.
- [294] J. Thompson, M. Clarke, T. Stewart and A. Puce, “Configural processing of biological motion in human superior temporal sulcus”, *The Journal of Neuroscience*, vol. 25, pp. 9059-9066, 2005.
- [295] E. Bonda, M. Petrides, D. Ostry and A. Evans, “Specific involvement of human parietal systems and the amygdala in the perception of biological motion”, *The Journal of Neuroscience*, vol. 16, pp.3737-3744, 1996.
- [296] F. Loula, S. Prasad, K. Harber and M. Shiffrar, “Recognizing People From Their Movement”, *Journal of Experimental Psychology*, vol. 31, no. 1, pp. 210-220, 2005.
- [297] G. Johansson, “Visual perception of biological motion and a model for its analysis”, *Perception and Psychophysics*, vol. 14, pp. 201-211, 1973.

- [298] G. Bingham, R. Schmidt, L. Rosenblum, “Dynamics and the Orientation of Kinematic Forms in Visual Event Recognition”, *Journal of Experimental Psychology: Human Perception and Performance*, vol. 21, no. 6, pp. 1473–1493, 1995.
- [299] S. Stevenage, M. Nixon and K. Vince, “Visual analysis of gait as a cue to identity”, *Applied Cognitive Psychology*, vol. 13, no. 6, pp. 513-526, 1999.
- [300] W. Wong and E. Rogers, “Recognition of Temporal Patterns: From Engineering to ”, *Psychology and Back Again*, vol. 61, no. 2, pp.159–167, 2007.
- [301] S. Brownlow, A. Dixon, C. Egbert and R. Radcliffe, “Perception of movement and dancer characteristics from point-light displays of dance”, *Psychological Record*, vol. 47, pp.411-421, 1997.
- [302] W. Dittrich, T. Troscianko, S. Lea and D. Morgan, “Perception of emotion from dynamic point-light displays represented in dance”, *Perception*, vol. 25, pp. 727-738, 1996.
- [303] S. Runeson and G. Frykholm, “Kinematic specification of dynamics as an informational bias for person-and-action perception: Expectation, gender recognition, and deceptive intent”, *Journal of Experimental Psychology: General*, vol. 112, pp. 585-615, 1983.
- [304] G. Mather and L. Murdoch, “Gender discrimination in biological motion displays based on dynamic cues”, *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 258, pp. 273–279, 1994.
- [305] N. Ambady, M. Hallahan and B. Conner, “Accuracy of judgments of sexual orientation from thin slides of behavior”, *Journal of Personality and Social Psychology*, vol. 77, pp. 538-547, 1999
- [306] M. Shiffrar and J. Pinto, “The visual analysis of bodily motion. In W. Prinz and B. Hommel (Eds.)”, *Common mechanisms in perception and action: Attention and performance XIX Oxford*, England: Oxford University Press, pp. 381-399, 2002.

- [307] M. Giese and T. Poggio, “Neural mechanisms for the recognition of biological movements”, *Nature Reviews Neuroscience*, vol.4, 179-192, 2003.
- [308] J. Cutting and L. Kozlowski, “Recognizing friends by their walk: Gait perception without familiarity cues”, *Bulletin of the Psychonomic Society*, vo;. 9, pp. 353-356, 1977.
- [309] F. Pollick, J. Kay, K. Heim and R. Stringer, “A review of gender recognition from gait”, *Perception ECVF Abstracts*, 2002.
- [310] T. Beardsworth and T. Buckner, “The ability to recognize oneself from a video recording of ones movements without seeing ones body”, *Bulletin of the Psychonomic Society*, vol. 18, pp. 19-22, 1981.
- [311] W. Prinz, “Perception and action planning. European Journal of Cognitive Psychology”, vol. 9, pp. 129-154, 1997.
- [312] M. Wilson, “Perceiving imitable stimuli: Consequences of isomorphism between input and output”, *Psychological Bulletin*, vol. 127, pp. 543-553, 2001.
- [313] H.V. Halteren, “Linguistic profiling for author recognition and verification”, *Proceedings of ACL*, 2004.
- [314] L. Brown, C. Moore and D. Rosenbaum, “Feature-Specific Perceptual Processing Dissociates Action From Recognition”, *Journal of Experimental Psychology: Human Perception and Performance*, vol. 28, no. 6, pp.1330-1344, 2002
- [315] A. Churchill, B. Hopkins, L. Roennqvist and S. Vogt, “Vision of the hand and environmental context in human prehension”, *Exp Brain Res*, vol. 134, pp.81-89, 2000.
- [316] J. Vaughan, D. Rosenbaum and R. Meulenbroek, “Planning Reaching and Grasping Movements : The Problem of Obstacle Avoidance”, *Motor Control*, vol. 2, pp. 116–135, 2001.
- [317] D. Rosenbaum, C. Coelho, J. Rhode and J. Santamaria, “Psychologically Distinct Classes of Motor Behavior Inferred from Individual

- Differences: Evidence from a Sequential Stacking Task”, *Journal of Motor Behavior*, vol. 42, no. 3, 2010.
- [318] M. Oram and D. Perrett, “Responses of Anterior Superior Temporal Polysensory Neurons to Biological Motion Stimuli”, *Journal of Cognitive Neuroscience*, vol. 6, no. 2, pp. 99-116, 1994.
- [319] E. Grossman, M. Donnelly, R. Price, D. Pickens, V. Morgan, G. Neighbor and R. Blake, “Brain areas involved in perception of biological motion”, *Journal of Cognitive Neuroscience*, vol. 12, no. 5, pp. 711-720, 2000.
- [320] T. Allison, A. Puce and G. McCarthy, “Social perception from visual cues: role of the STS region”, *Trends in Cognitive Sciences*, vol. 4, no. 7, pp. 267-278, 2000.
- [321] G. Rizzolatti, L. Fogassi and V. Gallese, “Neurophysiological mechanisms underlying the understanding and imitation of action”, *Nature Reviews Neuroscience*, vol. 2, no. 9, pp. 661-670, 2001.
- [322] N. Troje, “Decomposing biological motion: A framework for analysis and synthesis of human gait patterns”, *Journal of Vision*, vol. 2, pp. 371-387, 2002.
- [323] F. Pollick, “The Features People Use to Recognize Human Movement Style”, *Gesture-Based Communication in Human-Computer Interaction*, vol. 2915, pp. 10–19, 2004.
- [324] Y. Zhang and D. Wang, “Research on object storage-based intrusion detection”, in Proc. of 12th *International Conference on Parallel and Distributed Systems (ICPADS)*, vol. 1, pp. 68-78, 2006.
- [325] N. Muralidharan and S. Wunnava, “Signature verification: a popular biometric technology”, *Second LACCEI International Latin American and Caribbean Conference for Engineering and Technology (LACCEI 2004)*, 2004.
- [326] S. Argyropoulos, D. Tzovaras, D. Ioannidis, and M. G. Strintzis, “A Channel Coding Approach for Human Authentication From Gait Sequences”, *IEEE Trans. Inf. Forensics Security.*, vol. 24, no. 3, pp. 428–440, 2009.

- [327] T. Wadayama, “An authentication scheme based on a low-density parity check matrix”, in *Int. Symp. Information Theory*, pp. 2266–2269, 2005.
- [328] G. I. Davida, Y. Frankel, and B. J. Matt, “On enabling secure applications through off-line biometric identification”, in *IEEE Symp. Security and Privacy*, (Oakland, CA), pp. 148–157, 1998.
- [329] A. Juels and M. Sudan, “A fuzzy vault scheme”, *Designs Codes Cryptography*, vol. 38, no. 2, pp. 237–257, 2006.
- [330] J. D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources”, *IEEE Trans. Inf. Theory*, vol. 19, pp. 471–480, 1973.
- [331] E. Martinian, S. Yekhanin, and J. Yedidia, “Secure biometrics via syndromes”, in *43rd Annual Allerton Conf. on Communications, Control, and Computing*, pp. 1–11, 2005.
- [332] S. C. Draper, A. Khisti, E. Martinian, A. Vetro, and J. Yedidia, “Using distributed source coding to secure fingerprint biometrics”, in *Int. Conf. Acoustics, Speech and Signal Processing*, (Honolulu, HI), pp. 129–132, 2007.
- [333] L. Yao-Chung, D. Varodayan, and B. Girod, “Image authentication based on distributed source coding”, in *Int. Conf. on Image Processing*, (San Antonio, TX), pp. 5–8, 2007.
- [334] G. Cohen and G. Zemor, “Generalized coset schemes for the wire-tap channel: application to biometrics”, in *IEEE International Carnahan Conference on Security Technology (ICCST) Symposium on Information Theory*, (Chicago, IL), p. 46, 2004.
- [335] S. S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (discus): Design and construction”, *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 626–643, 2003.
- [336] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, “Distributed video coding”, *Proc. IEEE*, vol. 93, pp. 71–89, 2005.
- [337] W. Stallings, *Cryptography and Network Security: Principles and Practices*. 2006.

- [338] F. H. Álvarez, L. H. Encinas, and C. S. Ávila, “Biometric Fuzzy Extractor Scheme for Iris Templates”, in *World Congress in Computer Science, Computer Engineering and Applied Computing*, 2009.
- [339] Z. Zuang, and B. Bradtmiller “Head-and-Face Anthropometric Survey of U.S. Respirator Users”, *Jrnl. of Occupat. and Environ. Hygiene*, vol. 2, pp. 567–576, 2005.
- [340] R. Koekoek and R. F. Swarttouw, “The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue”, *Delft University of Technology, Faculty of Information Technology and Systems, Department of Technical Mathematics and Informatics*, tech. report no. 98-17, 1998.

A. Vision-based UpperBody Tracking Algorithm

The movements of the users are recorded by a depth enabled camera and the raw captured images are processed in order to track the their head and hands via the successive application of filtering masks on the captured image. Specifically, a skin-colour mask (see Section A.2) combined with a motion-mask (section A.4) can provide the location of the palms, while the head can be accurately tracked via a combination of a head detection algorithm combined with an object tracking algorithm (Section A.1). The 3D spatial information can be easily retrieved from the provided depth images, while the temporal information is derived from the timestamps at the moment each frame is recorded.

A.1. Face Detection - Tracking

The detection of a face in an image is the first step towards the behavioural tracking of a user. The face detection problem has been a major issue in the fields of image processing and computer vision for the last decades. At the core, face detection requires an effective discrimination function between facial and non-facial patterns. Generally speaking, there are nowadays mainly two methodologies for a face detection task:

- Knowledge-based methods [69] attempt to describe all the face patterns using rules based on human knowledge such as the fact that all faces have two eyes and a mouth. However, they are difficult to use to detect faces in real images as the translation of human knowledge into well formed rules is nontrivial. If the rules are too restrictive, many faces will be ruled out, resulting in false negatives; on the other hand, if the rules are too general, non-facia patterns will be included in the face class, resulting in false positives.

- The learning-based methodology, examples of which are Osuna et al.'s SVM method [70] and Kanade's Bayesian-rule method [71], tries to model the face pattern with distribution functions or discriminant functions under the probabilistic framework. Methods of this kind are not limited by our describable knowledge on faces but determined by the capability of learning model and training samples, hence being able to deal with more complex cases compared with the knowledge-based approach. Specifically, the breakthrough of learning-based methodology happened in 2001 when Viola and Jones proposed a novel boosted cascade framework [201]. This work showed amazing real-time speed and high detection accuracy, due to the fast calculation of Haar-like features via the integral image and the cascade structure of classifiers learned by AdaBoost.

Thus, the aforementioned algorithm of Viola and Jones has been implemented and further enhanced within the framework of the current research. Face detection is a rather difficult task due to the variability of the object of interest itself and the environment. In particular, the following requirements need to be considered [72] for a robust face detector:

- Size: A face detector should be able to detect faces in different sizes.
- Position: The detection of faces at different positions within the image is usually achieved by sliding a window over the image and applying the detection step at each window position.
- Number: An important issue here is to handle partially overlapping faces. The standard way to solve this problem is to apply a post-filter to remove multiple overlapping faces and derive a single representative face.
- Expressions: The changes in the appearance of a face for different facial expressions are usually considered within the training process of the face detector.
- Orientation: Faces can appear in different orientations within the image plane depending on the angle of the camera and the face.
- Illumination: Varying illumination and shadows can cause big problems to face detection since they change the color and the appearance

of the face depending on the color and the direction of the light.

The first four requirements, namely the Size, Position, Number and Expression requirements, are explicitly handled by the Viola and Jones face detector (see Section A.1.1). The Orientation and Illumination requirements have been further augmented by the Mean Shift algorithm (see Section A.1.2) and a skin detection algorithm (Section A.2), as it will be explained next.

A.1.1. AdaBoost learning and cascade structure of classifiers

The Viola and Jones face detector is based on the AdaBoost algorithm. AdaBoost is a general method of combining an ensemble of “weak classifiers” whose accuracy may be poor, but still better than random guess. Given a set T of weak classifiers, a “strong classifier” is obtained as a weighted linear combination of the weak classifiers as follows:

$$H(F) = \begin{cases} 1, & \text{if } \sum_{t=1}^T a_t h_t(I) \geq \delta \\ 0, & \text{otherwise} \end{cases} \quad (\text{A.1})$$

where F is the input image F , $h_t(F)$ is a weak classifier, a_t is the t^{th} corresponding weight, and δ represents a threshold value as it is calculated in [201].

In the following, the weak classifiers are selected from a large number of features, computed inside rectangular windows and treated as individual weak classifiers. For simplicity we call these features hereafter “rectangle features” Each weak classifier consists of a rectangle feature $f_t(F)$, a parity p_t that indicates the direction of the inequality sign and a threshold θ_t :

$$h_t(F) = \begin{cases} 1, & \text{if } p_t f_t(I) < p_t \theta_t \\ 0, & \text{otherwise} \end{cases} \quad (\text{A.2})$$

To find the best weak classifier in every boosting round, an exhaustive search is employed as in [73]. In Figure A.1 six types of rectangle feature are shown, that have been used in this study.

Types (a)-(e) are similar to basic Haar-like features proposed by Viola and Jones [201]. These features are computed by subtracting the sum of the pixel values in the dark rectangle from the sum of the pixel values in

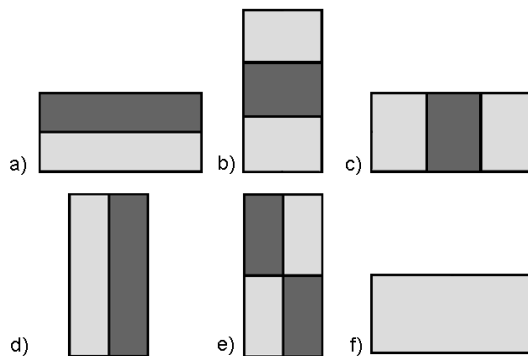


Figure A.1.: Types of rectangle features.

the bright rectangle. Specifically, the value of a *two-rectangle feature* is the difference between the sum of the pixels within two rectangular regions. The regions have the same size and shape and are horizontally or vertically adjacent (Figures A.1a & A.1d). A *three-rectangle feature* computes the sum within the two outside rectangles and subtracts it from the sum in a central rectangle (Figures A.1b & A.1c). Finally a *four-rectangle feature* computes the difference between diagonal pairs of rectangles (Figure A.1e).

Additionally, type (f) calculates the variance value of the pixels inside the rectangle (Figure A.1f). Thus, type (f) indicates the variance feature which expresses second-order statistics in the given region [74]. By utilizing second order statistics more information is available to distinguish the face pattern from the non-face pattern (Figure A.2a).

Further, the rectangle features can be calculated very rapidly using an intermediate representation for the image, namely the integral image [73]:

$$IF(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y') \quad (\text{A.3})$$

where $IF(x, y)$ is the integral image and $F(x, y)$ is the original image.

After calculating the feature values, they are all normalized to minimize the effect of illumination conditions, except type (f). Normalization is simply performed by dividing with the variance of the whole window.

Among the millions of possible subwindows, only very few subwindows are classified as a face. Window scanning techniques are used in most view-based detection methods. The potential frequency of faces and non-faces should be considered for real-time performance.

Further, the cascade structure of a classifier is a good framework for im-

plementing a fast face detector. Specifically, at each stage, a strong classifier is trained to pass almost all face training data, while discarding a certain portion (typically between 0.2 and 0.5) of non-face training data. The learning goal for the i^{th} stage is to satisfy detection rate (p_i) and false positive rate (q_i), i.e., the detection rate in each stage should be greater than or equal to p_i , and the false positive rate in each stage should be less than or equal to q_i .

Feature selection is performed until the strong classifier satisfies the learning goal. After a stage classifier is trained, a new negative training data set is collected for the next stage and the AdaBoost algorithm is applied in the same way.

In particular, given the current stage number s , the current stage classifier H_s can be organized by adding the previous stage classifier as follows for $s > 1$:

$$h(F) = \begin{cases} 1, & \text{if } a_0 H'_{s-1}(F) + \sum_{t=1}^T a_t h_t(F) > \delta_s \\ 0, & \text{otherwise} \end{cases} \quad (\text{A.4})$$

where H'_{s-1} is the previous stage classifier with new threshold δ'_{s-1} and a_0 is a corresponding weight coefficient found by the AdaBoost algorithm. By changing the threshold value from δ_{s-1} to δ'_{s-1} , the previous stage classifier is employed with no additional computational cost. Since every feature value is already computed at the previous stage, only the last threshold needs to be compared to the new one, and replaced, if different.

A.1.2. Head Tracking

The continuous tracking of a person's face is of vital importance for biometric monitoring of behaviour. The simplest idea to implement is the successive face detection in each frame. However, despite the maturity of frontal face detection, it is often inadequate to meet the rigorous requirements of general applications (e.g., visual surveillance systems, digital equipments that need autofocus on faces, etc.), since human faces in real-life images are seldom upright and frontal. Thus, there have been many works in recent years that developed new methods to enhance Viola and Jones' framework in various respects.

For instance, the detector structure has been extended to Wu et al.'s nesting cascade model [86] who transformed Viola and Jones' loose cascade

model into a more compact one. Moreover, a finer partition of the feature space was adopted in order to fit likelihoods more precisely through finer partition granularity [87], as exhibited by the Wuet et al.'s piece-wise function [86], and Mita et al.'s joint binarization of Haar-like features [88]. As for the type of feature, there have been the works of Liu and Shum's Kullback-Leibler features [155] and Baluja's pair-wise points [89].

In the framework of the current research, the Viola and Jones algorithm [73] has been extended by utilizing the Mean Shift algorithm. The idea behind this extension lies in the fact that each time the implemented face detector fails, the chromatic information of the last successful detected face rectangle will be passed over to a rigid object tracker. Thus, the face will be tracked as a simple object, using just the histogram information of the rectangle. In all other cases, the tracking of the face exclusively depends on the detection of the face on the successive images of the recorded sequence (Figure A.2a).

In the following, a short description of the Mean-Shift algorithm [212] is presented. Assuming that $\{x_i^*\}_{i=1,\dots,n}$ represent the pixel locations centered at $x_0 = 0$, function b can be defined: $R^2 \rightarrow 1, \dots, m$ which associates to the pixel at location $F(x, y)$ the index $b(F(x, y))$ of the histogram bin corresponding to the color of that pixel. The probability of a colour u in the target model is derived by employing a convex and monotonically decreasing kernel profile k which assigns a smaller weight to the locations that are further away from the center of the target. Thus, the robustness of the estimation is increased, since the peripheral pixels are the least reliable ones, since they are often affected by occlusions or the background. So, we can write:

$$\hat{q}_u = C \sum_{i=1}^n k(\|F(x, y)\|) \delta(b(F(x, y)) - u) \quad (\text{A.5})$$

whereby C is computed by imposing the condition $\sum_{u=1}^m \hat{q}_u = 1$, i.e. the summation of delta functions for $u = 1, \dots, m$ is equal to one.

Further, when the target model is passed onto the next frame, we calculate the probability of colour u in the target candidate with a center $F_0 = F(x_0, y_0)$ and a radius h as:

$$\hat{p}_u(F_0) = C_k \sum_{i=1}^{n_k} k \left(\left\| \frac{F_0 - F(x, y)}{h} \right\|^2 \right) \delta(b(F(x, y)) - u) \quad (\text{A.6})$$

The most probable location $F(x_a, y_a)$ of the target pixel area in the current frame is obtained by minimizing the distance $d(F(x_a, y_a))$ at a given location $F_a = F(x_a, y_a)$, which is described by:

$$d(F_a) = \sqrt{1 - \rho[\hat{\mathbf{p}}(F_a), \hat{\mathbf{q}}]} = \sqrt{1 - \sum_{u=1}^m \sqrt{\hat{p}_u(F_a) \hat{q}_u}} \quad (\text{A.7})$$

By defining now that $\hat{\mathbf{p}}(y) = (\hat{p}_1(F_a), \hat{p}_2(F_a), \dots, \hat{p}_m(F_a))$ and that $\hat{\mathbf{q}} = (\hat{q}_1, \hat{q}_2, \dots, \hat{q}_m)$, the distance $d(y)$ that is equivalent to maximizing the Bhattacharyya coefficient [75] $\rho(F_a)$, can be minimized as described in [76]:

$$\rho(I_a) \equiv \rho[\hat{\mathbf{p}}(F_a), \hat{\mathbf{q}}] = \sum_{u=1}^m \sqrt{\hat{p}_u(F_a) \hat{q}_u} \quad (\text{A.8})$$

A.2. Skin Colour

Detecting human skin tone is of utmost importance in numerous applications, such as video surveillance, face and gesture recognition, human computer interaction, human pose modelling, image and video indexing and retrieval, image editing, vehicle drivers' drowsiness detection, controlling users' browsing behaviour (e.g. surfing indecent sites, etc.), semantic filtering of web contents and steganography [77]. Detection of human skin tone is regarded as a two-class classification problem, and has received considerable attention from researchers in recent years [78] [79], especially those who deal with biometrics and computer vision aspects. Numerous techniques for skin color modelling and detection have been proposed in [80].

Generally, methods for skin detection and segmentation can be divided into the three following categories.

1. The first category of methods uses explicit rules on the color values and a metric, which measures the distance/proximity between each pixel's colour and the pre-defined skin tone [195]. Although these methods are very simple to implement and computationally inexpensive, they cannot cope with the complexity of the problem.

2. The second category uses a nonparametric model for skin color distribution. These methods estimate the skin color distribution from the training data without deriving an explicit model of the skin color [148]. Probability based classifiers are also developed to segregate skin tone regions such as the Bayes classifier used in [81]. This category includes methods that build and use the skin distribution map (the probability distribution of observed skin colors). Although they are fast, these methods require significant storage space and a careful selection of the training set.
3. The third category uses parametric models for the skin color distribution. These models usually consist of a Gaussian or a mixture of Gaussians and offer a more compact skin representation along with the ability to generalize and interpolate the training data [82].

An interesting approach of combining the problem of skin detection and model learning for an image has been presented in [83]. In particular, using an initial skin color model, the authors estimate the actual skin color distribution in an image and learn the non-skin distribution.

In general, apart from the methods in the first category, almost all other methods build an extra non-skin model. In this case, the image pixels are detected as skin by comparing their color's probability of being skin or non-skin, using the likelihood ratio. All these methods use a number of images to build their models and thus require significant storage, application specific adjustments and increased computational power in order to detect skin pixels in an image.

Contrary to the model learning based methods, a real-time skin detection algorithm has been implemented in the framework of the current tracker, which utilizes explicit rules on the colour values of the white balanced images. Two critical issues for colour-based skin detection have to be answered [84]: (1) "What colour space should be selected?" and (2) "What segmentation method should be used?"

Driven from the outcomes of the survey on pixel-based skin color detection techniques [80], we utilize a combination of two colour spaces, namely the *RGB* and the *HSV* colour spaces, as a very promising answer to the first question above. The problem that arises by the second question, however, is partially answered by the location of the skin coloured pixels, augmented by

both the existence of movement (see Section A.4) and the relative distance between the detected head and the expected location of the hands.

Theoretically, the only skin coloured pixels in a regular image are the ones of the face and the ones of the hands of all people. Practically, however, this is almost impossible, since there can appear any arbitrary number of objects, within the recorded area, with a skin tone very close to the actual colour of the human skin.

The decision rules followed in the current approach, realize skin cluster boundaries of two color-spaces. Namely a combination of the RGB [85] and the HSV [144] colour spaces is utilized, which forms a very rapid classifier with high recognition rates. Specifically, given that the r , g and b denotes the normalized colour space:

$$r = \frac{R}{R + G + B} \quad g = \frac{G}{R + G + B} \quad b = \frac{B}{R + G + B}; \quad (\text{A.9})$$

the skin color classification of the pixels in each frame F is implemented by setting constraints on the normalized values of both RGB and HSV color spaces, according to [195]. This way, a skin-mask image $S(F)$ is acquired (Figure A.2c):

$$\begin{aligned} R \geq G \text{ and } ||R - G|| > 11 & \quad S \geq 0.12 \text{ and } S \leq 0.7 \\ r \geq 0.33 \text{ and } r \leq 0.6 & \quad V \geq 0.3 \text{ and } V \leq 1.0 \\ g \geq 0.6 \text{ and } r \leq 0.37 & \quad H \geq 340 \text{ and } H \leq 50 \end{aligned} \quad (\text{A.10})$$

where the b component has the least representation of skin colour and therefore it is omitted in skin segmentation [145].

A.3. Background Removal

Another filtering towards the detection of the hands in the original frame F includes 3D filtering and background extraction. Background extraction methods are well known in silhouette extraction problems and usually require significant processing power, since they are based on segmentation or other image processing techniques [146], [147]. Contrary to these, a simple, real-time processing method for background isolation has been implemented

herein, based on the depth information of the utilized camera.

The acquired depth images (Figure A.2b) are gray-scale images, whereas the farthest objects are marked with darker tones while the closest ones with brighter ones. Given that the face detection was successful, the depth value of the head can be acquired. Any object with a depth value greater than the one of the head can now be discarded and thus excluded from our observation area. On the other hand, all objects in the foreground, including the user's hands, remain active. Thus, another mask image $B_{head}(F^n)$ is obtained that corresponds to the active foreground. A major contribution of this filtering will be the exclusion of all skin colored items in the background, including other persons or surfaces in skin tones (wooden floor, shades, etc) and other noise.

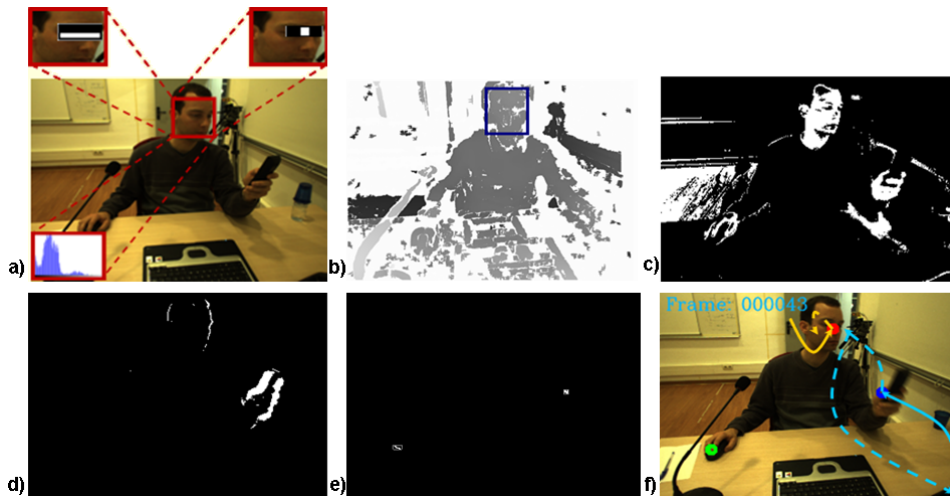


Figure A.2.: (a)Face detection, (b)Disparity Information, (c)Skin Colour Filtering, (d)Motion Detection, (e)Detected Hand Positions and (f)Tracking Result

A.4. Motion Detection

Given the n^{th} frame F^n of the recorded image sequence, a skin-colour mask $S(F^n)$ (Section A.2) combined with background extraction $B_{head}(F^n)$ with respect to the head's position can offer an initial approximation of the possible location of the palms. The head can be efficiently tracked via the head detection algorithm as described in Section A.1. Moreover, provided the pre-calibrated set of colour/depth sensors mounted on the camera, the real

depth information can be easily derived. Thus, it can be written that the derived filtered image $D(F^n)$ can be written as $D(F^n) = S(F^n) \cap B_{head}(F^n)$.

Although the mask image $D(F^n)$ has filtered out most of the unwanted information of the n^{th} frame F , it still contains a lot of foreground noise, which deteriorates the accuracy of the tracking. Given that the background has been removed, we can predict with much confidence that the only moving parts between the camera and the user are the user's hands. Thus, skin-coloured foreground objects, shadows, illumination variances, etc. could be removed, if the moving objects on $D(F^n)$ would be detected.

To this direction, the detection of the exact position of the two palms on each frame is augmented by a motion detection algorithm. In particular, the concept of Motion History Images (MHI) [196] is employed. Each MHI is a static image template, where pixel intensity is a function of the recency of motion in a sequence (recently moving pixels are brighter). Thus, by defining as $M(F^n)$ the pixel-wise subtraction of the two sequential filtered images $D(F^n)$ and $D(F^{n+1})$: $M(I^n) \equiv D(F^n) - D(F^{n+1})$, the remaining blobs on the image F_f^n provide a good estimation of the palms' positions (Figure A.2d).

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

The remaining active regions in each frame MHI_t , where $t = 1, \dots, N$, indicate the locations of the palms of the user with high confidence (see figure A.2e and A.2f), while the location of the head has already been estimated from Section A.1 (Figure A.2a).

By filtering all images from the recorded video sequence successively, the movements of the head and hands can be followed, while the user performs any activity (Figure A.2f). Last, by defining some a priori rules about the environmental setting (e.g. an office, where the users are expected to be seated or a corridor, where the users are meant to stand) the proposed tracking algorithm renders a very accurate system even under extreme conditions (e.g. large moving skin coloured regions, in the case the user wears a T-shirt).

B. Geometric Gait Features Extraction

The estimation of the Geometric Gait Descriptor is described herein, along with the algorithm for the extraction of the gait related soft biometric characteristics, i.e. “height” and “stride”.

B.1. Geometric Gait Recognition

Let the term “*gallery*” refer to the set of reference sequences, whereas the term “*probe*” stands for the test sequences to be verified or identified, in both presented modalities.

Assuming a static background and a moving foreground, the walking subject silhouette can be extracted by using a temporal median filter for the background estimation on the image sequence. Next, the binary silhouettes, B_k^{Sil} , are extracted by comparing each frame of the sequence with the background. By using a pre-defined threshold, the silhouette areas can be separated from the background. In order to denoise the silhouette sequences, morphological filtering, based on antiextensive-connected operators [247], is applied. Finally, potential shadows are removed by analyzing the sequence in the HSV color space [248], thus resulting in the final binary silhouette B_k^{Sil} .

Given the extracted binary gait silhouette images B_k^{Sil} at each gait cycles k , the gray level GEI (Figures B.1a-B.1c) is defined over a gait cycle as:

$$GEI_k = \frac{1}{C_L} \sum_{k=CycleStart}^{CycleEnd} B_k^{Sil} \quad (B.1)$$

where C_L is the length of the gait cycle and k refers to the gait cycles extracted in the current gait image sequence.

The feature extraction process of the gait sequences is based on the Radial Integration Transform (*RIT*) and the Circular Integration Transform (*CIT*), but instead of applying those transforms on the binary silhouette sequences themselves, the Gait Energy Images are utilized, which have been proven on the one hand to achieve remarkable recognition performance and on the other hand to speed up the gait recognition ([249] [250]).

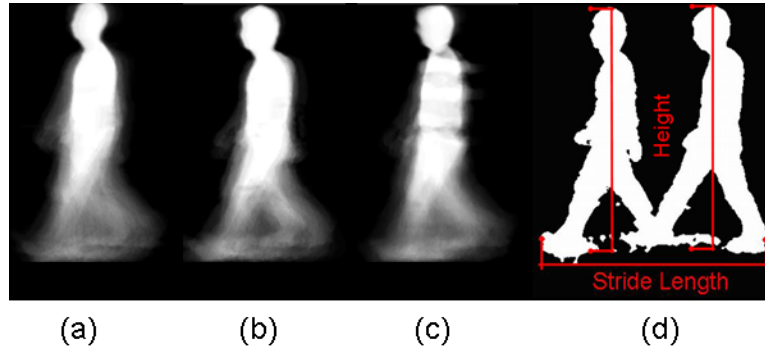


Figure B.1.: a-c) Gait Energy Images from several users; d) Estimation of Height and Stride length from Silhouette Images

The *RIT* and *CIT* transforms are applied on the *GEI* (Figure B.2(a) and Figure B.2(b), respectively), in order to construct the gait template for each user, as shown below:

$$RIT(t\Delta\theta) = \frac{1}{J} \sum_{j=1}^J GEI_k(x_0 + j\Delta u \cdot \cos(t\Delta\theta), y_0 + j\Delta u \cdot \sin(t\Delta\theta)) \quad (\text{B.2})$$

$$\text{for } t = 1, \dots, T \text{ with } T = 360^\circ/\Delta\theta$$

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

Similarly, *CIT* is defined as the integral of a function $f(x, y)$ along a circle $h(\rho)$ with center (x_0, y_0) and radius ρ . The *CIT* is computed using

$$CIT_{f(\rho)} = \oint_{h(\rho)} GEI_k(x_0 + \rho \cos \theta, y_0 + \rho \sin \theta) du \quad (\text{B.3})$$

where du is the arc length over the path of integration and θ is the corre-

sponding angle. Hereby, the silhouette's center of gravity is again used as the origin for the *CIT*.

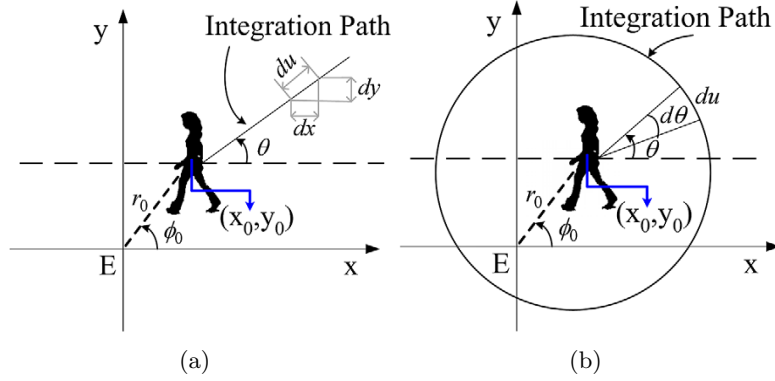


Figure B.2.: Applying (a) the *RIT* and (b) *CIT* transforms on a Gait Energy Image using the Center of Gravity as its origin.

In the same respect, the $KRM_{n,m}$ of order $(n + m)$ transform is applied as follows:

$$KRM_{n,m} = \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} \bar{K}_n(x; p_1, N_x - 1) \cdot \bar{K}_m(y; p_2, N_y - 1) \cdot GEI_k(x, y) \quad (\text{B.4})$$

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

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

$$w(x; p, N) = \binom{N}{x} p^x (1-p)^{N-x} \quad (\text{B.5})$$

$$\rho(x; p, N) = (-1)^n \left(\frac{1-p}{p} \right)^n \frac{n!}{(-N)_n} \quad (\text{B.6})$$

whereby the symbol $(-N)_n$ in is the Pochhammer symbol given by $(-N)_n = -N(-N+1)(-N+2)\dots(-N+n+1) = \frac{\Gamma(-N+n)}{\Gamma(-N)}$, whereby $\Gamma(n) = (n-1)!$ denotes the Gamma Function.

In the proposed framework the weighted 3D *KRM*s are estimated using

the recurrent relations suggested in [340], since their direct estimation is of heavy computational complexity $O(n^6)$.

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

$$d_E = \min_{i,j} (\|\mathbf{s}_i^G - \mathbf{s}_j^P\|) \quad \forall i, j; i \in [1, G_{GEI}] \text{ and } j \in [1, P_{GEI}] \quad (\text{B.7})$$

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

For reasons of convenience, the direct application of the *RIT*, *CIT* or *KRM* transformation of the silhouette images will be referred to as Baseline *BS – RIT* and *BS – CIT* algorithm from now on, while the transformation of the Gait Energy Images as *GEI – RIT*, *GEI – CIT*, *GEI – KRM* algorithm, respectively.

B.2. Height and Stride-length Estimation

A comprehensive analysis on the height and stride length estimation is out of the scope of this thesis. However, in order to make this thesis self contained, a brief outline of the algorithms follow. The height and the stride length soft biometric features are estimated by utilizing the calibrated stereoscopic sequences that were obtained by capturing the HUMABIO Gait Dataset (DB.G.1) and ACTIBIO Gait Dataset (DB.G.2) (see Annex 3). Since real world coordinates and absolute distances can be extracted through calibrated stereoscopic sequences, the problem of the estimation of the height and stride length features is trivially reduced to the selection of the features that correspond to the highest-lowest part of the subject, concerning height, and to the largest distance between the legs within a gait cycle (see Figure B.1d).

The world coordinates $(x_{wc,k}, y_{wc,k}, z_{wc,k})$ of the silhouette image of the k^{th} frame are calculated as

$$(x_{wc,k}, y_{wc,k}, z_{wc,k}) = WC(x, y, D_k(x, y)) B_k^{Sil}(x, y) , \quad (\text{B.8})$$

where D_k stands for the gait disparity data sequence and WC converts a disparity value $D(x, y)$ to 3D coordinates in the world coordinate system, using the already calibrated stereo-camera.

B.3. Detection of stops in a gait silhouette sequence

Initially, the walking human binary silhouette is extracted as described in [178]. Let I_i denote the i^{th} binary human silhouette (2^{nd} row in Figure B.3). In order to detect the stops during a gait sequence, motion estimation through the calculation of a Motion History Image (MHI) [196] is performed in the silhouette image sequence. Motion history template M_t at time instance t is estimated by counting the number of non-zero pixels in the difference image $D(I)$ of two sequential silhouette frames (I_i, I_{i-1}), as indicated by equation (B.9).

$$M_t(x, y) = \begin{cases} b, & \text{if } D(I(x, y))=1 \\ \max(0, M_{t-1}(x, y) - 1), & \text{otherwise} \end{cases} \quad (\text{B.9})$$

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

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

$$f_{motion} = \sum_{i=0}^{N_x} \sum_{j=0}^{N_y} M_t > \epsilon_1 \quad (\text{B.10})$$

whereby N_x and N_y are the resolution dimensions of the image M_t and ϵ_1 the noise threshold of a non-motion image. Similarly, stops in the user's walking are detected, when the motion indicator f_{motion} regarding the lower 25% of the silhouette image height -the part of the legs below the knees [176]- falls below ϵ_2 (3^{rd} row in Figure B.3). The values of both ϵ_1 and ϵ_2 have been experimentally defined.

Once the stop and (re)start frames are detected, the whole gait cycles that include stop frames are removed from the recorded sequence. Thus, a new set of silhouette sequence $\tilde{\mathbf{I}}$ is derived. In the following, the gait

periods are extracted, as described in [178], and the gait cycles indexes are estimated accordingly.

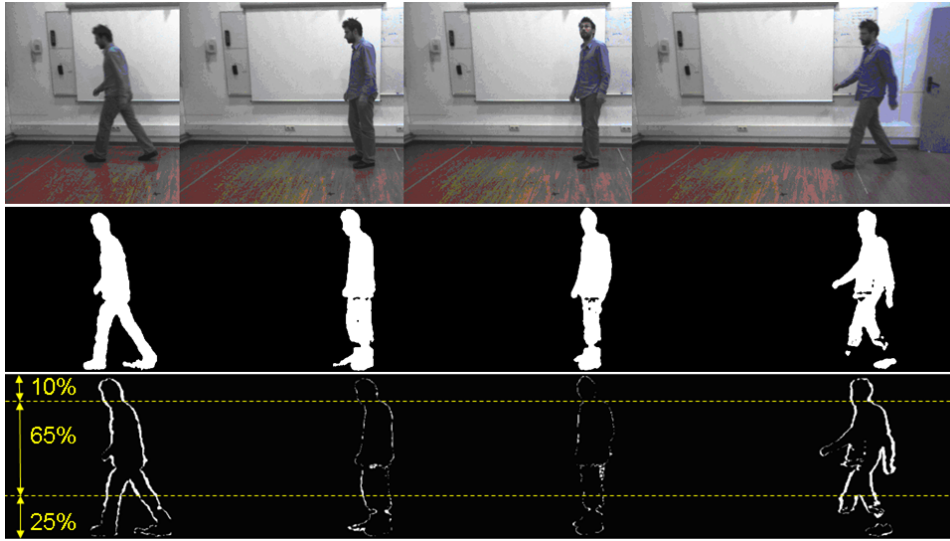


Figure B.3.: 1st row: The user walks along the corridor, makes a short stop (2 frames in the middle) and walks on - 2nd row: Silhouette extraction for the corresponding frames - 3rd row: Motion History Image (MHI) of two sequential silhouette images. The area of interest is restricted to the lower 25% of the image height. The upper region which covers 10% of the image, is considered to include the head, which is used for the estimation of the walking angle.

B.4. Walking angle estimation and compensation

Let the term “*gallery*” refer to the set of reference sequences, whereas the term “*probe*” sequence stands for the test sequences to be verified or identified. As reported in the literature, the gait recognition systems achieve high recognition rates when the gallery and the probe sequences demonstrate similar walking angles [38], with respect to the observing camera local coordinate system. On the contrary, in cases whereby people walk with arbitrary view angles or different model-based types of angle transformations are applied [179]. However, the accuracy of angle view transformations at model based approaches relies on small angle variations that are easily affected by slightly noisy images.

Thus, a novel feature-based method is introduced within the proposed framework that applies, prior to the feature extraction phase, 3D reconstruction on the silhouette itself, encoding at the same time shape information about the user’s body. Specifically, range data are utilized for the compensation of angular variation in the walking direction.

The first step is to estimate the relative walking angle. The walking direction with respect to the camera (Figure B.4) can be estimated in a straight forward manner under the assumption of straight gait within each gait cycle. Given that the head of the silhouette image can be trivially detected within the highest part of the silhouette (Figure B.3), the gait direction \mathbf{v}_1 in the 3D space can be explicitly estimated from the position of the subject’s head in the first \mathbf{h}_0 and last frame \mathbf{h}_L of the respective gait cycle $\mathbf{v}_1 = \mathbf{h}_0 - \mathbf{h}_L$. It should be clarified that the variance of the walking direction within the same cycle is very rare in practice and thus, it is not taken into consideration in the current context.

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

$$\vartheta = \arccos \left(\frac{\mathbf{v}_1 \bullet \mathbf{v}_2}{|\mathbf{v}_1| |\mathbf{v}_2|} \right) \quad (\text{B.11})$$

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

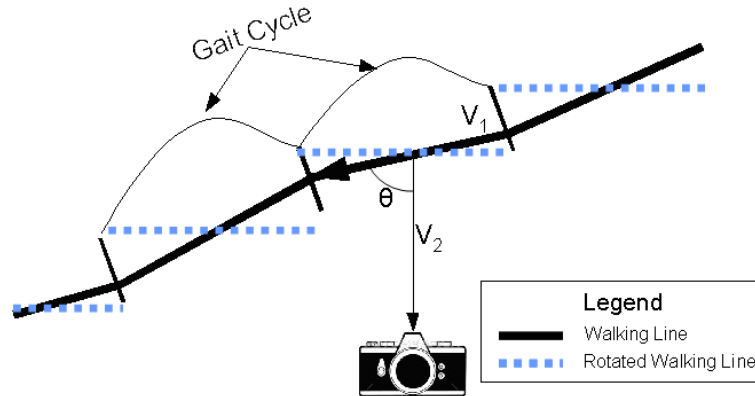


Figure B.4.: The walking angle determination is calculated by the across of the inner product of the walking direction vector and the parallel to image plane vector.

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

$$\tilde{\mathbf{p}}_i = \mathbf{p}_i \cdot \begin{bmatrix} \cos(\vartheta) & \sin(\vartheta) & 0 \\ -\sin(\vartheta) & \cos(\vartheta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (\text{B.12})$$

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

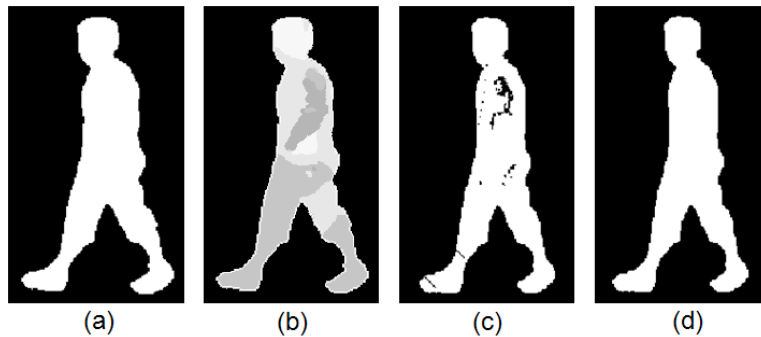


Figure B.5.: (a)Silhouette Image - (b)Depth Image of the Silhouette - (c)Rotated Silhouette - (d)Rotated Silhouette after Refinement.

Despite the notable simplicity of equation (B.12), its direct application in the generation of the virtual view includes some inherent problems, i.e. the reconstructed point clouds could generate non-consistent surfaces, including holes and non-realistic edges, when projected on new virtual views (Figure B.5c). Figures B.5a and B.5b depict the input and the depth silhouette of the user.

Therefore, in the proposed framework a 3D surface is initially formed from the 3D point cloud, so as to generate a consistent surface and silhouette image in the synthesized virtual view (Figure B.5d). The surface is created using only a subset of the points of the image, so as to reduce the redundancy and size of the triangulated surface to be generated. Then, the silhouette

for a particular view is generated by re-projecting it using the Z-buffering principle so as to rapidly perform depth culling in the new rendered image.

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

B.5. Genetic fusion algorithm

The genotypes or chromosomes for the current GA are provided by the concatenation of w_{RIT} , w_{KRM} and w_{HMM} . An initial population of m chromosomes is generated. Each of them denotes the weight for the gait features scores (RIT , KRM) and for the activity-related recognition scores (HMM), respectively. They all range between 0 and 1, similarly to the training patterns, which stand for the dissimilarity scores of the extracted feature. Then, the total similarity $Sim(x, y)$ of each person (gallery) in the database to the client (probe) is given by equation (6.10).

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

$$C = \arg \max_{y \in \mathbb{R}} Sim(x, y) \quad (\text{B.13})$$

Following, the quality of a specific chromosome for the subject C is measured with respect to its *fitness* function $f_{fitness}$, as follows:

$$f_{fitness} = \sum_{x=1}^{N_p} correct_id_x \quad (\text{B.14})$$

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

$$correct_id_i = \begin{cases} 1 & , \text{ if } Sim(x, C) = \max(Sim(x, y)), y = \{1, \dots, N_G\} \\ 0 & , \text{ if } Sim(x, C) < \max(Sim(x, y)), y = \{1, \dots, N_G\} \end{cases} \quad (\text{B.15})$$

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

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