SAPIENZA
UNIVERSITÀ DI ROMA

Biometric Walk Recognizer

Research and results on wearable sensor-based gait recognition

PhD School in Information Science

Dottorato di Ricerca in Information Science – XXXI Ciclo

Candidate

Alessio Mecca
ID number 1392443

Thesis Advisor

Prof. Maria De Marsico

A thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Information Science

October 2018

Thesis defended on 28 February 2019
in front of a Board of Examiners composed by:

Prof. Nome Cognome (chairman)
Prof. Nome Cognome
Dr. Nome Cognome

Biometric Walk RecognizerResearch and results on wearable sensor-based gait recognition
Ph.D. thesis. Sapienza – University of Rome
ISBN: 000000000-0

This thesis has been typeset by LaTeX and the Sapthesis class.

Version: January 30, 2019

Author's email: mecca@di.uniroma1.it

Dedicated to:

who never gives up and
keeps going on walking

and

to my grandmother who supported
and putted on with me throughout
the university course

Abstract

Gait is a biometric trait that can allow continuous user authentication, though being classified as a "soft" one due to a certain lack in permanence, and to sensibility to specific conditions. The earliest research in the domain on gait recognition relies on computer vision-based approaches, especially applied in video surveillance. More recently, the spread of wearable sensors, especially those embedded in mobile devices, which are able to capture the dynamics of the walking pattern through simpler 1D signals, has spurred a different research line. This capture methodology can avoid some problems related to computer vision-based techniques, but suffers from specific limitations. Related research is still in a less advanced phase with respect to other biometric traits. However, the promising results achieved so far, the increasing accuracy of sensors, the ubiquitous presence of mobile devices, and the low cost of related techniques, make this biometrics attractive and suggest to continue the investigations in this field. The first Chapters of this thesis deal with an introduction to biometrics, and more specifically to gait trait. A comprehensive review of technologies, approaches and strategies exploited by gait recognition proposals in the state-of-the-art is also provided. After such introduction, the contributions of this work are presented in details. The research deals with different strategies for gait biometrics, including preprocessing and recognition techniques, in order to allow both an automatic recognition and an improvement of the system accuracy.

Contents

Chapter 1

Introduction

Modern technologies facilitate everyday life, but also create unprecedented security issues. This observation become obvious thinking about modern smartphones and tablets: they are literally miniaturized computers that can access, modify and send documents (possibly private) from anywhere. More in general, a robust user authentication strategy can prevent unauthorized access to restricted physical areas (e.g., a bank caveau), remote services (e.g., home banking), or mobile devices (e.g., smartphones and tablets). The processes to authorize a user conventionally follow three different kinds of strategies: knowledge-based, token-based, and biometrics-based [1].

The **knowledge-based** approach relies on a secret that the user has to know/remember, e.g., the well-known username/password pairs or the PIN codes. Some studies investigated the habits of the users in managing their passwords, highlighting some usability limitations associated with them. For instance, to remember multiple usernames, passwords and PIN codes is not an easy task, especially if they are very complex or long. According to old literature, a heavy information technology user has to remember on average 21 passwords, and some of them even more than 70. From these studies, it comes out that 49% of the users write down or store their passwords in a file, and 67% never change them. More than ten years ago, Gaw and Felten [2] found that even undergraduate students had an average of about 12 accounts, and they found that password reuse was commonplace. Just some time later, Florencio and Herley [3] conducted a large scale study by collecting data on password use from more than 250000 users, during a six-month deployment of a Microsoft toolbar. They found that an average user had 6.5 passwords. However, users accessed 25 accounts over the six-month period, and logged into 8 accounts per day. Therefore, each password was shared across 3.9 websites. Finally, a 2011 diary study of password use by Hayashi and Hong [4] collected detailed records of password entries over a two-week period. They found that users accessed a mean of 8.6 accounts over two weeks, and estimated that most participants had about 11 accounts in total. In practice, all of their participants reported password reusing. In these studies the majority of the participants mentioned memorability as a reason for password reuse. Of course, the use of the same password for all services causes a possible security breach. The same happens for too short, or too obvious and/or easy-to-guess passwords.

The **token-based** authentication approach is characterized by the possession of physical objects, e.g., physical key(s)/card(s). Of course, as material objects, they can be stolen/lost by the user or duplicated by an attacker with enough time aiming at reusing the victim privileges (e.g., to access a secure zone more than one time).

When a higher level of security is required, these two approaches can be combined as, e.g., in the case of the well-known pairs of bank card and related PIN code.

The **biometrics-based** approach exploits either physical or behavioral characteristics of an individual, or a combination of them. This approach differs from the two mentioned above because it is not related to memory or possession but rather to the way an individual behaves. This is something that cannot be forgot, lost, or stolen, even if it is still possible, for an attacker, to try to imitate or forge it. Moreover, biometric traits have a more direct and explicit link with humans than passwords or tokens, since biometrics use physiological and behavioral features of human beings. Even thanks to this fact, nowadays the request for biometrics-based systems is continuously increasing.

Of course, also the biometrics strategies can be combined with the approaches in the other two categories, e.g., in a PIN code/fingerprint authentication strategy.

From the biological point of view, a number of human traits can be used as biometrics, including the popular fingerprints and face, the iris, the hand geometry, and so on. These traits must obey a number of conditions, especially universality (the trait must be owned by all subjects), uniqueness (the trait must present some unique/discriminative characteristics), and permanence (the trait must remain stable for a sufficient time), but also ubiquitousness (the trait can be collected anywhere) is quite desirable. Given the presence/absence of these properties, biometric traits are classified as either strong or soft [5]. Strong traits are generally related to physical/appearance characteristics and allow the recognition of individuals with a sufficient accuracy, though suffering from trait-specific problems. For example, face recognition suffers, at different extents, from modifications in Age, Pose, Illumination and/or Expression (A-PIE). Some soft biometrics are related to high level users' appearance characteristics and can be used to identify groups of individuals (e.g., the hair or the skin colors, and the height). Some others are related to user behavior and consequently lack in permanence, especially in long term analysis (e.g., most behavioral traits can be conditioned by both physical and/or emotional factors differing from time to time). It is also possible to find some traits that can be classified in the middle of these two "soft" biometric subcategories. For example, the way a person walks (gait) is both related to physical characteristics (such as the body conformation) and behavioral factors (e.g., energy saving strategies, see Section 3.1 below) of an individual. It is worth observing that literature has often demonstrated that behavioral traits are generally difficult to spoof (maliciously reproduce). For this reason, they can represent interesting candidates to be used in conjunction with strong traits, that are more accurate but generally less robust to this kind of attack. Even if the gait is included into the soft traits, several studies are still investigating its discriminative power with interesting results. Gait recognition can exploit computer vision techniques (by processing video acquisitions), the analysis of signals from floors equipped with pressure/weight sensors, and/or the data coming from wearable sensors (especially accelerometers).

This thesis deals with gait recognition by wearable sensors, especially focusing on data coming from accelerometers. This choice is mostly due to the nowadays widespread of smartphones (all including an accelerometer sensor) which allows to perform gait recognition practically anywhere. Moreover, mobile biometrics are becoming increasingly relevant in different scenarios during the last years.

The main goals of this research are the following ones:

- to investigate the best preprocessing techniques in order to enhance the quality of gait signals collected by worn sensors;
- to investigate the best way to compare gait signals in order to increase the recognition capability.
- to prototype different real scenario settings to exploit gait biometrics in order to allow a transparent, continuous, and ubiquitous subject recognition.
- to face interoperability problems, e.g., to recognize a subject using different acquisition devices.

The thesis continues as follows: Chapter 2 introduces the main concepts related to biometrics and defines both the different recognition modalities and the metrics used in order to evaluate the performances of a biometrics system. Chapter 3 describes the human gait, discussing the physiological characteristics that allow it to be considered as a biometric trait. It also presents in more details the possible approaches to gait recognition. Chapter 4 deals with a comprehensive presentation of the state of the art on gait recognition based on wearable sensors. In particular, the thesis analyzes the available datasets, some commonly exploited preprocessing techniques and the recognition strategies adopted in literature. Chapter 5 presents in details the original contributions of my PhD studies to the research community. Finally, Chapter 6 sketches some conclusions and possible future works.

Chapter 2

Biometrics

This Chapter aims at providing both general concepts related to biometrics and a common dictionary of the terms used in this research field.

Among the first definitions of biometric recognition, it is worth mentioning the one by Jain *at al.* in [5]: "Biometric recognition, or simply biometrics, is the science of establishing the identity of a person based on physical or behavioral attributes". Such definition has been further refined by ISO (International Organization of Standardization). According to ISO/IEC 2382-37 [6], biometric recognition is the "automated recognition of individuals based on their behavioural and biological characteristics".

As already mentioned in Chapter 1, biometric traits are generally classified, according to their discriminative power and long term permanence, as strong or soft. The former traits present properties (i.e., universality, uniqueness, permanence) that allow accurate recognition, especially in controlled conditions, while the latter lack in one or more of such properties. Some physical traits in soft category, e.g., hair color, height, weight, age, do not identify a single user but an entire group. However, they can reduce the search space for recognition. The majority of behavioral traits fall in the soft category, due to the lack of sufficient permanence. For instance, the way of walking, writing or signing, or the keystroke dynamics, may be affected by either physical or emotional status, especially over the long term. Nevertheless, they can be used in the short term for reidentification, to assure, for example, that the person working with the keyboard does not change during a session. These traits, being not akin to physical characteristics but rather to user's "way to be or behave", are more difficult to forge and spoof. For this reason, these traits can be also used as a further support for strong ones, in a multi-biometrics setting, in order to improve the global recognition accuracy. Among these soft biometrics, the people walking pattern and the related gait recognition, which is the main topic of this thesis, is increasing its popularity, especially due to the possibility to capture gait signals exploiting the built-in smartphones' accelerometer.

2.1 Biometric Applications

In general, biometric recognition entails a (set of) known subject(s), enrolled with one or more biometric templates (gallery), and an incoming template to compare against the reference (probe). According to the ISO/IEC 19795-1 standard [7], there are three kinds of biometric applications. Before start describing the biometric applications, it is worth to introduce the Failure to Acquire (FTA) error, that is used in some performance evaluation formulae.

Definition 2.1.1. Failure to Acquire (FTA) "The failure-to-acquire rate is the proportion of verification or identification attempts for which the system fails to capture or locate a sample of sufficient quality" [7].

2.1.1 Verification (VER)

Definition 2.1.2. Verification. The verification is the "application in which the user makes a positive claim to an identity, features derived from the submitted sample biometric measure are compared to the enrolled template for the claimed identity, and an accept or reject decision regarding the identity claim is returned" [7]. In other words, verification implies a 1:1 identity comparison between the probe and the gallery template(s) of the claimed identity.

Remark. A threshold value is chosen to regulate the acceptance/rejection of the claim.

Remark. "The claimed identity might be in the form of a name, personal identification number (PIN), swipe card, or other unique identifier provided to the system"[7]. In the case of wearable sensor-based gait recognition, this declaration can possibly be implicit, assuming that it corresponds to the identity of the smartphone's owner.

The performance measures for verification applications take into account the possible system errors.

Definition 2.1.3. FMR The FMR is the "proportion of zero-effort impostor attempt samples falsely declared to match the compared non-self template" [7].

Remark. "The measured/observed FMR is distinct from the predicted/expected FMR (the former may be used to estimate the latter)" [7].

Definition 2.1.4. FAR The FAR is the "proportion of verification transactions with wrongful claims of identity that are incorrectly confirmed" [7]. The FRR is computed by using the following formula:

$$FAR = FMR * (1 – FTA) \tag{2.1}$$

Remark. If the FTA is considered as 0 (i.e., no errors happens during the probe acquisition), the FAR is equal to the FMR.

Definition 2.1.5. FNMR The FNMR is the "proportion of genuine attempt samples falsely declared not to match the template of the same characteristic from the same user supplying the sample" [7].

Remark. As for the FMR, the measured/observed FNMR is distinct from the predicted/expected FNMR (the former may be used to estimate the latter).

Definition 2.1.6. FRR The FRR is the "proportion of verification transactions with truthful claims of identity that are incorrectly denied" [7].
The FRR is computed by using the following formula:

$$FRR = FTA + FNMR * (1 - FTA) \tag{2.2}$$

Remark. If the FTA is considered as 0 (i.e., no errors happens during the probe acquisition), the FRR is equal to the FNMR.

According to the standards, two curves are used to summarize verification performance: the one plotting False Acceptance Rate (FAR) at different thresholds, and the one plotting the False Rejection Rate (FRR) at different thresholds.

It is worth underlining that FAR and FRR are computed from time to time according to the same threshold t. This means that an attempt to reduce one of them unavoidably increases the other one (see Figure 2.1).

A summative measure used in order to describe the global accuracy of a verification system is the Equal Error Rate (EER). This value represents the intersection between the FAR and FRR curves.

A further possible evaluation parameter is the ZeroFAR, defined as the value of the FRR when the FAR is fixed at 0 (see Figure 2.1). It is generally used when a very high level of security is required, e.g., in a military application.

Similarly, but in an opposite contest, it is possible to use the ZeroFRR parameter, defined as the value of the FAR when the FRR is fixed at 0 (see Figure 2.1). It is generally used when the security level is less important than the possible delays created by false rejections. An example is the reidentification in a playground in order to allow an already paying client to re-access the site.

In addition, it is possible to exploit the Receiver Operating Characteristic (ROC) curve and its related Area Under Curve (AUC). This curve is obtained by plotting Genuine Acceptance Rate (GAR=1-FRR) vs. FAR, and its AUC is computed as the percentage of the area covered by the ROC. An example of ROC curve is reported in Figure 2.2.

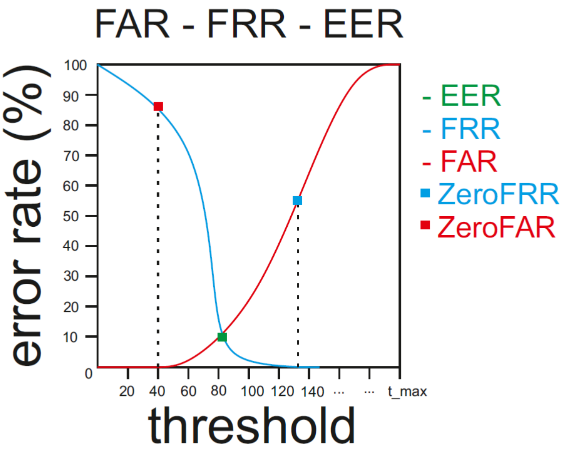


Figure 2.1. An example of FAR and FRR curves with the related EER, ZeroFAR and ZeroFRR points.

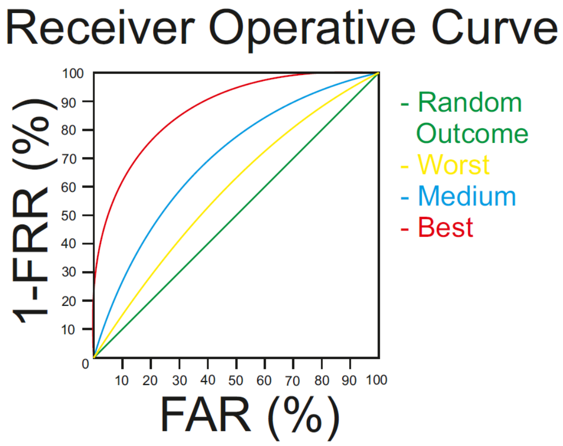


Figure 2.2. An example of ROC curves.

2.1.2 Identification

Definition 2.1.7. Identification. The identification is the "application in which a search of the enrolled database is performed, and a candidate list of 0, 1 or more identifiers is returned" [7].

Remark. The result is a list of the gallery templates, ordered either by increasing distance or decreasing similarity with the probe.

Differently from verification, in this case the subject submitting the probe to recognize does not issue any identity claim. For this reason, a 1:N comparison of the probe with all gallery identities and pertaining templates is entailed. The identification applications can be further distinguished in Closed Set Identification and Open Set Identification.

2.1.2.1 Closed Set Identification (CSI)

Definition 2.1.8. Closed Set Identification. The closed set identification is the "identification for which all potential users are enrolled in the system"[7].

One metric used for performance evaluation is the identification rate:

Definition 2.1.9. Identification Rate "The identification rate at rank r is the probability that a transaction by a user enrolled in the system includes that user's true identifier within the top r matches returned. When a single point identification rank is reported, it should be referenced directly to the database size" [7].

Even if not in the ISO/IEC 19795-1 standard, one of the most used metrics for performance evaluation is the Recognition Rate (RR), computed as the percentage of correct recognitions in the first position of the candidate ordered list returned by the system. A relaxed version of the RR is the Cumulative Match Score (CMS) at rank k. For each rank k, it represents the rate of correct recognitions within the first k positions (note that CMS(1)=RR). Generally used ks are 5 and 10. A comprehensive view of the recognition trend is given by plotting the Cumulative Match Curve (CMC). The points of this curve represent the achieved CMS values for each rank k, i.e., the rate of correct recognition within the first k positions.

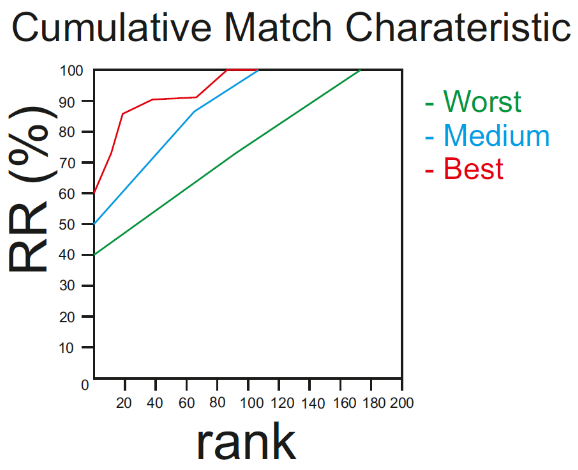


Figure 2.3. An example of the CMC curves.

2.1.2.2 Open Set Identification (OSI)

Definition 2.1.10. Open Set Identification. The open set identification is the "identification for which some potential users are not enrolled in the system"[7].

Remark. The acceptance must be regulated by a threshold, in order to allow a reject option.

In this application there are two possible kinds of errors: to accept a subject not enrolled into the system or to refuse a subject that is enrolled. It is worth noticing that in the latter case, the refusal of a subject can both depend on the fact that the correct identity is not in the returned list, or to the failure in the threshold test. The rate of this errors are used to estimate the performances of this kind of application.

Definition 2.1.11. (True-Positive) Identification Rate (TPIR). The (true-positive) identification rate, or briefly identification rate or TPIR, is the "proportion of identification transactions by users enrolled in the system in which the user's correct identifier is among those returned" [7].

Remark. "This identification rate is dependent on (a) the size of the enrolment database, and (b) a decision threshold for matching scores and/or the number of matching identifiers returned" [7].

According to this definition, in order to be accepted in an Open Set Identification application, a probe has to meet the threshold value and at the same time to be in the first k results in the returned ordered list.

Definition 2.1.12. False-Positive Identification-Error Rate (FPIR). The FPIR is the "proportion of identification transactions by users not enrolled in the system, where an identifier is returned"[7].

Remark. "The false-positive identification-error rate is dependent on (a) the size of the enrolment database, and (b) a decision threshold for matching scores and/or the number of matching identifiers returned" [7].

Assuming N as the number of templates in the database, the FPIR is computed by using the following formula:

$$FPIR = (1 - FTA) * (1 - (1 - FMR)^{N}) \tag{2.3}$$

Remark. Of course, in closed set identification is not possible to have this kind of error since each probe surely belongs to an enrolled subject by assumption.

Definition 2.1.13. False-Negative Identification-Error Rate (FNIR). The FNIR is the "proportion of identification transactions by users enrolled in the system in which the user's correct identifier is not among those returned" [7]. The FNIR is computed by using the following formula:

$$FNIR = FTA + (1 - FTA) * FNMR \tag{2.4}$$

It is worth noticing that, as for verification, the same threshold is used to determine both the FPIR and the FNIR, and therefore, trying to lower one of the two values consequently increases the other. At the same time, also the decided length for the returned list k is the same. Also in this case, the evaluation can rely on EER (defined as for verification, but using the FPIR and FNIR curves), ROC, and AUC, similarly to the verification applications.

Even if not mentioned in the ISO/IEC 19795-1 standard, it is possible to find in literature results presented in terms of the Detection and Identification Rate *DIR(k,t)*. It represents the percentage of identification transactions in which, given the acceptance threshold t, the correct identity of the subject is returned within the first k positions of the returned ordered list of candidates. In a similar way of CSI, it is also possible to compute the *DIR(1,t)*, which considers only the first identity returned in the ordered list. This is analogous to the RR for CSI but, of course, it takes into account the fact that an acceptance threshold is also required. Using this definition, it is possible to compute *FRR=1-DIR(1,t)* and the FAR as for verification.

In general, Open Set Identification is the biometric application raising more errors. Though being definitely more realistic than Closed Set Identification, it has been rarely reported in literature so far, but the U.S. NIST (National Institute of Standards and Technology) is encouraging its use.

Chapter 3

Gait as a Biometric Trait

This Chapter aims at presenting the main characteristics of human gait patterns. Moreover, it shows the motivations for which it is possible to consider the gait as a biometric trait. It also sketches the three possible categories of approaches exploited in gait recognition.

In the following, Section 3.1 presents the general physiological mechanisms underlying the gait, briefly describing how the idea of motion is converted into movement. Moreover, it introduces the concept of cycle and step used in the gait analysis (see Chapter 4) and shows the 8 gait phases that make up a gait cycle. Section 3.2 describes the 3 different research lines to approach gait recognition, aiming at highlighting the general pros and cons.

3.1 General Gait Physiology and Individual Characteristics

A classical authoritative source for understanding bio-physiological rules governing human gait is [8]. The locomotor planning starts in the brain. As for any kind of movement, the "idea" (in this case the desired walking movement) must be converted into the corresponding pattern of muscle activity [9]. The next phase towards the concrete walking entails the transmission of the neural output. The final implementation of the command can be divided into task planning and plan execution. The muscles activation produces tension, that in turn generates the joints movement. As reported in [8], "*the joint forces and moments cause the rigid skeletal links (segments such as the thigh, calf, foot, etc.) to move and to exert forces on the external environment*". The gait action triggers reaction forces from the ground. These are applied through the feet and, when sufficient, support the body and avoid its displacement. These actions represent a basic requirement for walking, together with the familiar periodic movement of each foot between two support positions. Even in the case of underlying pathologies that can distort walking, this pattern can not be scattered significantly [10]. "*This periodic leg movement is the essence of the cyclic nature of human gait*"[8].

In general, a gait period begins from the movement of the first leg and ends when the second leg completely touches the ground. It is characterized by 2 phases and 8 configurations [8], as in Figure 3.1, where arm swinging can also be observed.

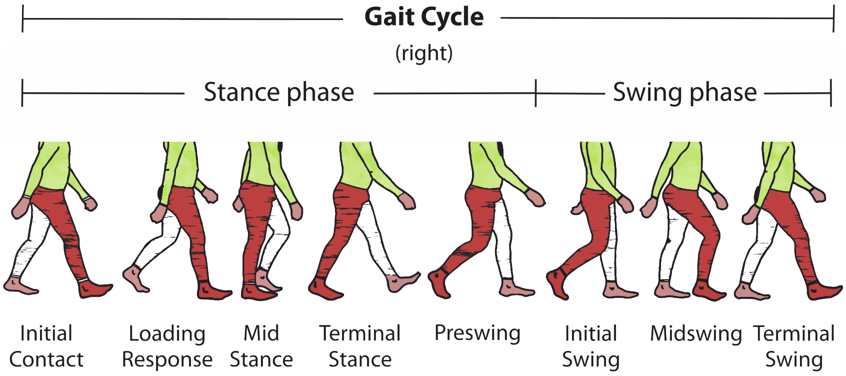


Figure 3.1. Walk cycle dynamics. Inspired by: [8]

The entire gait period is generally called cycle, while the first four configurations and the next four ones compose the right step and the left step (or vice versa). Even if not always true, it is possible to observe that women generally present shorter steps and more pelvic movements. Moreover, intra-subject differences can be caused by ageing and by other physical conditions (such as painful events involving leg(s) and/or foot(feet)).

Given the above aspects, it is worth wondering at which extent this trait can be considered distinctive for different subjects. Biomedical literature studies answer this question. As underlined in [11], the basic walking characteristics are naturally stereotyped, and the basic kinematic patterns are invariant across the normal range of speeds (slow, normal, and fast pace). Moreover, the locomotion must preserve postural stability and dynamic equilibrium [12]. It is also worth noticing that possible changes in support conditions call for an anticipatory adaptation involving coordinated synergies of upper limb, trunk and lower limb movements [13]. Moreover, in normal posture and locomotion conditions, the head is stabilized in space allowing the monitoring of gravity direction [14]. Notwithstanding the common aspects, different personal energy saving strategies produce qualitative and quantitative features, that make individual walking styles unmistakably recognizable. For example, the experiments in [15] by electromyographic (EMG) signals show that "*locomotion cannot be considered as a completely stereotyped function*" since "*despite the similar kinematics, the torque time courses of different subjects present significant differences in agreement with different temporal sequence of muscle activation.* A later work [16] investigates sequences of muscle forces providing for coordinated gait. Overall, posture and locomotion can be considered as a dynamic, personal body signature [17]. The behavioral variations producing this effect can be partially explained by kinematic principles, which are investigated in [18]. Different kinematic strategies are found to be among the factors that produce inter-personal differences in walking patterns, and that can therefore be exploited for identification.

It is worth pointing out that gait is also effected by external factors. For example, literature studies [19, 20] (see also below in Chapter 4) demonstrate that different kinds of worn shoes have an impact on the recognition capability. Moreover, such studies only consider man shoes. At the best of our knowledge, there are no studies concerning woman shoes, especially high heels, that produce totally different walk dynamics that can be also easily noted by humans. Another external factor that influences the gait dynamics is the ground conformation. In fact, to walk on a steep climb or descent requires different energy saving strategies than just walking on a flat ground. Moreover, also the stairs or a dismal ground have an impact: the first because going up and down the stairs is a different task than just walk normally and the second because, in order to avoid physical damages, the brain must actively control the gait dynamics.

3.2 Approaches to gait recognition

Considering the biometrics field, it is possible to classify the literature approaches to gait recognition in three main groups, according the division proposed in [21], depending on the way gait data are acquired and analyzed.

- **Machine Vision-based approaches**: use video sequences and generally aim at extracting some visual elements to model static/dynamic features of walking pattern.
- **Floor Sensor-based approaches**: generally collect data from pressure and/or weight sensors embedded in floors, and generally exploit properties more related to quantitative physical user's features than to user's appearance.
- **Wearable Sensor-based approaches**: acquire data using sensors directly worn by the user, capturing pure walking dynamics; in this case there is no need to equip the environment, allowing ubiquitous recognition.

Similarly to what happens with other biometric traits, both intra-subject differences and inter-subject similarity can condition gait recognition. For example, the gait pattern can be modified by walking speed, kind of worn shoes, ground slopes, and possibly by temporary physical problems. Different kinds of worn dresses and carried objects can be further factors disturbing gait recognition if the source of data is a camera [22]. In this case, these factors are added to the problems that generally affect image processing, e.g., different camera views, varying illuminations, and (partial/self) occlusions. Gait recognition based on sensor-equipped floors has been abandoned during the last years due to the low recognition accuracy achieved and the cost of acquisition devices/ambient set up. However, the biomedical field [23, 24] still exploits gait analysis by equipped-floors for diagnostic research. Wearable sensors appear as a promising alternative to reduce the recognition problems related to appearance.

Notwithstanding limitations, gait recognition offers some advantages too:

- it can operate at variable distances depending on the acquisition modality:
  - Machine Vision-based approaches need a distance up to 10 meters, to maintain a sufficient image quality;
  - Floor Sensor-based approaches ignore distance problems since the devices capturing relevant signals are inside the floor;

  - Wearable Sensor-based approaches ignore distance problems since the acquisition device is located on the user body.

- it is non-intrusive, since the data subject has not to perform any specific action but walk;
- it is non-invasive, since the data subject does not need any physical contact with an acquisition device;
- gait it is quite difficult to imitate (see Section 4.7);
- though being a soft biometrics, recent works report interesting performances [25], and it can be combined with other "strong" biometrics to improve recognition and/or as a presentation attack detection technique.

3.2.1 Machine Vision-based Approaches

The gait recognition approaches based on Machine Vision entail the acquisition of walking signals using one or more video-cameras from distance. Therefore, they require an ambient set up and, in order to monitor distant zones, this set up must be replicated multiple times in order to cover the desired areas.

Approaches in this category can be divided into two main groups: model-free and model-based. The model-free strategies (also often referred as "silhouette-based") usually train classifiers to analyze the motion of the silhouette across the scene. Model-based techniques, instead, rely on a precise model of the human movement, built by exploiting the limbs and joints composing human body. Generally, features extracted from the such model are compared against those stored for the user model.

Two common steps of machine vision-based strategies are the detection of a person in the scene, and the tracking of that person's walk. Of course, occlusions represent a delicate problem in machine vision-based approaches. In most cases, a preprocessing phase includes background subtraction and body silhouette extraction, eventually identifying the Degree of Freedom (DOF) points [26] (corresponding to body joints) in order to track user's gait. What differs from one system to another are the possible further preprocessing operations used to improve the quality of extracted data, and/or the kind of comparing strategies used in order to find the correct identity. The majority of machine vision-based works in the state-of-the-art convert the preprocessed data into the related Gait Energy Image (GEI) or its variations, and use these images as the base for feature extraction and/or comparing.

Among the other possible differences in the state-of-the-art proposals, it is to mention the use of different technologies for data acquisition, such as different kinds of cameras (fixed or Pan-Tilt-Zoom), that can possibly work in different conditions (e.g., visible light, infrared or thermal). Moreover, the quality of the acquisition device is another not negligible factor.

In addition, there are systems that exploit the fusion of the data acquired by more (possibly different) cameras, in any combination. In these cases, is it necessary to synchronize the acquired signals, and this generally requires a stereo calibration procedure, adding computational demand. Of course, the use of more than one kind of camera helps the recognition, e.g., for the previously mentioned problem of the carried objects.

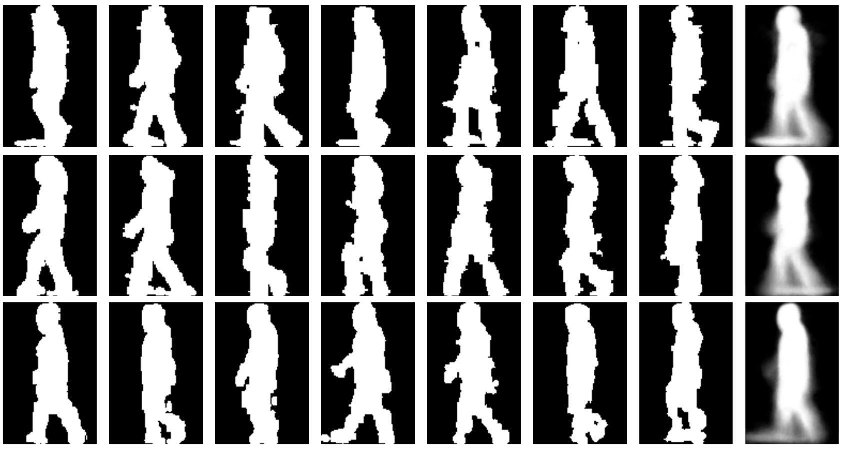


Figure 3.2. Three examples Gait Energy Image (the last image of each row) extracted by the corresponding gait sequences. Source: [27]

In general, the state-of-the-art approaches working in visible light may eventually suffer from pose, illumination, and occlusion problems, especially in outdoor environments. Moreover, another not negligible aspect is the perspective with respect to the camera. In this case, when the user is not consistently aligned with the camera, this creates anomalies and distortions, for example, in the extracted GEI. For these reasons, if only visible light cameras are involved, the comparing generally provides good results only in controlled scenarios.

Some possible solutions to the above mentioned gait analysis problems are proposed in literature. For instance, the problem of pose can be resolved by combining data from different cameras or choosing as video source only the frames in which the highest number of DOF points can be extracted from the images.

The problem of illumination can be reduced by introducing infrared cameras, that allow a more accurate silhouette extraction, especially in dark scenarios, even if they have problems with strong illumination sources if not combined with a visible light camera.

The possible occlusion of elements in the body silhouette represents a delicate aspect, because the performance of recognition algorithms significantly decreases if the subject holds an object or carries a backpack, due to an erroneous silhouette extraction. As shown in the already mentioned [22], a thermal camera can be a suitable solution to solve this problem in the majority of cases. In fact, thermal cameras can help in the identification of the subject body that possibly is warmer than the carried object(s), but of course, their use requires the combination and synchronization of more cameras.

Finally, the problem of perspective can be attenuate by geometric transformations, but this would increase the computational effort and it is not always possible to project data in a reliable way, to reconstruct an aligned view of the scene.

Complete surveys on machine vision-based gait recognition can be found in [28, 29, 30]. Such works provide comprehensive discussions about techniques for machine vision-based approaches. The review in [31], instead, provides a description of model-free machine vision approaches only.

3.2.2 Floor Sensor-based Approach

The Floor Sensor-based approaches rely on the use of a special equipped floor able to record pressure and/or weight variations. This allows a data acquisition that is not afflicted by the well-known and above mentioned machine vision problems. Moreover, the preproccessing algorithms, working generally on linear signals, have a very little impact in terms of computational effort. On the other hand, as machine vision-based systems, also in this case there is a lack of ubiquitousness, because the monitoring of multiple zones requires equipment set up and duplication. Besides this factor, the performances are generally lower with respect to the machine vision-based strategies. There are very few works about this kind of approach for user recognition, and the research in this field has probably been overwhelmed by the new and more practical wearable sensors.

Figure 3.3 shows an example of a floor equipped with sensors.

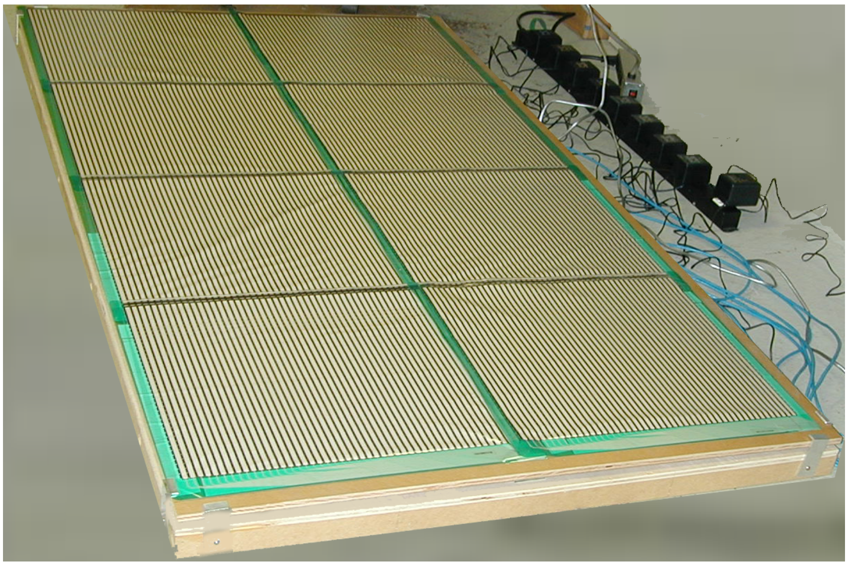


Figure 3.3. An example of a pressure sensor-equipped floor. Modified from: https://ame2.asu.edu/projects/floor

Three proposals concerning floor sensor-based recognition are presented in [32, 33, 34]. The latter especially points out how gait analysis [35] can be also exploited for diagnostic/clinical purposes (see [36]). In fact, as already mentioned, even if this kind of approach is nowadays rarely used for the recognition of individuals, the interest is still alive in the biomedical field. In such field, equipped floors are used for gait pattern analysis in the diagnosis of particular pathologies and as a rehabilitation support. Two examples can be found in [23] and in [24]. In the first, data from equipped floors are used for the diagnosis of Cerebral Palsy and for the evaluation of the outcomes from treatments, while in the second they are used in the study of Parkinson's disease.

3.2.3 Wearable Sensor-based Approach

"*The growing popularity of wearable devices is leading to new ways to interact with the environment, with other smart devices, and with other people. Wearables equipped with an array of sensors are able to capture the owner's physiological and behavioral traits, thus are well suited for biometric authentication to control other devices or access digital services.*" This quotation from the survey presented in [37], and a number of recent works [38, 39, 40, 41, 42] point out a new trend in biometric research, focusing on cheap, widely available, and easy-to-set-up wearable equipment(s). Such equipment(s), and in particular smartphones where they are possibly embedded, can further spur "*a non-intrusive autonomous sensing and context recognition*". This allows the development of a new class of applications defined as "*opportunistic user context recognition with mobile phones*". Such applications have been recently surveyed in [43]. As for some other biometrics (e.g., writing dynamics, signature, iris, face, fingerprints and so on), gait signals can be acquired by mobile devices by standard built-in sensors, especially accelerometer and gyroscope.

Being the main topic of this thesis, the wearable sensor-based approaches are described in more details in Chapter 4. For this reason, this Section only sketches the main lines of research in this field.

In literature, gait recognition by wearable has been faced following two different families of approaches. The strategies in the first group deal with the characteristics (such as shape, period, phase, and so on) of the gait signals. These approaches exploit signal processing techniques, trying to find the best way to compare the entire gait signal or pieces of it (generally corresponding to steps or cycles - see also Section 3.1). One of the most used techniques in this category is the well-known Dynamic Time Warping (DTW), also widely exploited in speech recognition [44], and its variations. The strategies in the second group deal with machine learning techniques, trying to extract the most relevant aggregative characteristics from the gait signals. The new trend in this category is the use of deep architectures, such as Deep Convolutional Neural Networks (Deep CNN). Of course, this kind of approach generally requires an elevate (and sometimes huge) amount of training data to properly work and generalize. Unfortunately, freely available datasets are very few and present some limitations (see Section 4.2).

Chapter 4

Related Work

This PhD thesis deals with gait recognition by wearable sensors. For this reason, only works in this category have been reported in the following. One of the contribution of this PhD study is a survey of wearable sensor-based gait recognition. This work has been submitted to ACM Computing Survey and is currently under review. This Chapter presents an extract of such survey, with an extensive discussion on the literature regarding acquisition sensors that can be used for gait recognition 3.2.3, freely available datasets 4.2, state-of-the-art preprocessing techniques 4.3, and recognition approaches 4.4. The topics in the last three Sections are less addressed in literature. Section 4.5 reviews some works facing gait by wearables; Section 4.6 discusses some particular approaches/aspects of the systems exploiting different kinds of wearable sensors; Section 4.7 presents some works demonstrating the robustness of gait with respect to impersonation attacks.

Interested readers can find reviews on wearable sensor-based gait recognition in [21, 45]. Moreover, gait often appears in researches dealing with human action recognition; extensive reviews are presented in [46, 47, 48, 49].

4.1 Wearable Sensors for Gait Recognition

Section 4.1.1 presents the main characteristics of the smartphones' built-in sensors (focusing especially on the accelerometer). Section 4.1.2 briefly describes other sensors either used in literature works or presenting useful characteristics for gait recognition.

4.1.1 Standard Sensors Embedded in Smartphones

Among the smartphones' built-in sensors, the accelerometer is the most used for gait recognition, so it will be further discussed in more details in the following Section 4.1.1.1. As for now, it is sufficient to say that it records acceleration values along three orthogonal axes.

The gyroscope is sometimes used in gait recognition too. This sensor is made up by a spinning wheel or disc, rotating around its axes. When the disk is rotating, the orientation of the axes tends to be always parallel to itself and to oppose any attempt to change such orientation, according to the law of conservation of angular momentum. For this reason, gyroscopes are generally useful for measuring or maintaining orientation. In gait recognition, when used, the gyroscope is mostly considered as an additional source of information to support recognition by accelerometer. However, differently from the accelerometer, that is a standard equipment for all smart devices (e.g., smartphones, smartwatches, and tablets), the gyroscope is sometimes missing.

For sake of completeness, it is worth mentioning the magnetometer too, because it is often another standard equipment of smart devices. For instance, this sensor is the one that allows geolocalization. It is used to measure magnetization, and the strength and possibly the direction of the magnetic field in a certain point. For this reason, it acts as a compass in consumer devices. At the best of our knowledge, this sensor is barely used in gait recognition, because it merely contributes to detect walking direction. Moreover, it can be negatively affected by external magnetic fields beyond the earth one.

A general consideration about these kind of sensors embedded in the smartphones is that, as for now, the Android standard does not allow to acquired data at a fixed sampling rate. For this reason, if the exploited techniques require samples with a constant frequency, it is necessary to interpolate the captured signal (see Section 4.3).

4.1.1.1 The Accelerometer Sensor

The accelerometer is a sensor able to record acceleration variations in time, reporting them in terms of $\frac{m}{s^2}$ or g. It measures the proper acceleration, i.e., the one relative to free-fall, also known as g-force. This is the acceleration felt by people and objects. Even if it is possible to find accelerometers with only one or two axes, the most common models have three of them. Nowadays, the widespread use of smartphones has significantly incremented their diffusion. In fact, as already mentioned before, the smartphones always have a built-in tri-axial accelerometer sensor, and the majority of them have a gyroscope and a magnetometer too. There are different kinds of accelerometers. Even if only those embedded in smart devices will be discussed here, the general underlying principle is always the same: a mass is taken hang up by some force, e.g., the one produced by direct attachment to an elastic element, such as a spring, and when an external force moves the sensor (and consequently the mass), the device measures the movement. Taking into account the direct proportionality among the movement and the acceleration, it is possible to coherently convert the variation in position into an electric signal. As a consequence, this signal will contain the converted acceleration variations during time. It is worth noticing that this sensor can reveal a different acceleration on each axis, so it is possible to access three different measurements at any time. Figure 4.1 shows a simplified schema of the accelerometer functioning: it is possible to see a spherical mass hung up by three springs, representing the three axes, which pass through it. Moving the cube, the mass will change its position, compressing and extending the spring lengths. These compressions and extensions allow to reveal the physical acceleration on each axis and its direction.

Though the majority of accelerometers use this kind of schema, the modern ones have a micro-manufactured silicon structure, as highlighted in [50]. This is done in order to reduce dimensions. In this case, the mass is not spherical and it

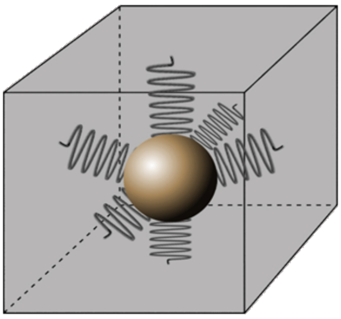


Figure 4.1. A simple schema of accelerometer functioning. Self-produced and presented in [118].

is substituted by a mobile plate in a capacitor, hung up between two other plates that are fixed in the structure in a way that avoids any contact between them. The sensor measures the mass movement exploiting the electric capacity variation in the capacitor, which directly depends on the distance between the plates.

Finally, the modern wearable devices generally use accelerometer sensors made up by a single silicon chip with an integrated electronic circuit. These chips are included in the MEMS (Micro-Electro-Mechanical-Systems) category and are microscopic, as big as a match tip. A complete description of all MEMS characteristics can be found in [51]. In addition to their microscopic sizes, they generally have a high sensibility (see below for a definition of this characteristic), are little influenced by temperature variations, provide a good accuracy, are able to reveal relatively small acceleration variations and, lasts but not least, they have a very low power consumption and are very cheap. For these reasons, they are perfect to be integrated in everyday usable devices such as smartphones and tablets.

Figure 4.2 shows the orientation of the accelerometer axes in a smartphone.

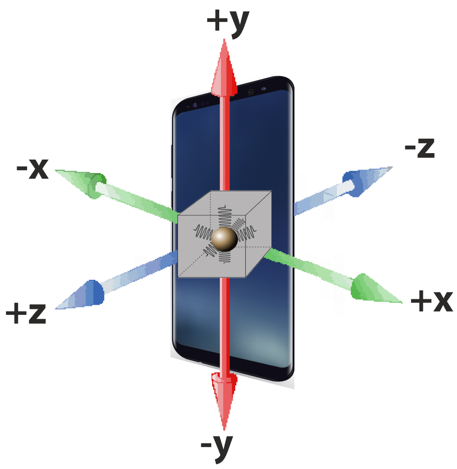


Figure 4.2. Accelerometer axes and their orientations in smartphones. Self-produced image.

In general, data provided by an accelerometer at the time instant t_i is a triplet of acceleration values recorded for the three axes, x_i, y_i, and z_i. As a consequence, a walking time series is represented by a series of these triplets. These triplets make up three time series, one for each axis.

It is worth pointing out two important factors. The former is a negative aspect: as it happens with all physical sensors, even two accelerometers of the same brand and production chain can present different values in identical positions and in identical conditions (e.g., walking speed and pace). This possible misalignment with respect to the expected values is negligible for the usual tasks this sensor is used for in smart devices (e.g., rotation of the screen, gaming applications, and so on), but it is relevant for biometrics applications. The second factor is a good property: the linearity. This means that the accelerometer data are directly proportional to the physical acceleration it is intended to measure. This is a very useful characteristic that can be used in various ways. An example is represented by the data normalization procedure developed during my PhD and presented in [52] (see Section 5) that aims at solving the problem of inter-sensor differences.

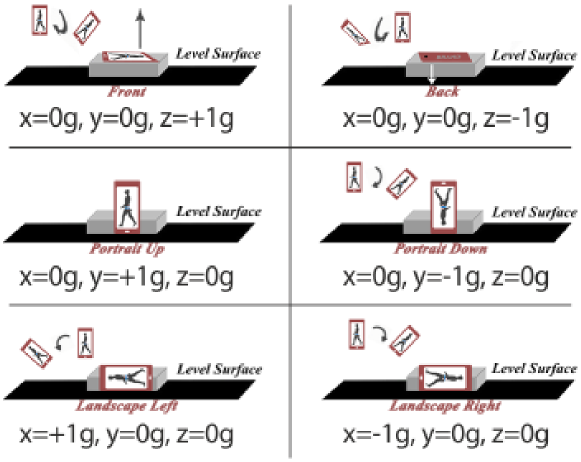
In general, when working with an accelerometer sensor, it is worth taking into account some important parameters that define its physical characteristics and help to better exploit its functionality. In the following, the most relevant ones are introduced.

The *maximum range* parameter describes the range of acceleration values that can be measured by the sensor: if the collected values are outside this range, the accelerometer will lose its linearity property. This parameter is normally expressed in terms of g (gravitational force or g-force, i.e., $9.81m/s^2$). Common built-in accelerometers have a range that varies from $\pm 2g$ to $\pm 8g$. For example, the recent Samsung Galaxy S9 contains a LSM6DSL iNEMO inertial module (the same of their previous S8 and S8+ model) that can be set up to acquire acceleration signals ranging from -8g to +8g. The same inertial module is also embedded into the Huawei Mate 10. The Iphone X includes Bosch BMI160 IMU, instead. This sensor can acquire data ranging from -16g to +16g.

The *bandwidth* expresses the maximum frequency of detectable variations and it is better known as *sampling rate* or *Output Data Rate (ODR)*. This value is measured in $Hz(1/s)$ and, in accelerometers built in mobile devices, it is generally about $100Hz$, while it is possible to find high quality accelerometers with a sampling rate of more than $500Hz$. For example, the already mentioned LSM6DSL iNEMO inertial module has an ODR of $400Hz$, while the Bosch BMI160 IMU can reach an ODR up to $1600Hz$.

The *sensitivity*, sometimes denoted as *resolution*, describes the minimum detectable acceleration variation. This value is generally expressed in terms of LSB $(LeastSignificantBit)/g$. This means that if an accelerometer has a *sensibility* of x, it can provide only measurements that are multiples of x.

The *Offset* (often referred as *Zero-g Offset* or *Zero-g Bias*) value describes the difference between the real output and the ideal output when no acceleration is applied to the sensor. Considering sensors built in smartophones, the X axis is the one co-planar with the screen, parallel to the short side and with positive direction rightwards; Y axis is the one co-planar with the screen, parallel to the long side and with positive direction upwards; and Z axis is orthogonal to the screen with positive direction frontwards (see Figure 4.2 and Table 4.1). In an ideal scenario, when an accelerometer sensor is placed on a horizontal flat surface with the front



Level Surface Front x=0g, y=0g, z=+1g	Level Surface Back x=0g, y=0g, z=-1g
Level Surface Portrait Up x=0g, y=+1g, z=0g	Level Surface Portrait Down x=0g, y=-1g, z=0g
Level Surface Landscape Left x=+1g, y=0g, z=0g	Level Surface Landscape Right x=-1g, y=0g, z=0g

Table 4.1. Ideal accelerometer values with respect to the smartphone positions.

surface facing up, the accelerometer values should be $0g$ on X and Y axes and $1g$ for Z axis. Rotating the sensor by 180°, the values for X and Y axes would remain unchanged while the one for Z would change in $-1g$. Table 4.1 reports the ideal values in all of the six "flat" positions when the sensor is embedded in a mobile device, a smartphone in this case.

4.1.2 Other Kinds of Wearable Sensors used in Gait Recognition

Further sensors are used in gait analysis, though they have not, or not yet, been exploited in gait recognition too. They can be part of wearable devices or can be implanted inside a shoe toe or sole.

Force sensors return a current or voltage measure which is proportional to the pressure, i.e., to the Ground Reaction Force (GRF) under the foot.

Pressure sensors (mostly capacitive, piezoresistive , or resistive piezoelectric) measure the force applied on the sensor neglecting its spatial components. Different types offer, e.g., a different range of pressure they are able to stand and measure, or a different sensitivity. These sensors have been recently used for gait recognition too (see Section 4.6.3). Pressure sensors can also be exploited in the footstep recognition. An example can be found in [53].

Goniometers measure the angles, for example, of ankles or knees. Strain gauge-based ones work with the resistance, modified by sensor flexion (the material stretches, and the current traversing it must complete a longer path, causing a proportional resistance increases). These sensors can be fitted into instrumented shoes but have never been exploited for gait recognition.

Ultrasonic sensors are used to analyze short steps and stride length and the feet distance.

Electromyograph (EMG) measures the electrical manifestation of the either voluntary or involuntary muscle contraction. The signals can be obtained either from surface electrodes (non-invasive), or from wire/needle electrodes (invasive). They can measure different gait features, e.g., kinematic plots of joint angular motion. A recent work uses this technique (see Section 4.6.3).

4.2 Gait Datasets Acquired by Wearable Sensors

There are few public benchmark datasets for wearable gait recognition. Works published so far in this field are generally tested on different datasets which possibly differ for number of subjects, length of the templates, sampling rate and other important characteristics, so that it is not so easy to sketch a meaningful comparison across them. This Section describes the only three (to the best of our knowledge) public freely accessible datasets relevant for this field. For the works exploiting in-house collections of walking samples, a description of the characteristics of such collections is provided during the presentation of the recognition strategy and their characteristics are summarized at the end of Section 4.4.

OU-ISIR dataset[1] [54] collects data from a rather large number of subjects (744) with a 2-78 years of age range. No further demographic information is provided. The data is captured with both accelerometer and gyroscope, though collected from a similar body location (all acquisition sensors are located in a belt in the hip and back waist zone). A single signal per subject is collected during a single walk (therefore recording very few intra-class variations) along a path with two 3 meters long slopes (one ascending and one descending), and a 9 meters level floor (4.5 meters in each direction). The diversity of ground inclination is a positive feature. The devices to record acceleration are 3 IMUZs, each equipped with accelerometer and gyroscope; in addition, the authors use a Motorola ME860 smartphone with a triaxial KXTF9 Kionix accelerometer. There are 7 signals for each walk (3 from the accelerometer and 3 from the gyroscope of IMUZ, and one accelerometer signal from the smartphone). The total number of acceleration samples per walk is about 1400 and no data preprocessing is applied. Each collected walk is manually segmented to extract 4 fragments, 2 from the level floor path and 2 from the ground slopes. They appear probably too short to provide a reliable benchmark. In fact, there are only about 400 samples for the level floor fragments and about 250 for the slopes ones, and this can negatively affect recognition.

ZJU-gaitacc dataset[2] [55] includes gait signals from 175 subjects. Though the number of subjects is lower, walks are much longer and variated. Data for 153 subjects is captured in two different sessions (6 walks per session), with a delay ranging from a week to six months. The remaining 22 subjects have 6 walks collected in a single session. No individual demographic information on either gender or age is provided. Walk signals are collected along a 20 meters long hallway. This corresponds to signals with a sufficiently high number of samples per walk (about 1400). The acceleration data are collected using 5 Wii Remote controllers, located on the right ankle, the right wrist, the right hip, the left thigh, and the left upper arm. The positive features of this dataset are the high number of walks per subject and per sensor, and the sufficient length of the signals for a significant comparing. Since the two different sessions are sufficiently separated in time, it is also possible to consider more unpredictable time-related variations. However, the signal quality is not optimal, due to the low accuracy of the controller. In addition, the data are interpolated and, at the best of our knowledge, it is not possible to access to the raw

[1] http://www.am.sanken.osaka-u.ac.jp/BiometricDB/InertialGait.html

[2] http://www.cs.zju.edu.cn/~gpan/database/gaitacc.html

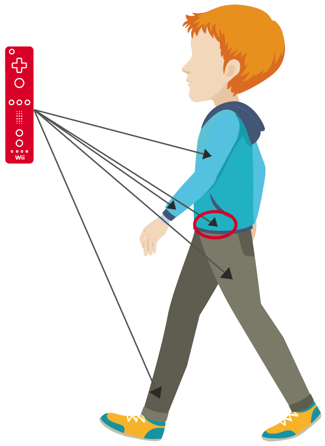


Figure 4.3. Body locations available from ZJU-gaitacc dataset. The red circle (pelvis zone), is the one exploited in the proposals in Section 5. Self-produced image.

signals. This is not ideal to test approaches that do not entail interpolation. Figure 4.3 shows the body locations available from the dataset. The red circle indicates the one exploited to test our proposals, presented in Section 5. The reason for this choice is twofold: first of all, the use of all 5 accelerometers is not suitable for a real system, and in second place the pelvis location is the one that achieves better results (since such bodily zone is the closest one to the gravitational center). Moreover, this location is close to the trousers pocket location, which is a reasonable place to put a smartphone in daily life/real scenario situations.

Concluding, OU-ISIR can assess inter-class variations, thanks to the high number of subjects, while intra-class ones are better captured by ZJU-gaitacc. However, for both of the datasets, the lack of demographic data does not allow further analysis, for example, to study gait-based gender recognition.

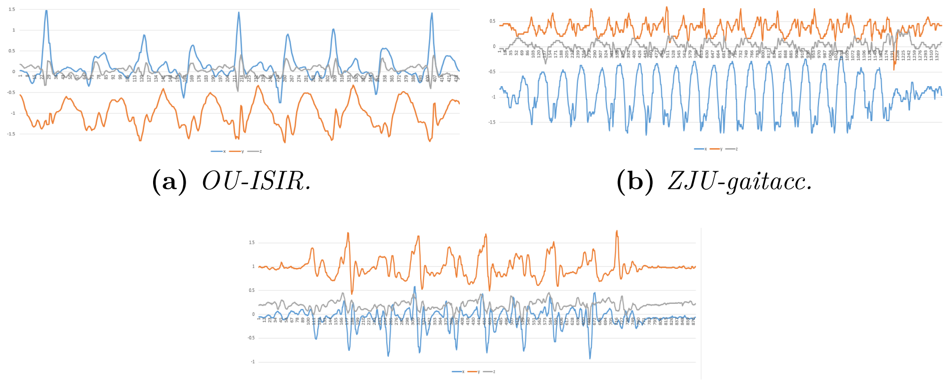
BWR-MultiDevice dataset[3] exclusively collects data for smartphones-based experiments. This dataset has been collected during my PhD [52] and a description of its use is presented in 5 in more details. This dataset is smaller than the previous ones. It aims at highlighting the differences among signals collected by different devices. It contains walking signals from 25 subjects during 2 different sessions with at least 15 days of time elapse. The data acquisition is carried out along a hallway with 3 different accelerometers embedded in different smartphones: a Bosch Sensortec BMA250 (Sony Xperia S), a ST Microelectronics K330 (Samsung Galaxy S4 Active), and a ST Microelectronics LIS3DH (OnePlus One)[4]. The acquisition device is positioned on a belt in a lateral position. Each walk is about 10 steps long with an average number of 1300 samples. Each session contains 2 walks per subject per device. Only a single device at time is used to acquire data and after each acquisition the device is detached and repositioned to add further variations.

[3] https://sites.google.com/a/di.uniroma1.it/biometric-interaction/home/gait-recognition/datasets/bwr-multidevice

[4] We decided to do not include Iphone models for two reasons. The former is a time constraint. The acquisition of accelerometer data using an Iphone would require a possibly completely new application. The latter is a technical constraint. As for the Iphone standards, the management of applications working in background is very strict and barely allowed.

This means that all walks collected separately and no data belongs to the same acquisition momentum. In summary, each session contains 6 walks per subject, for a total of 12 walks (300 walk signals in total). Each walk is annotated with some demographic information (gender and age range, ethnicity is always Caucasian). At the moment, this dataset is the only one that provides cross-sensor signals from different smartphones, and this allows investigating cross-identification approaches. It is worth considering the different capture accuracy of the embedded accelerometers, and the need to somehow normalize signals before comparing them.

For a summative representation, Table 4.2 describes the main features of these three datasets. Figure 4.4 visually exemplifies the signals differences. Figure 4.5 shows 2 templates for the Y axis only (the most relevant one) from ZJU-gaitacc, OU-ISIR, and the in-house dataset collected in my PhD research.



(a) *OU-ISIR.* **(b)** *ZJU-gaitacc.*

(c) *BWRMultiDevice.*

Figure 4.4. Examples of acceleration signals from OU-ISIR, ZJU-gaitacc, and BWR-MultiDevice datasets.

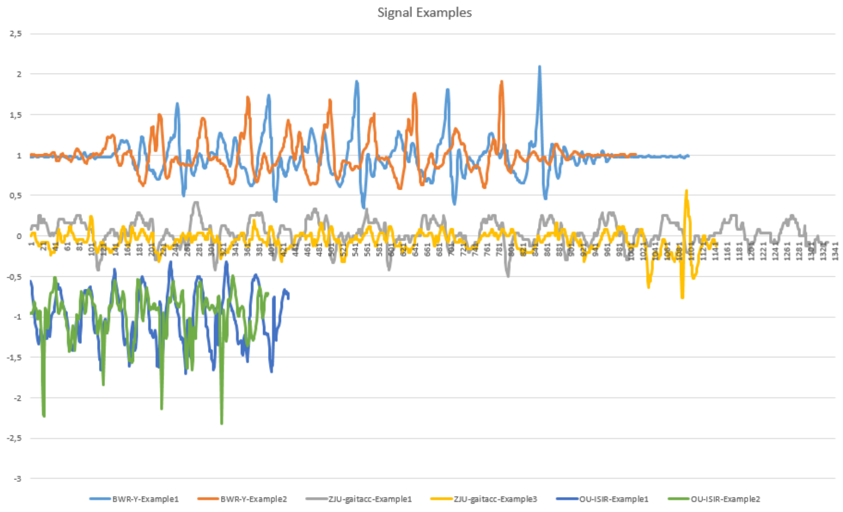


Figure 4.5. Examples of y axis signal from 2 template taken by ZJU-gaitacc, OU-ISIR, and the in-house dataset collected during my PhD research.

Table 4.2. Summary of freely available datasets that were collected by wearable sensors.

Data			Dataset						
Paper	Year	Dataset	Device(s)	Device(s) Position(s)	# Subjects	Walk Length	# Walks per Subject	# Sessions	Notes
[54]	2014	OU-ISIR Inertial Sensor	3 IMUZ, 1 Motorola ME860 (both with accelerometer and gyroscope)	Hip, Back Waist	744	15m (2 slopes of 3m, 1 level of 9m divided into 2 of 4.5m)	1 (manually segmented into 4 fragments)	1	Age ranges from 2 to 84 years Short signals
[55]	2015	ZJU-gaitacc	5 Wii Remote	Left Upper Arm, Right Wrist, Right Hip, Left Thigh, Right Ankle	153 (2 sessions) + 22 (1 session)	20m	12	2 (with 6 walks per user each)	Data are interpolated
[52]	2016	BWR-MultiDevice	1 Samsung S4 Active, 1 OnePlus One, 1 Sony Xperia S	Hip	25	10 steps	12 (4 for each smartphone)	2 (2 walks per smartphone per user)	3 different acquisition devices

4.3 Preprocessing Techniques

This Section deals with preprocessing techniques often used in state-of-the-art proposals. The majority of them are general purpose signal enhancing strategies used in different signal processing fields.

Denoising aims at enhancing the quality of the signal produced by a device. In the case of wearable sensors, the noise can be static (constant and device-dependent) and/or dynamic (because of the slight fluctuations produced by the gait). The two most exploited algorithms are the Weighted Moving Average (WMA) [56], (used in [57, 19, 20]), and the Wavelet Denoising (WD) [58] (used in [59, 60]). WMA substitutes signal points with the weighted average value in a fixed neighborhood, with closest points having highest weights. It requires choosing whether to apply the algorithm symmetrically or only forward, and the number of values making up the sliding window (entering the computation). As concerns WD, it entails convolution of the chosen wavelet with the original signal. The key factors are the choice of the mother wavelet and its amplitude.

Figure 4.6 shows an example of the effect of denoising by wavelet and by moving weighted average. The wavelet used in the figure is the same of [60, 59].

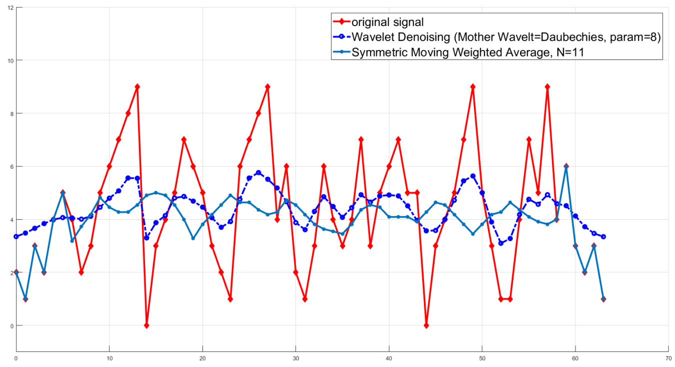
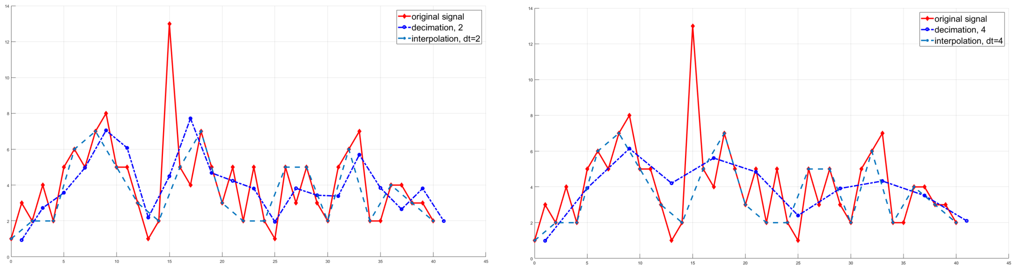


Figure 4.6. An example of the effect of wavelet denoising and moving weighted average.

Time interpolation aims at re-sampling the signal to have data points at fixed time distance. It selects only the values at suitable points and approximates the missing ones. Linear interpolation is the most used version of this technique [55, 57, 61, 62, 63, 64, 65, 19, 20]. Each cited work chooses a possibly different re-sampling rate, from 100 to 500 Hz. Therefore, there is no evidence of a common standard for this operation, which actually depends on the characteristics of the original signal at hand. Non-linear time interpolation is used in [66], in a version based on piece-wise cubic spline. The different sampling rate and interpolation strategy impede the comparison of the achieved results, since the experiments use different datasets.

Figure 4.7 shows an example of the effect of interpolation. Figures 4.6 and 4.7 show that the signal is regularized but also loses some apparently characterizing elements. Looking at the reported results in works using the mentioned techniques, such elements seem to be redundant in terms of recognition accuracy.

Amplitude and/or period normalization are further preprocessing proce-



(a) *Effect of interpolation and decimation with dt=2.* **(b)** *Effect of interpolation and decimation with dt=4.*

Figure 4.7. Examples of interpolation and decimation with different parameters on the same original signal.

dures. Amplitude normalization reduces the signals within a common range of values. A popular formula is the one used for standardization:

$$v_n(i) = \frac{v_o(i) - \mu(v_o)}{\sigma(v_o)} \tag{4.1}$$

where $v_o(i)$ and $v_n(i)$ are respectively the original and normalized signal values, while $\mu(v_o)$ is the mean of the original signal, and $\sigma(v_o)$ its standard deviation.
Period normalization reduces signals to the same length (number of samples) by either up-sampling (e.g., by interpolation to increase the sampling rate) or down-sampling (e.g., by decimation to decrease the sampling rate)[67]. Examples of works using this technique are [64, 65, 68, 69, 70, 71, 59]. Differently from interpolation, period normalization does not set the time distance between subsequent samples, but the compression or extension ratio needed to obtain a signal with fixed time length/number of samples: the acquired values are evenly distributed according to their total number and to the desired final length [70, 71]. Figure 4.7 shows the different effects of these two procedures. The final effect depends on the time interval chosen as parameter. Alternatively, the methods in [59, 60] normalize cycle length by Dynamic Time Warping (usually exploited for comparison only), while [72] exploits self-DTW [73], taking into account the characteristics of a pseudo-periodic signal.

Further preprocessing techniques deal with the relations/combinations of the three signal axes and the way in which they are used for comparison. The magnitude vector is often used instead of the values from the three single axes [55, 63, 74]. For each instant of time i, this vector (also known as g-force vector) provides an acceleration value:

$$v_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \tag{4.2}$$

where x_i, y_i, and z_i are the original values of the signal at time t_i. The main advantage of the magnitude vector is the invariance with respect to the sensor orientation; this aspect is better discussed in Section 4.5.2. In fact, the different orientation of the accelerometer may affect the values of the signals over the single axis. A second gain is the dimensionality reduction from 3D to 1D. However, since this kind of preprocessing is a kind of aggregation, it causes a loss of possible relevant peculiar characteristics over each single axis. In particular, in gait recognition it always happens that one axis (the dominant one) captures more relevant information than the others, depending on the orientation. As a consequence of aggregation, also correlation information is lost. This can be observed in practice and it is reported in literature. Two works [69, 68] combine the axes using a different formula:

$$v_i = arcsin(z_i / \sqrt{x_i^2 + y_i^2 + z_i^2}) \tag{4.3}$$

Other proposals use either the single identified/hypothesized dominant axis [57, 59], or two of them [70, 71]. Finally, as in the proposals presented during this PhD research [75, 76], it is possible to use the three axes as independent data sources and compute results giving different weights to each of them (see Chapter 5).

Table 4.3 summarizes the algorithms discussed so far and schematizes the works exploiting them. It also accounts for the use of magnitude values or other combinations of the axes values. When the signals are not exploited in their original form, normalization is the most frequent preprocessing step. Time is the most used normalization element (see Section 4.4.1). Amplitude normalization mostly relies on mean subtraction. A few works use both normalization strategies. Most works carrying out interpolation exploit its linear form. Finally, Weighted Moving Average is the most used method, if present, for signal denoising. The table does not report any performance measure, since different comparing algorithms and different datasets are used for the experiments. Moreover, no paper compares results achieved using preprocessing with those obtained without it.

4.4 Recognition Methods

As already mentioned, a possible classification of gait recognition in state-of-the-art methods exploiting wearable sensors includes two main categories.

The proposals in the first category generally use algorithms to compare signals. Manhattan distance or Euclidean distance can be used to compute a distance score, but they can be dramatically affected by time misalignment problems, especially if no kind of preprocessing is carried out in advance. Therefore, they are mostly applied as distance measures for DTW and its variations. Some works preliminarily segment the signals to compare into either steps/cycles or fragments/chunks, while a few ones deal with unsegmented signals. Though providing good results, the latter has the limitation to require a sufficient similarity of the lengths of the probe and of the gallery walks. As a matter of fact, DTW often achieves poor recognition results if the signal length difference is too large. For this reason, the majority of the methods in this first category use some kind of segmentation. Afterwards, DTW or similar algorithms are used to compare the detected steps/cycles following different strategies. Considering the importance of step/cycle segmentation and the relevance of DTW in literature works and in this thesis, Section 4.4.1 presents two state-of-the-art segmentation strategies and Section 4.4.2 describes the basic formulation of DTW algorithm.

The proposals in the second category generally exploit k-Nearest Neighbour (k-NN), Support Vector Machine (SVM), Hidden Markov Model (HMM), Convolutional Neural Network (CNN) or other Machine Learning techniques. Even these methods can either use entire signals, or fragments; in this case, a fragment (or chunk) is a part

Table 4.3. Summary of the preprocessing techniques used for wearable sensor-based gait recognition

Paper	Year	Preprocessing			Magnitude (M)/ Other Combinations
		Denoising	Interpolation	Normalization	
[57]	2010	Weighted Moving Average	Linear	-	-
[19]	2008	Weighted Moving Average	Linear	Time (on cycles)	-
[63]	2010	Weighted Moving Average	Linear	Time (on cycles)	M
[20]	2010	Weighted Moving Average	-	Time (on cycles)	-
[60]	2007	Wavelet Denoising	-	Time	-
[59]	2007	Wavelet Denoising	-	Amplitude ([-1,1]) and Time	-
[77]	2007	Direct Form II Transpose filter with a Gaussian window	-	-	-
[78]	2016	Low Pass Finite Impulse Response filter	Cubic Spline	Time	-
[79]	2017	Low Pass Butterworth filter (on cycles)	Linear	-	-
[62]	2011	-	Linear	Amplitude (Mean Subtraction)	-
[64]	2011	-	Linear	Amplitude (Mean Subtraction)	-
[65]	2012	-	Linear	Amplitude (Mean Subtraction)	-
[61]	2011	-	Linear	-	-
[55]	2015	-	Linear	-	M
[66]	2012	-	Piece-wise Cubic Spline	Time	-
[74]	2009	-	-	Time	M
[69]	2006	-	-	Amplitude (Mean Subtraction)	$R_i = \frac{Z_i}{\sqrt{x_i^2+y_i^2+z_i^2}}$
[71]	2005	-	-	Amplitude and Time	-
[70]	2005	-	-	Amplitude and Time	-
[80]	2016	-	-	-	M

of the walking signal of a certain fixed length (either in time or in number of samples). In general, it is possible to notice that the term "segment" is mostly referred to steps/cycles, that have a strict relation with gait physiology, while "fragments" may lack this characteristic.

As already mentioned, the two following Sections deal with step/cycle segmentation strategies and DTW. Section 4.4.3 describes works applying step/cycle segmentation. Section 4.4.4 presents proposals using fragments/chunks, Section 4.4.5 shows works dealing with the whole gait signal. Finally, Section 4.4.6 summarizes the results in tabular form and provides a comparative discussion.

As a further note, it is not possible to present the results achieved by the literature works by using a common evaluation criteria because the different research teams possibly uses different evaluation metrics. For this reason, the results are reported as they are presented in the paper.

4.4.1 Step/Cycle Segmentation Procedures

The possible step/cycle segmentation is a procedure of non-negligible importance. This Section describes two state-of-the-art cycle segmentation procedures [63, 59]. These methods characterize different possible approaches to the problem, and are clearly explained and therefore repeatable, while in other cases the adopted choices are not completely clear.

4.4.1.1 Cycle segmentation based on cycle extremes identification

Rong *et al.* [59] propose a segmentation algorithm which computes results only on the most significant axis, and then projects them onto the other two axes. In the work proposing the discussed segmentation method, the acceleration data are acquired from a sensor attached on the hip. It is worth pointing out that, depending on the sensor orientation, the most significant axis possibly changes. In general, independently from that, the axis that is perpendicular to the ground can be assumed as the principal one. The mentioned work deals with the Z axis. The preprocessing procedure includes signal values normalization to map acceleration data onto the interval [-1,1], and wavelet denoising (Debauchies order 8 as mother wavelet). The phases of the proposed segmentation algorithm are:

1. search local minima in the Z signal;
2. for each local minimum, search the first next value with a different sign (candidate starting/ending points of cycles);
3. segmentation: each gait cycle includes all samples within four consecutive points found above.

This proposal is quite fast and gives sufficiently accurate results. However, it is error prone, especially if the signal falls around the zero value many times. This can be due, e.g., to noise.

4.4.1.2 Cycle segmentation based on cycle length estimation

The algorithm by Derawi *et al.* in [63] is made up by two parts: the cycle length estimation and the segmentation. Preprocessing entails linear time normalization, in order to have samples at a constant $\frac{1}{100}s$ distance, and noise reduction through the weighted moving average algorithm. Finally, the magnitude vector (mv) is computed by the standard formula, and used by cycle length estimation and segmentation phases.

Cycle Length Estimation:

1. pick a sequence sw of 70 samples at random in the approximate center of the mv;
2. compute and store the DTW distances between sw and all the possible sequences in mv with the same length;
3. search local minima among the above distances (call this set LM);
4. the average of the distances of consecutive pairs of local minima is computed as γ (the average cycle length).

Segmentation:

1. choose a point in LM approximately in the middle of the mv (call it P_{start});
2. look for the closest local signal minimum around the point $P_{start} + \gamma$ (and symmetrically for $P_{start} - \gamma$), using Neighbour Search; each found point is the start/stop of a cycle;
3. repeat the search in both directions until to reach the end and start of the signal respectively.

Cycle length estimation is smart and flexible yet computationally demanding. Sometimes the segmentation might not work correctly because it relies on a fixed number of samples, that could be rather chosen depending on the sampling rate of the accelerometer. Being a cycle composed by two steps, it may happen that a user walk starts from a cycle with left step first, while in a different walk by the same user, the right step comes first (and vice versa). Unless controlled during acquisition, this can create not negligible problems in the recognition phase.

4.4.2 Dynamic Time Warping

Dynamic Time Warping is a well-known and robust algorithm. Though not being recent, it is still a widely used signal comparing strategy. It is based on the dynamic programming optimization method. Along the years, it has been and it is still used in many fields such as biometrics, medicine, bioinformatics, gesture recognition, image processing, bioacoustics, finance and so on. It aims at finding the best correspondence between two (time)series, computing the "warp" required to overlap them.

More formally, let $X = (x_1, x_2, ..., x_N)$ with $N \in \mathbb{N}$ and $Y = (y_1, y_2, y_M)$ with $M \in \mathbb{N}$ be two series and $c: X \times Y -> \mathbb{R}$ such that $c_{i,j} = |x_i - y_j|; \quad i \in [1:N], j \in [1:M]$ the cost (or distance) function. In general, different types of distance functions can be exploited. The presented one is just an example of the most used one, based on a simple absolute distance. The DTW algorithm aims at finding the alignment path (or warping path, or warping function) that minimizes the cost from $[x_1, y_1]$ to $[x_N, y_M]$. This alignment path creates the correspondence between elements $x_i \in X$ and $y_j \in Y$. Let k be the length of this path. The searched correspondence has to follow at least two conditions:

1. **Boundary condition:** $p_1 = (x_1; y_1)$ and $p_K = (x_N; x_M)$. This means that the starting and ending points of the alignment path have to be the first and the last pair of points of the aligned sequences, respectively.
   In gait recognition, this aspect is a crucial. In fact, if the step segmentation is not robust, the starting and the ending point will be possibly misaligned, drastically lowering the performances.

2. **Monotonicity condition:** $n_1 \leq n_2 \leq ... \leq n_K$ and $m_1 \leq m_2 \leq ... \leq m_K$. This condition allows to preserve the time ordering of the signal points.

Following these rules, the DTW algorithm creates an incremental distance matrix *DTW* defined as follows:

$$DTW_{i,j} = c_{i,j} + min(DTW_{i-1,j}, DTW_{i,j-1}, DTW_{i-1,j-1}) \tag{4.4}$$

An algorithmic overview of the basic version of the DTW algorithm formulation is presented here:

Algorithm 1 DTW algorithm in its basic formulation.

```
procedure DYNAMICTIMEWARPING(TIMESERIES X, TIMESERIES Y)
    $DTW[length(X)+1][length(Y)+1]$
    for i in X:
        $DTW[i][0] \leftarrow 0$
    for j in Y:
        $DTW[0][j] \leftarrow 0$
    for i in X:
        for j in Y:
            $DTW[i][j] \leftarrow X[i]-Y[j]+min(DTW[i-1][j], DTW[i][j-1], DTW[i-1][j-1])$
    return: $DTW[length(X)+1][length(Y)+1]$
```

As it can be also observed by Algorithm 1, the computational cost of the algorithm is $O(N \times M)$. Assuming that the two signals generally have more or less the same number of samples, the global cost can be considered as $O(max(N, M)^2)$. This can be, in some cases, one of the reasons against the use of DTW. However, considering our application scenario, long-enough gait signals have about 1500 samples and the extracted steps have about 150 samples. So that, even considering to use the algorithm 3 times (one per axis), it requires a computational time in the order of milliseconds. Evolutions of DTW algorithm have been presented along the years in order to speed it up and to increase its flexibility. Some possible adjustments are presented in Section 5.3.3.

A review on the Dynamic Time Warping algorithm can be found in [81].

4.4.3 Systems comparing step/cycle-segmented signals

This Section presents some works in the state-of-the-art which rely on a preliminary segmentation phase. From now on, outliers are defined as all those steps/cycles that are too different from the others (according to some threshold), because of noise or punctual variations. Unless differently specified, reported experimental results are obtained in verification mode.

The system in [57] uses a Google G1 phone equipped with an accelerometer with a low sampling rate. The gait signal is acquired with a round walk in a plain hallway of 37m, with a stop of 2 seconds between roundtrip, so as to consider two separate walks. This work precedes [63], cited above in Section 4.4.1.2. The cycle detection and signal segmentation procedures rely on the use of an average cycle length (ACL) value, which is experimentally estimated here to be about 40-60 samples long (but this would correctly work only for an accelerometer with the same sampling rate). Afterwards, the segmentation proceeds as in [63] (see Section 4.4.1.2). The system computes the average cycle to use it as feature vector and applies the DTW algorithm for comparing the average cycles of a pair of walks. The procedure only processes the x-axis because, with the device orientation used in this work (the long side parallel to the floor), it is the dominant one. This system achieves ERR=20% using a dataset with 51 subjects, each with two roundtrip sessions.

An evolution of the previous work can be found in [63]. The authors use the Motion Recording 100, a better and dedicated accelerometer. Moreover, preprocessing is improved by noise reduction and by detection and removal of outlier steps. The step segmentation algorithm used is the one described in Section 4.4.1.2. For outlier removal, the system computes the DTW distance among all cycle pairs belonging to the same walk signal. Then, the procedure computes the average distance of each cycle from all the others, and finally discards all cycles whose average distance is not in the $\mu \pm 2\sigma$ interval, with μ being the average of the average distances, and σ their standard deviation. All the remaining cycles are used to build the gait template. In recognition phase, the comparison between cycles in the probe template and cycles in the gallery template is carried out using a cross comparison strategy that exploits a Cycle Rotation Metric (CRM). In a first step, using Manhattan distance, the most similar probe cycle is found for each gallery cycle, and then DTW is computed for the best probe cycle/gallery cycle pair. This decreases the computational burden. With respect to the previous work, the performance improves to an EER=5.7% on a dataset of 60 subjects with 12 gait acquisitions each, collected in two different days.

The proposals in [59, 60] present and exploit the segmentation algorithm described in Section 4.4.1.1. In the first one, the recognition relies on a time warping network, based on DTW. The reported results are RR=70% (identification) and EER=6.7% (verification). In the second work, the recognition is performed in both time and frequency domains. A DTW based algorithm in time domain achieves EER=5.6%, while in frequency domain EER=21.1%.

The work in [19] and its evolution in [20] face the problem related to the different shoes worn by the subjects. The experiments are carried out with 4 models of shoes with different weight. The same dataset is used for both works and collects 4 walk signals for each shoe model from 30 male subjects along a 20 meters' hallway, for a total of 16 walks per subject. Both works apply a preliminary interpolation procedure to re-sample the signal at a fixed frequency, and reduce noise through weighted moving average. The part of signal before the true start of walking motion is discarded by using a fixed threshold on the acceleration value, and symmetrically for the end of motion. As claimed for other works, this could limit the portability of the system. Signal segmentation only relies on the up-down axis (here it is x), using three different fixed thresholds to find the cycles starting/ending points. In the first proposal the authors use the average cycle as gait template, and then use the Euclidean distance as dissimilarity measure. In the second proposal, instead, they keep all cycles. In this case, the dissimilarity measure is based on cross comparisons, again based on Euclidean distance, and only the minimum result is taken as distance score. The second approach seems to show a global improvement in terms of performance, achieving an EER down to 1.6% in the best scenario. A first aspect highlighted by these works is that in a mixed scenario involving all kinds of shoes, the results decrease drastically to EER=16.4% in the best test case. As a second aspect, it is possible to notice that the higher is the weight of the shoes the lower the performance, passing from an EER of 1.6% to 6.1%.

An original cycle segmentation algorithm is presented in [80] and tested on an in-house dataset with 4 different recognition strategies. The in-house dataset collects gait signals from 15 subjects (10 males, 5 females), acquired with a Samsung Galaxy Note 4 positioned in the hip zone and attached to a belt, with the long side parallel to the ground. Each walk is captured along a 40 meters hallway in 3 different moments of the day (morning, mid-day, and late afternoon) during two sessions. All subjects are required to wear sweatpants and the same model of shoes. If this on one hand improves the homogeneity of the data, on the other hand it is not suitable for a real system where the users are generally free to wear anything. The system performs a cycle detection procedure based on the identification of high peaks. First, the procedure finds all local maxima. Then removes all those below the maxima average, re-computes a new maxima average, and eliminates again all the maxima below such value. Finally, it discards all maxima that do not have a distance of at least 800ms among them. After that, only the 10 cycles with the lowest average distance from the others are kept. The recognition strategies all exploit well-known comparison methods. The first uses the Euclidean distance and achieves an up to 26.67% EER. The second uses DTW and reaches an up to 28.07% EER. The third is Manhattan Rotation. Manhattan distance is used to compare a template with all the possible rotations of another template, and the best result is kept as the distance between the two. This strategy achieves an up to 16.49% EER. The last comparison is with the CRM, presented in [63] and previously discussed in this Section. This procedure achieves an up to 16.38% of EER.

The system presented in [68] uses two different kinds of approaches to recognition. The walks by 21 subjects are collected with AVR Buttery Accelerometer - Motion Recording 100 along a 70 meters long hallway, and manually divided into two segments (probe and gallery). Both proposed recognizers use the same preprocessing in which the values from each axis are combined as $v_i = arcsin(Z_i / \sqrt{X_i^2 + Y_i^2 + Z_i^2})$, producing a mono-dimensional signal. The first comparing strategy uses histogram similarity, where histogram bins correspond to quantized values of the original signal. This strategy achieves EER=5%. The second approach uses cycle group comparison, i.e. comparison of all cycles in a walk. To identify a cycle, it uses the 0-cross points in order to segment the signal. This, in general, is not guaranteed to work properly, especially with noisy data. The recognition is performed by comparing signals cycle by cycle, and the final score is the number of cycle pairs whose distance that meet a fixed threshold. This strategy achieves a 9% of EER.

The gait recognition proposal in [66] exploits the gyroscope too. The data acquisition is performed by an Android smartphone (the model is not specified) equipped with accelerometer and gyroscope. The smartphone is oriented vertically, facing outwards, and is carried in the right pocket. Walk signals produced by 36 volunteers make up the dataset, collected along a hallway of about 25 meters long in a roundtrip pattern, at different speeds (pace, fast). After the acquisition, the raw data are re-sampled at 2ms intervals (or 500Hz) with piece-wise cubic spline interpolation. Then, the signal is segmented into cycles exploiting the high peaks.

Feature extraction relies on continuous wavelet transform; the recognition phase exploits the analysis of both time frequency spectrogram, and of ciclostationarity. Results are also provided for cross-speed comparisons. The tests at 0.1% of FAR achieve 99.4% of verification rate for pace vs. pace, 96.8% for fast vs. fast, and 61.1% for pace vs. fast respectively. It is noteworthy that the very good results are achieved using the gyroscope too.

The system in [70] is based on the concept of *gait_code*. The used benchmark is a dataset of 36 subjects (2 sessions per subject with 3 walks each, captured at 5 days distance). The *gait_code* computation stems from the fact that it is not interesting to distinguish the right from the left steps, but rather to take into account that the signals that they produce are different. Therefore, it is sufficient to classify the two categories of steps as *a* steps and *b* steps. The *code* is a 4-tuple, composed of the average of *a* steps and the average of *b* steps over two out of the three accelerometer axes. This leaves out the averages over the less significant axis, which in this work is identified as Y, therefore producing $gait\_code =< xa, xb, za, zb >$. In the recognition phase, the signals are compared taking the best correlation of probe and test codes, considering the possible correspondences between probe and test code pairs of components on the same axis. This strategy produces a 6.4% of EER. Variations of this approach are presented in [71]. The walk signals are acquired from 36 subjects at three speeds (slow, normal, and fast). A first variation entails signal preprocessing by Fast Fourier Transform (FFT) after dividing the signal into frames that are 256 samples long and with an overlap of 128 samples. A feature vector is computed by chaining the first 40 FFT coefficients per channel. The second variation entails computing the feature vectors by chaining of 10-bin histograms of accelerations over x and z axes, after normalization by signal length. A final variation entails concatenating moments of third and fourth order for both axes. The achieved EER are respectively of 7% (with the correlation-based method from the previous work), 10% (FFT coefficients), 19% (histograms), and 18% (moments). It is worth underlining that recognition of the right subject but with a different level of speed is considered as a wrong decision.

An approach based on Signature Points (SPs) is described in [74]. The idea behind the SPs is to search for informative points in gait acceleration signals and takes inspiration by SIFT for 2D images. Data are collected with 5 Wii Remote (from the well-known console) located in different body locations. These sensors have lower accuracy than the others. The dataset used for experiments contains 30 subjects with 6 walks each in 2 sessions (12 walks total per subject). Acceleration data are combined in the magnitude vector form, using the standard formulation. After that, the walk signals are segmented into cycles (but the used procedure is not explained), and then normalized in length. Finally, the proposed procedure extracts the SPs, corresponding to the positions of the starting and ending points of gait cycles. The positions of such points are used for recognition, with majority voting. The proposed system achieves up to 74.5% of RR with a single accelerometer (the one located in pelvis zone) and a 96.7% of RR combining all five accelerometer signals.

Machine learning techniques are applied in [77] on a preliminarily segmented signal. Acceleration data are collected using Wii Remote controller in three different walk modalities: straight-line, circuitous and multi-directional, and straight-line with distorted speed. The dataset includes only 7 subjects (up to 3 acquisitions each for each kind of walk). Starting and ending points for cycle segmentation are the relative maxima values in the signal after the application of Direct Form II Transpose filter [82]. The recognition is done by three different machine learning techniques, namely k-NN, Naïve Bayes and Quadratic Discriminant Analysis, and achieves an up to 95% of accuracy with k-NN, which is generally the best, on straight-line walks, where all methods work obviously better. However, this result is obtained with a very small dataset.

The work in [78] presents IDNet, an authentication framework relying on Convolutional Neural Networks (CNN). The CNN is used as feature extractor, and then One-Class SVM (OSVM) is used for verification. The experiments exploit an in-house dataset with walking signals acquired from 50 subjects over six months. Different smartphone models (namely Samsung S4, Samsung S3 Neo, LG G2, LG G4, Google Nexus 5, and Asus Zenfone 2) are positioned in the trousers' right front pocket. Subjects walk at their normal pace in different walking sessions of about 5 minutes. Information from accelerometer, gyroscope, and magnetometer are stored into the walking file (magnetometer data are not used in the recognition process). The preprocessing phase exploits cubic spline interpolation at 200Hz and denoising by a low pass Finite Impulse Response (FIR) filter with a cutoff frequency of 40Hz. The walking cycles are extracted from the walk magnitude vector, after low-pass filtering with a cutoff frequency of 3HZ. Then the minima are found and refined in order to better fit with the original signal. An orientation invariant transformation is then applied to the identified cycles, using the values taken from the original signal. The transformation maps all data onto a vector space defined by 3 identified invariant versors. Finally, data are normalized both in time (with spline interpolation to reach 200 samples per cycle) and in amplitude (with the standard Gaussian normalization). A CNN is trained for feature extraction using data from 35 out of the 50 subjects. The CNN is trained using as input the extracted and normalized cycles. It is made up by a stack of 2 convolutional layers, a max pooling layer, and 2 fully connected layers. 10 different networks are trained to find out the best parameters. After feature extraction, a OSVM is trained for each of the remaining 15 users with the derived feature vectors, using part of the cycles from their walks. Each out of the 15 users is taken as target in turn (and the other 14 are impostors), and then results are averaged. The best OSVM configuration achieves FAR and a FRR less than 0.15% with the appropriate parameters. The authors also investigate how many walking cycles are needed for a reliable biometric recognition, and report that using less than 5 cycles can be sufficient to recognize the target or reject the probe in the 80% of the cases. Finally, the use of gyroscope in addition to accelerometer is tested, providing a slight improvement.

Deep convolution neural networks are exploited in [79]. Experiments are carried out on the ZJU-gaitacc dataset, already mentioned and described in Section 4.2. It is worth reminding that the dataset contains 6 signals per walk coming from 5 different accelerometers. The processing phase consists of three parts: cycles extraction, filtering, and normalization. The cycle extraction is carried out exploiting the strong changes of values (peaks) on the z-axis occurring when the heel impacts the ground, which causes peaks within the magnitude vector. Noise reduction and signal enhancement are obtained by a Butterworth low pass filter, and by linear interpolation. The deep network consists of two convolutional layers, a max pooling layer, two fully connected layers, and a final softmax layer for template classification. The implementation exploits the Caffe framework. The tests are performed using 5 out of the 6 walk signals per subject as training set, and one for test experiments. Only a single session in the dataset is used. The creation of artificial training samples aims at preventing overfitting and improving the accuracy of the deep network results. This data augmentation is obtained by deforming the labeled data, while maintaining the meaning of the labels. The re-identification accuracy of the proposed scheme is quantified as the average number of correctly recognized cycles for each identity, i.e., appearing in the first, second and third place of the list of results ordered by similarity. The authors report an accuracy of 0.94% without augmentation and a little improvement up to 0.95% of accuracy with data augmentation. It is worth pointing out that the paper probably reports an error, because the correct values seems to be either 94% and 95% respectively, or 0.94 and 0.95 respectively.

4.4.4 Systems Comparing Timed Chunks of Signal

The procedures presented in this Section do not rely on a preliminary segmentation phase, but rather use some other strategies, taking differently determined fragments of the signal for comparisons.

The paper [83] deals with accelerometer signals captured by a smartphone. The raw time series are first transformed into "examples", because the exploited classification algorithms [84] do not process time series data. In order to collect experimental data, 36 volunteers perform a specific set of activities, i.e., walking, jogging, climbing upstairs, and climbing downstairs, for a specified time, while keeping an Android smartphone in their pant front pocket. Data collection stops at subjects' activity switches. The data is segmented into 10-second chunks (referred as example duration - ED). Afterwards, 43 features are extracted from the accelerometer values of each segment (600 total acceleration values per segment, 200 per axis, given by taking 20 samples per second from each ED). Features are mostly computed on a per axis basis, and are variations of six basic ones: Average, Standard Deviation, Average Absolute Difference, Average Resultant Acceleration, Time Between Peaks, Binned Distribution. Details on the measures can be found in the paper. Data is used to generate six distinct datasets. Four of them contain examples separated by activity, to assess their effect on recognition. An "aggregate" dataset is created by removing the activity labels and mingling examples from all activities (the most realistic scenario). The last "aggregate (oracle)", contains the same examples as the "aggregate" but activity-labeled, allowing to evaluate the utility of a preliminary activity detection to estimate the class label. Decision Trees and Neural Networks are used to train single subject models and for recognition. Authentication results seem missing. Identification results are also reported on a per subject basis, and seem to be evaluated in homogeneous conditions, except for the aggregate dataset. The best performance is achieved by walking and jogging data. Ascending/descending stairs provide significantly lower accuracies, but there is much less data available for building the models.

The works in [61] and in [62] provide two solutions for gait recognition, both tested on the same dataset of 48 subjects with 4 walks each, that was collected throughout two sessions with a Google G1 phone. In the first work, the data are re-sampled at 200Hz and then divided into fragments (walk sections) of 3 second without overlap (for sake of homogeneity, 28 fragments are stored per subject). For each subject, a Hidden Markov Model (HMM) is trained using 20 of such fragments as positive examples, while the remaining fragments are used for testing. For each HMM, the set of negative samples includes 840 fragments taken from 30 subjects (that will not be included in testing). A test is carried out for each HMM/subject, to imitate the situation of the single enrolled subject, i.e., the phone owner. Testing is carried out with the remaining 8 fragments as genuine samples, and 476 samples from the 17 subjects not involved in training as impostor samples. The experiments exploit the HVITE tool and each subject is used one time as genuine and forty-seven times as negative training/impostor. This strategy reports an EER of about 10%. In the second work, the signals are interpolated at 100Hz and divided into fragments (segments) of 7 seconds with a 50% overlap. Since SVM is used for classification, the acceleration data from each fragment is is used to build a feature vector of fixed size. Features are mostly statistical ones, e.g., mean, maximum, minimum, binned distribution, etc., with the addition of both the Mel and the Bark frequency cepstral coefficients. Radial Basis Function (RBF) kernel is chosen for the SVM. Walks from the first day make up the training set and those from the second day are used for testing. The presented strategy initially achieves a very high FNMR (above 61%) while the FMR is reasonable (above 1% but below 2%) for all tested configurations (e.g., subset of features or acceleration axis). In order to reduce FNMR, the system exploits a quorum voting (at least a certain number of positive responses must be obtained for the probe to be accepted) using all test fragments of a user, treating each of them as an independent test. The same authors had experimentally demonstrated that this strategy works better than majority voting when the difference between FMR and FNMR is very high. In fact, the system achieves a down to 5.9% FMR and a down to 6.3% FNMR.

In [65] a k-NN approach is used. The walk signals are collected from 36 subjects during two sessions made up by 12 walks at normal pace, 16 walks at fast pace and further 12 walks at normal pace on a flat hallway. Walks are separated by stop periods (when the subject is not walking) by an automatic procedure and then manually corrected when necessary. The extracted signals are interpolated at 127Hz. Walking signals are divided into fragments. Each round of tests uses a different fragment size (3s, 5s, and 7.5s respectively) always with a 50% overlap. Similarly to [62], the feature vectors are built by some statistical parameters and both Mel and Berk coefficients extracted from the three axes and the magnitude vector. The comparison subsystem exploits the k-NN algorithm included in the WEKA library. As in [62], a quorum voting approach is introduced to reduce FNMR. The best reported performance among the tested configurations is EER=8.24%

The work in [85] proposes a system where the mobile acquisition device can be kept anywhere. This approach presents some peculiar aspects; therefore it will be described in more detail.
Three datasets are collected (using Intel Xolo, Samsung Galaxy S3, Samsung Galaxy S4, and Google Nexus 5) for different aims. The first one collects data from 47 subjects (19 females and 28 males), naturally performing different activities with the same frequency, while keeping the phone in different body placements. These

acquisitions sum up to 18 hours of data. This dataset is used to train a walking detector, determining whether the input frame is a walking one. The same dataset is used also to train a Gaussian Mixture Model – Universal Background Model (GMM-UBM), representing a distribution of the gait patterns which is independent from subjects under various conditions (stereotypical patterns). Two alternatives are tested for a second step of training, either supervised or unsupervised, to recognize single subjects from more specific features. The supervised alternative is evaluated using the second collected dataset, with data from 12 subjects (5 females and 7 males); they carry the phone in 2 or more positions (e.g., pant front pocket, jacket pocket, and bags), in multiple sessions and at different speeds. Data is annotated accordingly. These walking acquisitions sum up to about 1 hour of data per subject. For each user, an individual gait model is trained by 1/3 of such data; the remaining part is used for testing. The third dataset is used for evaluating the unsupervised version of the second step of training. Unlabeled accelerometer data are collected from 8 subjects 24/7 during two or three weeks. On average, the mentioned walking detector returns a total of 5 hours of data per subject.

All datasets undergo the same preprocessing phase. Accelerometer data are fragmented into frames of 512 samples each with an overlap of 50%. All stationary frames are discarded using an energy threshold. After this first selection, a low-pass mean filter is applied to each sample in the frame in order to estimate the direction of gravity, represented as the triplet of average values of respectively x, y, and z axes over all samples in the frame. If this triplet changes drastically with respect to the immediately preceding frame, the current frame is discarded, since it is possible that an abrupt change in device orientation happened. The low pass filtered remaining points are considered as a vector and remapped into two rotation invariant vectors, one for its vertical component and one for the horizontal one. The frame is represented as the set of component pairs of its samples.

The proposed system extracts from each frame and combines time and frequency domain features, and auto-correlation features. A small set of these features is used for walking detection. This is carried out exploiting a decision tree classifier achieving a precision of about 98% and a recall of about 95%.

After walking detection, the relevant features for gait analysis are extracted from walking frames only, such as the compressed sub-band cepstral coefficients, inspired by MFCC (Mel-frequency cepstral coefficients) feature set for audio analysis, widely used in both speech recognition research as well as for speaker identification.

The system achieves about 14% of EER of when user gait training exploits 20% of labeled data (using the second dataset). With unlabeled data (third dataset), the results get worse by about 5%. On one hand it seems strange that the difference is not so high, on the other hand this is probably due to the larger amount of data being exploited to train the system when using unlabeled data.

4.4.5 Systems Comparing Unsegmented Signals

One of the main advantages in the use of the entire signal is that in this case the co-articulation between the single steps/cycles is maintained, whose effect might be considered similar to the phenomenon observed in speech, where the pronunciation of a vowel or consonant can be influenced by the preceding one [86]. Similarly, we may assume a different step/cycle starting or ending shape according to the continuation of the movement flow characterizing the previous step/cycle. This might especially hold with short walks (see Section 4.6.1).

The proposal in [55], already mentioned in Section 4.2, presents an evolution of [74]. The collected data are converted into magnitude vectors using Equation 4.2. Differently from the previous work, the steps segmentation procedure is substituted by an improved version of the signature points extraction. The signature points are taken as the extrema of the convolution of the gait signal with a Difference of Gaussian pyramid. An extremum is a point being greater or smaller than all of its eight neighbors. These points are then stored as vectors for a sparse representation. Then they are clustered, linearly combined and saved into a dictionary, therefore obtaining a single element for each cluster. The system considers the recognition as a conditional probability problem and uses a sparse-code classifier. The results are very interesting and reach an up to 95.8% of RR (identification), and a down to 2.2% of EER (verification), even if is worth considering that the approaches in both works uses 5 accelerometer worn in different body locations.

4.4.6 Summary tables of state-of-the-art proposals

Table 4.4 summarizes the state-of-the-art approaches to gait recognition by wearable sensors. It schematizes the works in the same order as they have been presented in this Section, maintaining the subsection subdivision. The only work exclusively exploiting unsegmented walk signals is the last one in the table. For sake of space, details about the datasets used by each work are reported separately in Table 4.5, where the same ordering of papers is maintained for an easiest cross-reference.
Table 4.4 raises some interesting considerations. State-of-the-art proposals based on fragmentation, at the best of our knowledge, or according to the reviewed literature at least, always use a Machine Learning-based recognition strategy. On the contrary, those which rely on step/cycle segmentation generally, but not always, exploit signal comparing strategies based on distances or similarities. All works aim at improving the quality of the signal and/or at processing data of an equal length and/or with a same frequency. These works use some preprocessing. As already mentioned in Section 4.2, and as it is even more evident from Table 4.5, the majority of the proposals use in-house datasets as test-bed, which (often significantly) differ for number of subjects, acquisition device(s), device(s) position(s) and possibly number, and length of the acquired signals per subject. Moreover, some datasets are collected during a single session, and sometimes with data coming from a single walking sequence divided into two or more parts. This is the case, e.g., with the huge OU-ISIR dataset [54]. This factor drastically limits the inter-subject differences and makes the recognition task easier than the comparison of data really collected in more than one session. For these reasons, though performance measures are reported in the text for readers' interest, their full and significant comparison is not easy and often even unfeasible. Regarding device positioning, the most popular locations are the central parts of the body (namely the ones described as Be, W, H, and TP in Table 4.5). This is probably due to the observation of the intrinsic physical characteristics of human body and of gait kinematics, that make this zone a privileged center of gravity and therefore a good candidate for acquisition. Moreover,

they are the more realistic positions where to put a smartphone, especially for the TP, even if it can not be true for women. Another factor in Table 4.4 is that most works test the proposed methodology in verification applications, assuming an implicit identity claim corresponding to the owner of the device. Even though identification applications are sometimes reported, it is always closed set. This trend is quite common to most biometrics, due to the fact that open set identification usually dramatically affects performance in a negative way. However, this kind of application is the most realistic too, therefore recent works are facing it [87, 88].

The results of the proposed recognition strategies, even if not comparable, are generally quite interesting, especially considering gait as a soft biometrics. This makes it a good candidate for complementing other strong biometrics. However, it is to say that extensive datasets and evaluation are still missing for completely uncontrolled conditions and for massive variations regarding ground slope, shoe type, and other sources of variation that may affect gait signal.

4.5 Gait Recognition in less controlled conditions

This Section summarizes some outcomes related to the proposals either involving smartwatches as source for gait signals or facing the acquisition of the gait signal in less controlled scenarios. These two topics are joined in the same section because the smartwatches acquisition implicitly involve less controlled conditions possibly due to both the gesticulative arms movements and other actions involved during the walking, such as the response to a phone call.

4.5.1 An alternative to smartphones: smartwatches

The preceding literature review shows that several studies on gait recognition based on wearable sensors exploit smartphones, since accelerometers and gyroscopes are practically ubiquitously embedded in these devices. This allows to exploit them for free. Smartwatches are an apparently promising alternative, yet not significantly explored till now. They are included in this Section, because, if used normally, the captured signals are affected by a variety of factors extraneous to gait kinematics itself. Arm movements, either voluntary or involuntary, create an acceleration effect that overlaps, if not completely cancels, the one produced by walking. Due to the nature of the captured data, works in literature started using these devices mostly for activity detection and recognition. One of the first (hypothesized) uses of smartwatches in addition to body-worn sensors for activity recognition is described in [89]. The extension to gait recognition is more recent.

The work in [90] does not exploit a smartwatch, but it is interesting to discuss it in this context, because it adopts a kind of processing that can be extended to such devices. The experimental set-up includes a pair of Shimmer3 IMU sensors, one positioned on the right tight (like a smartphone) and one positioned on the left wrist (like a smartwatch). The captured signals are processed and classified independently: only the comparison of the achieved performance is reported, with no attempt for fusion.

The dataset is collected from 15 subjects, walking six times along a hallway,

Table 4.4. Surveyed methods divided by recognition strategies

Paper	Year	System				Dataset	Results	
		Machine Learning	Preprocessing	Methodology / Feature Extraction	Comparison Strategy		Biometric Modality	Performance
[57]	2010	NO	D, I	Uses only x-axis, CS, ACE	DTW	In-house	V	20.1% EER
[63]	2010	NO	D, I, M	CS, OR	AaA using CRM with DTW	In-house	V	5% EER
[60]	2007	NO	D, N	(TD) CS, (FD) Discrete FT	(TD) DTW, (FD) FT coefficient similarity	In-house	V	(TD) 5.6% EER, (FD) 21.1% EER
[59]	2007	NO	D, N	CS	DTW with ED	In-house	V CSI	6.7% EER, about 70% RR
[19]	2008	NO	D, I, N	CS, ACE using median	ED, weighted ED	In-house	V	7.2%-5% EER, 5.6% EER, 15-12.8% EER, 8.3-7.8% EER
[20]	2010	NO	D, I, N	CS	CrC ED taking MIN	In-house	V	1.6% EER, 2.8 EER, 5% EER, 3.3% EER
[80]	2016	NO	M	CS, keeps the best 10 cycles	ED, DTW, MR, CRM	In-house	V	ED 26.67% EER, DTW 28.07% EER, CRM 16.38% EER, MR 16.49% EER
[68]	2006	NO	N, O	CS, H	CC, HS	In-house	V	HS: 5% EER, CC: 9% EER
[66]	2012	Both	D, I	CS, OR, CWT (MHW)	TFSA using CA with 3 different procedures SVM	In-house	V	99.4% VR n vs. n, 96.8% VR f vs. f, 61.1% VR n vs. f, at 0.1% FAR
[70]	2005	NO	N	SS, Uses only x and z axes (GC)	CrC	In-house	V	6,4% EER

[71]	2005	NO	N	SS, Uses only x and z axes (GC), FFT feature, H	(TD) CrC taking MAX, HS, HOM, (FD) FFT feature comparison	In-house	V	7% EER NOTE: Right user yet recognized with the wrong speed is considered a wrong response
[74]	2009	NO	N, M	CS, SPs	MV of SPs label prediction	In-house	CSI	Wrist 66.8% RR, Upper arm 74.5% RR, Waist 70.1% RR, Thigh 67.5% RR, Ankle 72.9% RR, All five 96.7% EER
[77]	2009	k-NN NB QDA	D	CS, k-NN, NB, and QDA classifier training	k-NN NB QDA	In-house	V	up to 95% of accuracy with k-NN
[78]	2016	CNN	I, N, O	CS, OIT, CNN training	CNN-based Classification	In-house	V	FAR and a FRR less than 0.15% with appropriate thresholds
[79]	2017	Deep CNN	D, N	CS, CNN training, DA	CNN-based Classification	ZJU-gaitacc	V	0.94% (?) accuracy (w/o DA) 0.95% (?) accuracy (w DA)
[83]	2010	Decision Tree, Neural Network	F	Feature Extraction Training of Decision Tree Training of Neural Network	Decision Tree Classification Neural Network Classification	In-house	V CSI	Results provided only for single user
[61]	2011	HMM	I	F, HMM Training (with Baum-Welch algorithm).	HMM	In-house	V	10,42% FNMR, 10,29% FMR, about 10% EER
[62]	2011	SVM	I	F, MFCC and BFCC extraction, SVM training	SVM	In-house	V	5,9% FMR with 6,3% FNMR
[65]	2012	k-NN	I, N	F (3 dimensions), Feature extraction.	k-NN with ED	In-house	V	8,24% HTER

[85]	2014	GMM	D, O	F, Walking frame detection, Feature Extraction (TD and FD), GMM-UBM training, GMM training (specific user)	GMM	In-house	V	Walk stance detection accuracy of 98%, 14%-20% EER labeled, about 25% EER unlabeled
[55]	2015	SPs with Clustering	I, M	SPs extraction, Sparse Representation of SPs, Clustering	Sparse Code Classifier	ZJU-gaitacc	CSI, V	up to 73,4% RR (H) and down to 8,6% EER (T), 95.8% RR, 2.2% EER with all five sensors

Legend of acronyms:

AaA=All against All
ACE=Average Cycle Extraction
BFCC=Bark freq. cepstral coeffs
CA=Cyclostationarity Analysis
CC=Cycle Comparisons
CrC=Cross Correlation
CRM=Cyclic Rotation Metric
CS=Cycle Segmentation
CSI=Closed Set Identification
CWT= Continuous Wavelet Transform
D=Denoising
DTW=Dynamic Time Warping
ED=Euclidean Distance
F=Division into Fragments
FD=Frequency Domain
FT=Fourier Transform
GMM=Gaussian Mixture Model
H=Histogram
I=Interpolation
k-NN=k-Nearest neighbors
M=Magnitude
MFCC=Mel freq. cepstral coeffs
N=Normalization
NB=Naive Bayes
O=Other kind of combination
OR=Outlier Removal
QDA=Quadratic Discriminant Analysis
SA=Spectrogram Analysis
SPs=Signature Points
SS=Step Segmetation
TD=Time Domain
TF=Time Frequency
UBM=Universal Background Model
V=Verification

Table 4.5. The main characteristics of the datasets exploited in the state-of-the-art works. The datasets before the double line are used in works relying on step/cycle segmentation, while the others are used in works that process fragments of signals.

Data		Dataset						
Paper	Year	Device(s)	Device(s) Position(s)	Subjects	Walk Length	#Walks per subject in each session	#Sessions	Notes
[57]	2010	Google G1 (AK8976A) 50Hz	R.H.	51 (41M-10F)	37m+37m	2 (from the same walk)	2	-
[63]	2010	Motion Recording 100Hz	Be (L.H.)	60	20m+20m	6	2	-
[60]	2007	MMA7260	B.W.	21 (11M-10F) age [19,40]	30m	5	1	1 per user for training and 4 for test (randomly chosen)
[59]	2007			35 (19M-16F) age [20,45]				
[19]	2008	Motion Recording 100Hz	R.A.	30 (30M)	20m	4 (per shoes model)	1	4 different shoes model
[20]	2010							
[80]	2016	Samsung Galaxy Note 4 100Hz	H	15 (10M-5F) age [21,38]	40m+40m	6 (different day time)	2	Same shoes model for all subjects, all subjects wear sweatpants
[68]	2006	AVR Butterfly (ADXL202) 16Hz	R.A.	21 (12M-9F) age [20,40]	35m+35m manually divided	2 (from the same walk)	1	-
[66]	2012	Not specified Android device	R.TP	36 (28M-8F)	25m+25m	3 pace + 3 fast	1	2 levels of speed
[70]	2005	ADXL202JQ 256Hz	B.C.W.	36 (19M-17F)	20m	3	2 (5 days delay)	-
[71]	2005							
[74]	2009	Wii Remote 100Hz	L.Ar., L.Wr., R.H., L.T. R.A.	30	20m	6	2 (from 1 to 2 months delay)	-
[77]	2009	Wii Remote + Motion Plus (gyro) 50Hz	TP	7	not reported	up to 3 times per different type of walk	1	normal straight, circuitous, multidirection with start&stop, distorted speed

[78]	2016	Asus Zenfone 2, Samsung S3 Neo, Samsung S4, LG G2, LG G4, Google Nexus 5	F.R.TP	50	not reported	variable	variable (sessions last about 5 minutes)	-
[79]	2017	ZJU-gaitacc dataset						
[83]	2010	not reported	TP	36	variable	variable	1	Walking, Jogging, Stairs Up, Stair Down
[61]	2011	Google G1 (AK8976A) 50Hz	R.H.	48 (38M-10F)	37m+37m	2 (from the same walk)	2	26.5s per walk 1108 samples (less then 42Hz)
[62]	2011							
[65]	2012	Motorola Milestone 100Hz	R.H.	36 (29M-7F)	about 15 minutes	12 normal pace 16 fast pace 12 normal pace	2 (24 days delay)	-
[85]	2014	Samsung Galaxy S3, Samsung Galaxy S4, Google Nexus 5, Intel Xolo	variable	3 different datasets	variable	variable	variable	The 3 datasets are used for different tasks

Legend of acronyms:

A=ankle
Ar=arm
B=back

Be=belt
C=center

F=front
H=hip

L=left
R=right

T=tight
TP=trousers pocket

W=waist
Wr=wrist

twice at their normal preferred pace, twice at fast pace and twice with hands in their pockets. Finally, subjects are asked to perform random gestures to simulate false cycles. Capture appears to happen in a single session. The first step of the common workflow entails walking detection, through peaks of the magnitude vectors. It is interesting to point out that, while the general trend of the two signals is very similar, the wrist one has a much lower amplitude. The same kind of sensor is used in both positions, therefore this is probably due to the attenuation of the acceleration produced by gait when the accelerometer is farther from feet or from a body part directly involved in the movement, especially if the arms do not swing. As a consequence, the threshold used for magnitude peaks detection is lower. Afterwards, since the wrist movement is more irregular than the movement of legs, an autocorrelation-based filter is used for the wrist signal to discard too irregular cycles. The following processing is the same for the data from both sensors. A Butterworth filter at 20 Hz is applied and three additional vectors are computed for each cycle besides those provided by the accelerometers on the x, y and z axes, namely magnitude, and horizontal and vertical components computed as in [85]. The six vectors are used in different combinations for statistical feature extraction (max, mean, median, standard deviation, skewness, kurtosis, autocorrelation of feature samples, etc.). Common features are added, e.g., signal length, to compose the final feature vector. Three approaches are compared. In the first one, data from the pocket sensor is used, considering magnitude, horizontal and vertical components (mhv). In the second and third approach, data from the wrist sensor are used, considering either magnitude, horizontal and vertical components (mhv), or x, y, z axes and magnitude (xyzm) respectively. During verification (against the owner of the device), the same procedure is carried out and an anomaly score is assigned to each probe cycle with respect to subject gallery cycles using Euclidean distance and Nearest Neighbor. A first part of results reports a comparison between walking detection from data from the two positions, showing that the algorithm (with the necessary described modification) produces a similar number of detected cycles, sufficient for subject recognition. This is also due to the autocorrelation filter applied to the wrist signal that is able to discard all the random hand movements. Subject verification is both evaluated by AUC from ROC curve, and EER. As for pocket sensor, the system achieves AUC=99.6% and ERR=2.5%. The interesting information comes from comparing the results from the two approaches using wrist sensor. For mhv the results are AUC=97.3% (only slightly lower than pocket), but EER=8%, with a dramatic worsening. On the contrary, for xyzm AUC=99.6% (overall, user verification is scarcely affected either by sensor position or by the choice of feature axes) but EER=2.9%, therefore returning to values comparable to the pocket setting. This seems to demonstrate that information coming from wrist sensors is more effectively exploited using data from x, y, z axes and magnitude. The results are improved by applying the auto-correlation filter, with a higher impact on wrist data. Identification is also assessed, to take into account a scenario where different subjects can wear the sensor from time to time. To this aim, 5 classifiers are compared, namely 1-NN, Multilayer Perceptron, Random Forest, Rotation Forest, and Multinomial Logistic Regression. The best result is achieved by 1-NN with pocket sensor, while on average pocket (97.3%) and wrist with xyzm (97.4%) are almost identical.

The paper [91] proposes to pair a smartphone and an LG G smartwatch on the non-dominant hand. This choice allows less noise created by common actions usually carried out by the dominant one. Data are captured from both accelerometer and gyroscope of the smartwatch, while the smartphone is exclusively used to trigger data collection, receive data and retransmit them to the server for recognition. A dataset is collected from 59 subjects during a 5 minutes' walk. The overall approach, both the strategy for signal segmentation and the number and type of feature extracted, are the same used for smartphone [83]. The classifiers tested in this case are Multilayer Perceptron, Random Forest, Rotation Forest, and Naive Bayes. A 59-class training is carried out for experiments in identification modality. Half user data is exploited to train an individual authentication model. Results are mostly reported on a per user basis, but in general results provided by accelerometers data are better than gyroscope ones.

The paper [92] uses a real smartwatch (a Samsung Gear Live Smart Watch) for implementing the Gait-watch proposed system. The authors propose a kind of context-aware gait-based user recognition, that takes advantage of a preliminary activity detection. Training dictionaries are collected for the different actions taken into account, so that identification is performed on the corresponding dictionary. An activity dataset from 15 subjects is collected to train the system for the preliminary action recognition. The 7 different activities include 3 ones entailing arm swing (normal walking, walking upstairs and walking downstairs) and 4 ones without (walking while writing text on the phone, walking by making a phone call, walking with hand in jacket pocket, and walking with hand in pant pocket). Each subject is asked to walk for 3 minutes for each activity, for a total of about 5 hours of training data. Gait signals are segmented in cycles for training purposes, and dictionary creation is improved by a projection optimization algorithm [93]. The gait dataset used for subject recognition includes 20 volunteers in two sessions of data acquisition separated by a week. The terrain includes plain, grass and asphalt (but slope are not mentioned) and subjects' dresses and shoes may change across sessions. Subjects perform the considered activities in equal proportion, for a total of 10 hours of gait data. A 10-fold cross-validation is carried out. During testing, the first processing step is walking detection. The best classification accuracy is achieved by adopting a K-NN classifier with K=3 and a window size of 2s (98.6%). After activity recognition, gait cycle segmentation is carried out. The raw accelerometer signals along the three axes fluctuate in a significant manner due to arbitrary body movements. However, a kind of regular pattern is produced by gait along the gravity direction, so that gravity coordinates can be used, obtained applying the rotation (R) matrix provided by Android to transform acceleration data over x, y, and z axes into E(st), N(orth) and G(ravity). After applying a Butterworth filter, gravity-related peaks allow walking detection using the approach in [94]. A sparse fusion method combines information from different gait cycles. Subject recognition is carried out both without and with preliminary action recognition (after a corresponding kind of training), with an improvement of up to 20% of Recognition Accuracy (the rate of correct classifications) with respect to the so defined conventional method that uses a common pool of undifferentiated gait cycles to train subject recognition.

Regarding the best choice of features to exploit for smartwatch-based gait recognition, the experiments in [95] explore both time-related and frequency features.

Moreover, both accelerometer and gyroscope signals are captured. The analysis is carried out on a dataset of 36 subjects, with gait captured during 6 sessions of about 2 minutes each at normal speed in different days. Besides walking, the subjects also carry out actions like opening a door and turn. The reported analysis compares data for single subjects, taking into account elements like same/different session and the number of features exploited. The proposed solution is a dynamic feature vector for each subject.

4.5.2 Device Orientation and Phase Changes

One of the possible problems in gait recognition via wearable sensors is strictly related to the orientation of the acquisition device, that can cause huge differences among signals even for the same subject. The simplest solution (see Section 4.3) is to use the magnitude vector instead of the three separated axes values, but the correlation between the three axes is lost. Other kinds of solutions are generally more complicated and involve geometric transformations of a time series within a 3D space.

The approach in [96] computes orientation invariants for both accelerometer and gyroscope data. These are exploited to extract sensor orientation-robust features, able to characterize locomotion dynamics. A two dimensional matrix (Gait Dynamics Image) stores interactions within single cycles and time-invariant motion dynamics, extracted from an acceleration (rotation) time series with fixed intervals. Matrices are compared after extracting i-vectors, as for speech processing.

A different approach to the sensor orientation problem is proposed in [97], where signal correspondence is computed using cyclic dynamic programming and sensor-orientation estimation. Even in this case the magnitude vector is exploited as a rotation invariant feature.

Phase registration is a possible problem that may affect gait signal comparison, especially with unsegmented signals. The work in [72] proposes to exploit a linearized Time Warping Function (TWF) on the gallery samples to normalize signals and improve the comparing accuracy.

In almost all the above cases, comparisons with other approaches in literature are carried out using OU-ISIR dataset and achieve a significant decrease of the EER.

Other examples of approaches are those reported in Section 4.5.1 for smartwatches.

4.6 Less investigated topics

This Section presents some interesting works that address very specific processing steps, not necessarily related to the recognition process, so that it is worth describing them separately.

4.6.1 Optimizations for Gait Recognition Systems

State-of-the-art optimization techniques for gait recognition via wearable sensors presented here do not only regard the computational demand, but also the problems that it can rise or worsen. As for any other kind of mobile recognition system, the problem of the energy consumption is not negligible and must be taken into consideration when designing a real system.

The work in [98] investigates the correlation between the recognition results and either the position in the walk sequence, or the number of the cycles used for the comparisons. The two parameters are also considered together. The authors exploit the same techniques in [80] (see Section 4.4.3) for preprocessing (plus noise reduction by Low Pass Filter), cycle segmentation and recognition. The walk cycles are divided into three groups (Beginning, Middle, and End). After that, the system choses the k most representative cycles by DTW (those that achieve the lowest average distance with all the others). This is the most computational demanding part but it is independent from k, which is changed from time to time varying from 1 to the number of total cycles. This procedure is repeated and accuracy is computed both for the entire walk and for each of the three groups of cycles. The results show that a higher k has a positive impact on recognition, but the gain is not that significant, while the comparing computational time is directly proportional to k. As an interesting outcome, the cycles in the End group provide better results with respect to the other two groups, even though lower than those from the entire walk. This is probably due to gait stabilization, i.e., to a tendency to better regularity after some cycles. Therefore, it seems possible to reduce computational time by taking into account only the last representative cycles, with just a little accuracy decrease.

The approach introduced in [99] presents a possible strategy to reduce energy consumption during continuous gait authentication (see also Section 4.7). At present, the acquisition of accelerometer signals by any kind of mobile device requires the use of CPU activation and processing, causing a power consumption greatly impacting on the battery life. The situation taken into account entails acquiring and processing accelerometer data, even when the subject is not moving. The authors suggest to exploit a pedometer sensor, i.e., a low power consumption sensor able to detect device movement without passing through CPU activation. This allows triggering the recording phase only when needed. However, the pedometer is not a standard sensor in present mobile devices. Examples of smartphone embedding this pedometer sensor are those equipped with the already mentioned LSM6DSL iNemo inertial sensor (included into the latest Samsung and Huawei smartphones).

4.6.2 Gender Recognition by Gait

It could be interesting to investigate gender recognition by wearable sensors data. Recent literature discusses the role of demographic information in increasing the accuracy of recognition [100] and in decreasing the required computation time. Though related to face recognition, it is reasonable to assume that a similar consideration can hold for the other biometric traits too.

The disadvantage of an approach aiming at determining demographic information first is that it would require either a human operator intervention or a preliminary step of soft biometric recognition. In this latter case, if the gender estimator is not reliable enough, the performances can decrease. A possible solution to this problem would be to train different classifiers for different demographic features or combinations. Each classifier uses a separate gallery, created according to the same decomposition of the population (classes may overlap, e.g., gender and age range). An incoming probe is submitted in parallel to all classifiers, and the response with the highest score is taken as the final recognition result [101, 102]. In particular, gender recognition without further identification of the subject can be also used in a smart ambient to provide gender-specific functionalities (e.g., advertisement). Concerning gait, as it has already been mentioned in Section 4.2, the largest publicly available datasets do not provide any gender information, so that addressing this research topic first requires collecting new datasets. At the best of our knowledge, a single work [103] has tackled the problem so far. Walking data from both the accelerometer and the gyroscope are captured from 109 subjects (46 with a Samsung Galaxy S-II GT-I9100 and 63 with a Note-II N7100, with about the same number of males and females). The subjects are asked to keep the smartphone in one of the frontal trousers pockets and to walk twice at three speeds (slow, normal, and fast), for a total of 654 walking signals. The gait signals are normalized both in amplitude (by z-mean normalization) and in time (by cubic spline interpolation to re-sample at 100 Hz), and denoised (moving average). Minima in the z-axis are used to identify cycles, refining their positions by the data from x and y axes. The approach uses Histograms of Oriented Gradients (HOG) features [104], generally used in image processing. The gradients are computed using different masks and then a 6-bin histogram is computed. Classification and regression rely on a bootstrap aggregation, an ensemble method aggregating multiple predictors. In this case, the authors use an ensemble of decision trees. An aggregated model is produced by bagging, that combines the outputs from individual models using plurality voting for classification. The accuracy of this strategy for gender recognition ranges from about 70% to about 94% depending to the test scenario (e.g., single speed or cross-speed, and using accelerometer, gyroscope or both).

4.6.3 Wearable Sensors Capturing Different Kind of Physical Measures Exploited in Gait Analysis

The study in [105] proposes a myography (see Section 4.1.2) sensor to detect and extract gait cycles. Experiments consider the 4 gait phases starting with Initial Contact, Mid Stance, Pre-Swing, and Swing (see Figure 3.1). Data are acquired by a force myography ankle band (an array of 8 force-sensing resistors), and a high-speed camera is used for ground truth labeling. The 9 healthy volunteers involved in the acquisition (5 males and 4 females) are asked to walk on a tapis roulant at three different speeds (1, 1.5, and 2 km/h). After a 2 minutes' trial, followed by a rest of 2 minutes, each subject walks for 1 minute at each speed. This second part is repeated five times with 2 minutes breaks. A total of 14 features are extracted using a sliding window of 125ms with an overlap of 93ms, i.e., root-mean-square, sum of absolute values, mean absolute deviation and others. The features are normalized for comparing purposes. Linear Discriminant Analysis (LDA) is exploited for data points classification. The gait phases are used as classes. Two independent observers label the start and the stop of the phases used for the training set according to the video sources. The testing phase determines the phase for each data point. Ad-hoc rules detect suspect phase transitions (not compatible with normal walk) to discard them. The accuracy of walking phase identification is evaluated by cross-trial validation: for each subject and for each walking speed, one of the five trials is maintained in turn as testing data, while the remaining ones are used for training. The reported overall accuracy ranges from 89.5% ± 2.5% to 92.9% ± 1.2%, depending on speed, for an average accuracy of 91.3% ± 3.3%. The slowest the walk, the less accurate the phase classification. Results could be used to segment the signals from the accelerometer, but this approach requires a non-conventional sensors, and sensor synchronization. No subject recognition attempt is reported.

The method in [106] exploits gait data by pressure sensors put in a shoe's sole. The dataset used had been collected for a diagnostic gait analysis approach presented in [107], and consists of 50 pressure signal samples recorded by 10 subjects in an ambulatory. For the verification phase, a first method exploits Euclidean Distance between couples of samples. The best reported result is a FRR of 7.2% with a FAR 0.4%. A second method exploits a Gaussian Mixture Model (GMM) for each user, computed during the enrolment. Each sample is considered as a feature vector. GMM training relies on expectation maximization (EM), while, during verification, the system uses the new information to automatically update both the individual GMM model and the acceptance threshold(s). The new experiments achieve 80% positive identification rate and 86% impostors' rejection rate.

The acquisition technique in [108] relies on plantar pressure imaging (PPI), capturing the way each individual interacts with the walking ground following a common stereotyped sequence: "*heel strike, roll to the forefoot, then push-off with the distal forefoot and toes*". At normal walking speeds (about 1.2 ms^{-1}), this cycle lasts about 0.7 s. PPI systems can be included among floor-sensor based approaches. The typical equipment consists of a huge number of pressure sensors (hundreds/thousands) integrated into an array, with spatial and temporal resolutions respectively in the order of 5mm and 100Hz. These measures are rendered in the form of images, where low-pressure thresholding allows easily isolating the foot, and that can be easily aligned. PPI-based biometric identification had already been proposed, but with very small datasets (a maximum of 30 subjects in [109]). The goals of this study are to exploit a larger sample of subjects (104) to assess PPI-based gait recognition, and to compare different spatial alignment strategies, features and feature extraction procedures. Computer vision-based processing and classification are applied to PPI images; therefore it is out of the scope of this section to provide further detail on this approach. It is mentioned because a similar weight-based strategy has been recently made "wearable" by the work in [110]. The latter is worth mentioning just because of the novel strategy to gait recognition, though somehow related to the above one, that requires 8 weight sensors to be embedded directly in the shoe insole. It seems that values from the sensors are mapped onto corresponding time series and processed as such. Weight is used as a feature together with the way of walking (time series of weight distributions produced while walking). A very small dataset is used, with apparently 14 gallery and 14 probe samples, a pair per subject, for testing. No indication is given about the length and conditions of the walk. Both a Gaussian Radial Basis Function (GRBF) SVM and a Naive Bayes (NB) Classifier are tested, with an average verification accuracy ratio of 83.88% for NB and 99.60% for SVM-GRBF. The problem is represented by false acceptances. Apart from the results obtained over a small dataset, this capture modality might be interesting to consider, to solve the problem of device orientation and position. However, special equipped shoes are needed (not such unrealistic compared to a sensors band all around the body).

4.6.4 Multibiometric Systems including wearable sensor-based gait recognition

The following two works exploit the synchronization of video and inertial data for gait recognition.

The proposal in [111] presents Proprio-Extero Matching IDentification (PEM-ID). Data are acquired by a 640x480 camera at 30fps and by the SparkFun 6DoF IMU board, with a MMA7260Q accelerometer, embedded into a small computer on the person's body. A motion signature consists of landmark features extracted from camera and from accelerometer. Data from both devices is acquired from 2 subjects in 12 walks of about 10 steps (for sake of image quality), while outside this FOV, only the accelerometer captures the gait. The timestamps of the relevant events (heel-strike and midswing) are stored as a feature vector for both video and acceleration data. To this aim, device-specific strategies are exploited. For the video sequence, hell-strike and midswing events are characterized by detecting feet at their farthest or closest distance, respectively. As for the accelerometer signal, the vertical acceleration of the body's center of mass is known to reach its maximum at the midswing of the gait cycle, while heel-strike produces its minimum. In a first set of experiments, a single subject at a time walks 12 times, changing the entry and exit points, across the FOV of the camera. Two different subjects repeat this process, giving 24 walk sequences. The simulation of scenarios involving multiple subjects is possible by permuting the experimental data in different ways. The comparison step exploits a suitable distance metric. In the following experiments (8 experimental runs) three subjects cross the camera FOV, entering/exiting from different directions each time. Only one person in each scene wears the accelerometer, and is therefore identified by the PEM-ID system. The recognition procedure searches for the best comparison, globally minimizing the distance between video and accelerometer features. The average recognition rate is 87.5%.

The system in [112] combines RGBD (RGB data with depth information) and accelerometer data. The acceleration signals are collected by an Android smartphone (unspecified model) positioned in one of the front pockets. The video sequence is collected by a Kinect 2.0 placed 0.5m over the ground. The RGB and the depth stream are combined into a single VGA stream of 8 bits per pixel. The authors define a "*group of acceleration data as the sequence of acceleration values resulting from the entire walk (from one end of the hallway to the other)*"[112]. Three different datasets are collected and exploited. The first one includes data from 10 subjects each with 100 *groups of acceleration* and RGBD data, half and half with normal and fast pace, along a 18 meters hallway. The data from the accelerometer is collected, divided into steps and interpolated at 50Hz. The video data is captured asking the subjects to walk 100 times from 1m to 5m away with respect to the Kinect. Video and smartphone data are not synchronized. The second and the third datasets contain data from 50 subjects and the acquisitions of video and accelerometer are synchronized. The third dataset contains more acquisitions in different conditions (e.g., arms positions). For accelerometer-based signals, events are midswings and heel-strikes, identified as above. Separate features are extracted from accelerometer (magnitude vector) and video data. For accelerometer data, the system exploits Principal Component Analysis, in order to find the uncorrelated orthogonal basis of the gait curves exploited as training feature vectors; the system uses this basis (*EigenGaits*) to map a new gait signal onto the derived space. For the video-based part, the system uses Dense 3-D Trajectories-Based Gait Representation, and encodes and then clusters the extracted trajectories by a bag-of-words strategy, using k-means algorithm. For the recognition phase, the system concatenates the features extracted from accelerometer and from video data and uses this vector in order to train a SVM classifier (measuring signal similarity by Pearson's correlation coefficient). SVM reaches an accuracy above 90%, with peaks of 96%, in different experimental scenarios.

To the best of our knowledge, a single work so far has fused wearable sensor-based gait and voice recognition to improve their accuracy [113]. Gait recognition is possible when walking users carry mobile devices, and also recognizing a speaker is also a viable operation while talking via the mobile device or close to it. Performance of the combined approach is demonstrated to be significantly better. Experiments are carried out using voice samples and gait signals from 19 males and 12 females (31 subjects total) acquired during two different sessions. Two Analog Devices ADXL202JQ accelerometers are positioned perpendicularly to make up a three-dimensional accelerometer module, used for acquisition together with a laptop with a National Instruments DAQ 1200 card. The gait signals are recorded at 256 Hz. During each session, test subjects walk along an about 20 meters long corridor with normal, fast, and finally slow walking speed, keeping the accelerometer module in each of three varying positions: hip pocket, chest pocket and suitcase handle. The accelerometer module cannot freely move inside the pockets, but pockets are attached to mock-ups of clothes and their shifting leads to differences in accelerometer positioning during training and testing data collection. The speech database used for experiments contains five utterances per speaker, each being the pronunciation of string of eight digits. Four utterances are recorded in the same session and make up the training data. The fifth speech sample is recorded during a second session, and is used for testing. The speech samples (at 8000 Hz) are collected in a quiet environment and normalized. They are then contaminated in different ways: with white, pink, city and car noise, with three SNRs: 20, 10 and 0 dB. City and car noise are taken from NTT-AT Ambient Noise Database[5] while pink and white ones are artificially generated. Gait processing and recognition are carried out by the already discussed methods in [70, 71] (see Section 4.4.3), providing both a score determined by the correlation of gait segments and a further score computed via FFT. The speaker recognition exploits the text independent system MASV (Munich Automatic Speaker Verification) speaker verification environment[6]. Gait and voice similarity scores are fused by Weighted Sum fusion, after normalizing scores of each modality to the range [0, 1]. The weights are derived from the expected reliability of each modality, which is estimated in advance according to the achieved error rates.

[5] http://www.ntt-at.com/products_e/noise-DB/

[6] http://www.bas.uni-muenchen.de/Bas/SV/

The best performance using the separate traits is EER=14.1% for gait recognition using correlation with the accelerometer module in hip pocket, EER=13.7% using FFT with the accelerometer module in chest pocket, and EER=2.82% for speaker recognition with 20dB SNR with city noise. The best result achieved with fusion is EER=1.97% with city noise 20dB and the accelerometer module in the chest pocket.

4.7 Robustness to Presentation Attacks

Solutions for the problem of presentation attacks are a generally relevant topic of biometric research. In the case of gait patterns acquired through wearable sensors, it is interesting to know at which extent a presentation/spoofing/impersonation attack might be successful. Different attack scenarios have been analyzed that relate to this topic. The experiments in [114] include the evaluation of minimal effort mimicking for gait biometrics. The datasets used for this work and for the following ones by the same group are collected using a Motion Recording (MR) sensor. The magnitude vector is computed from the acquired signals, linearly interpolated and denoised by moving average filter. Cycle detection exploits local minima, starting from the first group of 250 samples and considering that a cycle contains about 100 samples. Cycles are normalized in time in order to have exactly 100 samples each. The median value of corresponding samples in walk cycles represents an average cycle, which is computed for each walk. Euclidean distance between average cycles provides the comparison score. In the kind of attack analyzed in this work impostors have only a limited number of attempts, and can spend a short time to study the target. It is not a zero-effort attack, but it entails very low engagement. The results are quite optimistic about the robustness of gait against this kind of attacks, but experiments involve a limited number of samples.

The experiments in [115] are more specifically focused on presentation attacks, and exploit a larger dataset of 100 subjects. The paper investigates two types of attack scenarios. The first one is similar to the preceding work, dealing with a minimal effort impersonation. In a more complex scenario the attackers know who is the data subject in the dataset with the most similar gait. The results confirm that "*a minimal effort impersonation attack on gait does not significantly increase the chances of impostors being accepted*", but they also demonstrate that "*an attacker with knowledge of the closest match in the database can be a serious threat to the gait authentication system*"[115].

The study in [116] demonstrates that knowing the gender of the impersonated data subject helps attacks. One point in the final discussion in [115] is whether attackers' chances to succeed can be improved by training.

The conclusions reached by different authors are not univocal. More recent experiments in [99] take also some amount of attacker training into account. The work exploits energy consumption reduction (Section 4.6.1), since the experiments hypothesize continuous verification of the identity of the owner of the mobile device. Preprocessing applies mean subtraction for accelerometer signal denoising and then the system computes the magnitude vector. This undergoes linear interpolation and filtering by Savitzhy-Golay algorithm. The cycle length l is estimated and cycles are extracted exploiting local minima at approximately l samples distance. Outlier cycles are computed by DTW and removed. Recognition is based on majority voting over DTW pairwise comparisons of cycles. A user is recognized if at least 50% of such comparisons passes the (fixed) voting threshold. Two datasets are used for different experiments: zero-effort, minimal effort (reenact), and coincide. In both cases data are collected by one Sony Xperia Z3 and two Sony Xperia Z5-Compact smartphones. The first dataset collects data from 35 subjects along a 68 meters long hallway. It is used in all-against-all comparison, that simulates in some way zero-effort attacks (simply a false identity is claimed). The system achieves an EER of 13%. The second set of experiments is specifically designed to simulate impersonation attacks. It involves a group of 5 attackers, that are all actors trained and specialized to mimic body language and motions. The 4 victims are normal users that have similar physical characteristics to those of at least one of the attackers (one of the victims is suitable for two attackers). During the enrolment, a second dataset is collected asking the victims to walk twice at normal speed along a 16 meters long hallway (3-4 minutes), under the respective attacker's observation (especially for walking dynamics and phone location). Afterwards, the attackers are allowed to further observe the respective victims also noting the body movements and other characteristics useful for imitation. Then, an attempt to be recognized as the victim is carried out. No attacker reaches the requested amount of votes to be accepted. The last scenario (coincide) is specifically designed to analyze whether attackers can be helped to learn and reproduce a victim's gait by the "*synchronization of steps and of other body movements*". An attacker and its victim walk side by side at victim's normal pace. The rationale for this is that it seems that "*walking side by side normally tend to adapt each other's walking pace*", which could make the attack more effective. A Bluetooth channel connects the smartphones of the victim and of the attacker, while the gait recognition application runs in verification mode on both for 15 minutes. Probe feature vectors are compared in real time: the signal captured by by the victim's smartphone is transferred to the attacker's device, where it is compared against the attacker's template. Notwithstanding the more favorable condition, no impostor reaches the needed confidence score to be accepted. Even if few subjects are involved, considering that attackers are well trained in impersonating behaviors, this results seem to prove the robustness of gait to impersonation attacks. Another interesting observation is that attackers' imitation capability generally decreases along the walk.

At the best of our knowledge, a single work has investigated spoofing of a silhouette-based gait biometrics systems [117]. Experiments testify that it is possible, though not as easily as for face. A photograph of the target person can allow to easily achieve face spoofing, whereas gait may be more demanding for the attacker, and require, e.g., to select a victim with a similar bodily structure, imitate the clothing and mimicking the walking style. Given the possibility to combine video and wearable sensors data (see Section 4.6.4 above), it would be interesting to measure the possible increase of robustness provided by simply combining gait recognition by the two different modalities.

Chapter 5

The Investigated Approaches To Wearable Sensor-based Gait Recognition

This Chapter describes in details the contributions of this PhD research. As mentioned in Chapter 1, the first goal is to improve the state-of-the art of gait recognition by wearable sensors. The additional goal is to design a prototype and to develop a complete and automatic system for gait recognition. Section 5.1 presents some introductory concepts. The proposals presented in Section 5.2 deal with this project. Section 5.3 presents further attempts related to different strategies for preprocessing, step segmentation and gait recognition methods. Section 5.4 presents a novel normalization procedure for data coming from different devices. It aims at producing more homogeneous gait signal by attenuating the calibration errors of involved accelerometers. Finally, Section 5.5 presents other published works in this field not exclusively related to the main topic of the thesis.

5.1 Introductory Concepts

Dataset Distance Matrix. When it is not mentioned otherwise, the experimental set up does not involve any training, and therefore there is no subdivision of the exploited dataset into training, testing, and/or validation sets. Instead of dividing in different ways the biometric templates into probe and gallery, we exploit an all vs. all distance matrix. Such a matrix is used from time to time in a way which is suited to the different tested biometric applications, as it will be detailed below. In this way, it is possible to obtain results that conceptually aggregate the statistics achieved by different subdivisions. The results are then averaged. Using this procedure it is possible to present performance statistics that are independent from the chosen subdivision(s) of the dataset into gallery and probe set. In the following, we will assume a number N of subjects, each with a number of S templates in the dataset (for simplicity we assume the same S for all subjects to maintain a fair comparison), for a total size of $G = S \times N$ templates in the dataset. Therefore, the distance matrix is a square $G \times G$ one. Since the experiments are carried out on a static dataset, we do not have either Failure To Acquire (FTA) or Failure to Enrol

(FTE), and therefore, according to [7], the FAR=FMR and the FRR=FNMR.

Experimental Set Up for Different Applications. As it can be seen in Chapter 4, the majority of the works in this field exploit gait recognition in verification applications. This is a common procedure, for example, in order to verify if the owner of the device is the one that is carrying it. In the experiments presented in this dissertation, when it is not differently specified, the tests consider two possibilities for verification.

In the first case, **Verification with a Single Template (VER)**, each walk template is taken in turn as a probe. For each probe, the full set of possible claims of identity are assumed in turn. Each test compares the probe with a single gallery (enrolled) template for the claimed identity. Different tests use in turn different gallery templates for each enrolled identity, except of course the one used as probe. This means that each probe is considered once as genuine and the remaining times (i.e., the number of subjects minus one) as impostor. The errors are then averaged over all the possible tests. In more detail, each row represents a set of $G-1$ operations, namely $S-1$ genuine attempts and $(N-1)\times S$ impostor attempts. This provides a total of $G\times(S-1)$ genuine attempts and a total of $G\times(N-1)\times S$ impostor attempts. Of course, for each comparison (matrix cell), according to the value in the cell, the distance threshold being considered, and probe label and the cell label either an error or a success is recorded.

In the second case, **Verification with Multiple Templates (VER_MULTI)**, a single sample is taken in turn as a probe, but all the templates in the gallery belonging to the same subject are used in group and only the best comparison is taken into account for the acceptance/rejection choice. Apart for this, the protocol is the same of VER, with of course a different total number of tests. In more detail, in this case, each row represents a set of N operations with 1 genuine attempts and $N-1$ impostor attempts. As a consequence, there is a total number of G genuine attempts and $G\times N-1$ impostor attempts.

The results for both the verification applications are presented in terms of EER.

In this thesis, also identification applications have been taken into account, including both Closed Set and Open Set. In these cases, if the application pertain to a real world scenario, the acceptance decision must be taken by an external service, that can be either a single server or a cloud service. This is due to the fact that for these kinds of applications it is not possible, for obvious security issues, to maintain the entire gallery of subjects in all the subject smartphones. An identification application can be useful, for example, in order to trigger the opening of a "secure" door only for authorized subjects. In both open and closet set applications, only multiple template tests have been always considered.

For the **Closed Set Identification (CSI)** applications, each walk template is taken in turn as a probe. It is compared against all the other walk templates in the dataset, in order to increase variability and possible errors. The results are reported in terms of Recognition Rate (RR). In some cases also the results achieved at different rank (generally 5 and 10) are reported. Each row represents a single operation and contains a single genuine attempt. Therefore there is a total of G genuine attempts and no impostor attempts. Moreover, no distance threshold is exploited.

For the **Open Set Identification (OSI)**, a similar protocol to CSI is used. Of course, in this application is also taken into account a reject option, according to an acceptance threshold. For this reason, each probe is assumed in turn either as genuine (the templates that belong to the same identity are present in the gallery) or impostor (such templates are not included in the gallery). The results for this application are presented in terms of EER. It is worth noticing that such EER is taken using a returned list of length *k=1* for the *TPIR(t)*, that is therefore computed as *DIR(1,t)*. This can be considered a strict constraint, especially for a soft biometrics. Each row represents two identification operations, one considering the subject as genuine and the other as impostor. In this case, we have the same number of G genuine and impostor attempts.

General Data Protection Regulation (GDPR) Related Issues. The new GDPR has a relevant impact on biometric data acquisition, management and processing. In general, for gait recognition by wearable, the user is always concious that the device can capture data for authentication/recognition purposes. In fact, the subject must install the app in order to collect data and authorize the acquisition. This differs from floor sensor-based and machine vision-based gait recognition, where the data are captured and processed by a third party, possibly without subjects awareness/consensus. Considering the verification application, the data are acquired directly from the device and processed on it. To ensure protection, the stored gallery templates can be also secured by encryption. As already mentioned before, for the identification applications, the data must be processed by an external service. Even in this case, the subject is aware of capture (unless a malicious application has been injected in the mobile device). However, in this scenario, in order to protect data on the way towards and inside a remote data center, it is possible to encrypt the acquired gait signal and to send it to the external service by using a secure protocol such as *https*. Of course, this means that also the external service must guarantee the security of biometric templates. More details can be found in Section 5.5.

Beacon Technology. The beacons are Bluetooth 4.0 emitting sources. They have the only aim to transmit their ID to a listening (mobile) application. In general, they can have different kinds of usage in different fields. An example is a smart-museum: in this case, they can be used to allow the museum app to show the description of the artwork close to/seen by the mobile device owner, changing it to the right one when the user moves to another location. In some of our experiments, beacons are exploited in pairs in order to trigger the start and the stop of the gait signal acquisition (see Section 5.2.4 for more details).

5.2 Biometric Walk Recognizer System

The work described in this Section has been peer-reviewed and has been published as conference paper in [75] and it has been presented at the 18th International Conference on Image Analysis and Processing - *ICIAP2015*. It reports the adopted strategies and the results achieved in the first project phases. It is worth describing this work, even if published before the start of this PhD, because it represents the basis of the present BWR approach.

The proposal in [75] exploits data captured in a hallway by using a One Plus

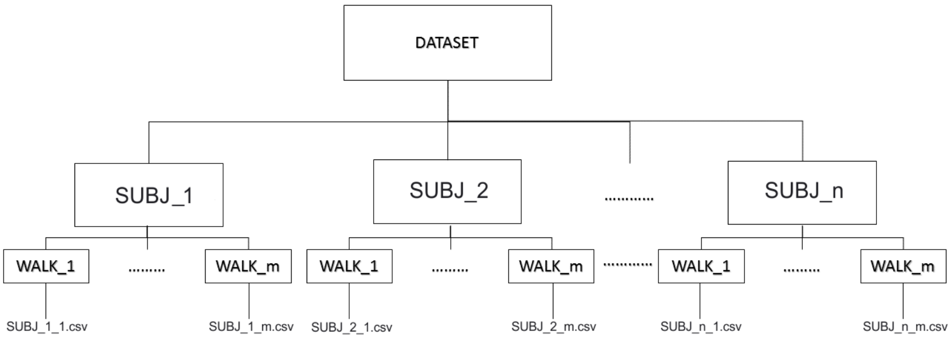


Figure 5.1. Data structure for the dataset used in the experiments.

One smartphone attached in the left or right hip zone (arbitrarily chosen by the subject) by a belt. Gait signals are acquired by the Physics Toolbox Accelerometer[1] application available in the Play Store. Each out of the 26 subjects involved in the acquisitions is asked to walk for 10 steps in a normal way 3 times (all data for a subject is collected in the same session). The dataset structure (see Figure 5.1) is studied in order to be easily loaded/updated.

Differently from the majority of literature works, no preprocessing algorithm is applied to the original signals, except for a "purification" of the starting and ending parts of the signal. These portions are mostly flat and therefore not significant, and they are consequence of the manual user operation required for triggering the acquisition. Given the smartphone orientation, the dominant axis is the Y one. Therefore the purification is performed on this axis by discarding the signal samples until the first relative maximum found after the first value greater than 1.05. This threshold is empirically chosen as the best option. A symmetrical operation is carried out on the signal tail. Figure 5.2 shows a graphical example of the effects of this purification procedure.

As for recognition, the proposal compares 5 different methods. In particular, 4 out of these 5 strategies (see below) require a preliminary step segmentation procedure. To this aim, a novel step segmentation procedure is presented. It is

[1] https://play.google.com/store/apps/details?id=com.chrystianvieyra.android.physicstoolboxaccelerometer

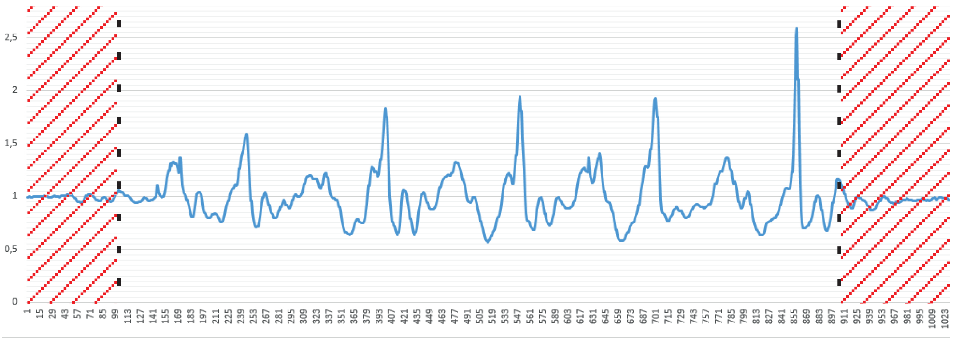


Figure 5.2. Example of extraction of the relevant segment from a signal

based on the previously knowledge of the number of steps composing the walk. After the purification, such procedure looks for all the maxima in all the gallery gait signals for the user at hand. These maxima are sorted in decreasing order and the $(10 \times \#gallery\_walks\_per\_user) - th$ maximum is taken as the step segmentation threshold parameter. Finally, the signal is divided exploiting this threshold. The procedure looks for the relative maximum in the walk at hand that follows a value greater or equal to such threshold. It is worth reminding that the number of steps requested to the users during the acquisition phase is 10. Since this strategy provides about 10 steps per walk, it is possible to conclude that it works quite correctly. Actually, in the following of the thesis it will be shown that the knowledge of the exact number of steps has been proven not to be a strict requirement. Therefore, the evolutions of this first proposal have been successfully applied to public datasets where the number of steps is unknown. In order to increase the accuracy, the next phase entails an outlier steps detection and removal strategy. As already mentioned in 4.4.3, an outlier is defined as a step or cycle which is very different from the other ones. In this case, we use DTW to compute pairwise differences between steps/cycles of the same walk and we discard those showing a distance from the others greater than the average distance plus the standard deviation. It is worth highlighting that, in this first work, operation parameters (such as the step segmentation threshold and the values to identify outlier steps in a single walk) are computed considering all the gallery templates of the same user as a whole. In particular, the step segmentation threshold is stored in the gallery for each user as a distinctive feature. This is different with respect to the following work. In fact, in the evolutions of the proposed segmentation strategy the analysis is carried out separately for each to the single walk (see below 5.2.1 for details).

The 5 recognition strategies are: WALK, BEST STEP (BS), BEST STEP VS. ALL (BSvsA), ALL STEPS VS ALL (ASvsA), and STEP SLIDING WINDOW (SSW). They are all based on the basic formulation of Dynamic Time Warping algorithm. They are presented once and for all here because they are exploited in the following works too.

WALK: It represents the simplest method. After the purification, it exploits the DTW algorithm to compare two entire walking signals. Clearly, this method does not need to carry out step segmentation. Figure 5.3 shows a graphical representation of the method. The limitation of this algorithm is a strong constraint: the probe and the corresponding template(s) to compare must have a sufficiently similar number of steps, otherwise the performances are very low. This problem, however, is not so relevant if the system acquires data always in the same points, for example between two beacons (as described in [118] - see below in Section 5.2.4). As a matter of fact, if beacons are suitably located, it is probable that the number of steps performed by users will be fairly equal to that decided for enrolling.

For all the recognition methods in the following that entail segmentation, this first version of the algorithms carries out a different segmentation of the probe signal for each comparison operation, by using each time the threshold stored for the current gallery user to compare. The underlying assumption is that the correct user will provide the best segmentation parameter.

BEST STEP (BS): This method represents the first attempt to avoid the constraint of a fixed number of steps and exploits step segmentation. The idea behind

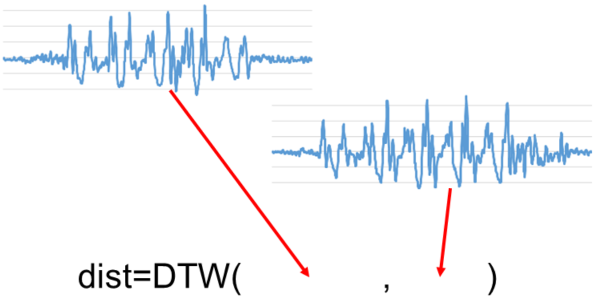


Figure 5.3. Graphic example of WALK comparing strategy.

this procedure is to compute the "best step" for each subject, defined as the centroid of the cluster composed by its steps. More in details, we consider as centroid the step that has the minimum average distance, computed in terms of DTW, with all the other ones contained in all the walks from the same data subject. Computing this best step requires no additional computational costs: when the system looks for the outlier steps it obtains the centroid step for free. In recognition phase, the distance is computed by applying the DTW algorithm between the two best steps (from the probe and the gallery walks of the user to compare). Figure 5.4 shows a graphic representation of the method.

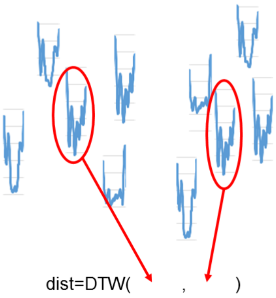


Figure 5.4. Graphic example of BEST STEP comparing strategy.

BEST STEP VS. ALL (BSvsA): This method uses the same concept of BS. The difference with BS is in the recognition phase: this procedure requires computing the average DTW distance between the best step of the user to compare and all steps in the probe. The best result is returned as the final distance. Figure 5.5 shows a graphical overview of the method.

ALL STEPS VS. ALL (ASvsA): A different kind of approach has been tried with this method. It exploits the purification and the step segmentation as the previous methods. In the comparing phase, for each step extracted from the probe, it computes the DTW distance with each step in the gallery for the data subject to compare and it takes the minimum. Then it computes the average of such minimum distances, and returns it as the final distance score between the probe and the gallery walk. Figure 5.6 shows a graphical example of this method.

STEP SLIDING WINDOW (SSW): it is similar to WALK. It requires step segmentation, but only to mark starting and ending points for each step. A comparison between two walks entails taking the longer sequence (in terms of number of

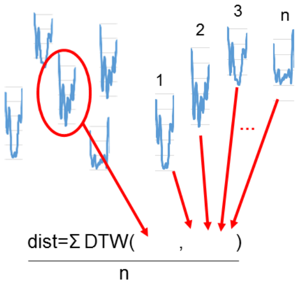


Figure 5.5. Graphic example of BEST STEP VS All method. In the figure, *n* represents the number of steps extracted from the probe.

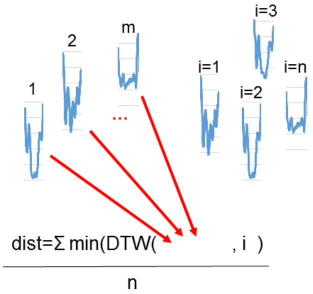


Figure 5.6. Graphic example of ALL STEPS VS ALL method. In the figure, *m* and *n* are the number of steps extracted from the gallery templates and from the probe, respectively.

steps) as the main stream, while the other one is used as a sliding window. Such window is aligned exploiting as starting points the start/ending points of the steps composing the walk chosen as the main stream. The comparison is performed over the overlapping region of the two signals. The returned distance is the minimum of all DTW comparisons performed while sliding. This method, differently from the other ones that use step segmentation, does not entail discarding the outliers since, as described above, segmentation results are only used for alignment purposes. We argue that a phenomenon similar to co-articulation in speech can hold for gait signals too. For this reason, to take steps as independent sources for comparison might cause to lack some characteristic features related to the transition from one step to the following one. SSW is designed according to this consideration. It does not suffer for the problem of a constrained number of steps, which is the limit of WALK, while retaining full information about step co-articulation, differently from the methods entailing step segmentation. Figure 5.7 shows a graphical example of SSW method. It is worth remarking that all these methods entail 3 separate comparisons, one for each axis. After the three comparisons, the results are aggregated using a weighted sum, in order to assign higher influence to the more relevant axis (in this case, y).

The proposed method has been tested both in Closed Set Identification and in Verification scenario and the results are summarized in Table 5.1. The best performing method is WALK, achieving a RR=95% and an EER=7.69%. The other methods achieve RR=38.5% and EER=30.46% (BS), RR=23.1% and EER=30.29% (BSvsA), RR=88% and EER=10.46% (ASvsA), RR=84.6% and EER=15.38% (SSW).

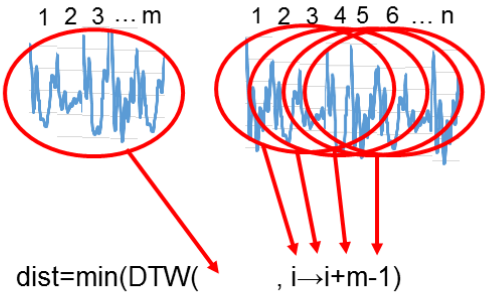


Figure 5.7. Graphic example of STEPS SLIDING WINDOW method. In the figure, *m* and *n* are the number of steps extracted from the gallery templates and from the probe, respectively.

	RR	EER
WALK	95%	7.69%
BS	38.5%	30.46%
BSvsA	23.1%	30.29%
ASvsA	88%	10.46%
SSW	84.6%	15.38%

Table 5.1. Results achieved with the first version of the system.

These results, especially those achieved by WALK, ASvsA, and SSW, encouraged our further investigations in this field.

5.2.1 The Evolution of the Biometric Walk Recognizer Approach

An evolution of the BWR approach [76] is proposed as a journal extension of the previous one, published on *Multimedia Tools and Applications*. The in-house test dataset is extended to 30 subjects. The general lines for acquisition and recognition as well as the system architecture do not change drastically but some improved techniques are presented in order to both increase the accuracy and the flexibility. The major improvements regard the step segmentation procedure and the recognition methods. The main reason behind the proposed changes is the choice of using external publicly available datasets as benchmarks for testing. The walking signals in such datasets are acquired with different kinds of devices and the use of the same (fixed) threshold cannot work properly and a redesign of some algorithmic steps is required.

The first and more relevant change is the choice to consider each single walk in the gallery as a totally independent template. This differs from the previous work where the extracted steps are used together as global features related to the enrolled data subject. This redefines the concept of a comparison: it now (and for all the following works) represents a comparison of single walks and no global characteristics for the data subject are extracted from the single walks. The reason is the execution of possibly different acquisition sessions. In fact, data slightly changes from one session to another. In this case, taking for example a single step as the only feature (as for BS and BSvsA) can possibly decrease the performance. As a matter of fact, the experimental results show an increases in term of performances processing the templates separately. It is worth remarking that, for each data subject in the gallery, the new comparison phase produces a number of distance results equals to the number of walks used as gallery templates. When more templates are stored in the gallery, this allows a better possibility for the user to be recognized.

The other relevant change regards the step segmentation algorithm. In this case, the goal is to allow it to adapt automatically to different datasets, increasing its flexibility. Experimental results demonstrate that this new version provides an improvement of the performances with both external datasets and with our in-house one. In fact, even when using the latter one, this adaptation produces a more accurate segmentation and a consequent increase in terms of performances. As for the case of recognition methods, the segmentation is now performed considering the single gait signals separately, i.e., without extracting any user-specific parameter, after carrying out the usual "purification" procedure. The new algorithm is divided into four phases. They are carried out only on the y axis (the dominant one) and the results are then projected onto the other axes (as for the previous version). The first and the second phases are used only for walks that make up the gallery, and only once, namely the first time the dataset is loaded. As for probes, these two phases are substituted by a different operation, that will be detailed in the following. The third phase represents the core of the step segmentation algorithm. The fourth one is the outlier removal phase, and it is used just to improve the recognition results by deleting steps very different from the others due to noise or other factors. These last two phases are carried out both for the walks in the gallery (only at enrolment time) and for the probe ones (according to the procedure defined below).

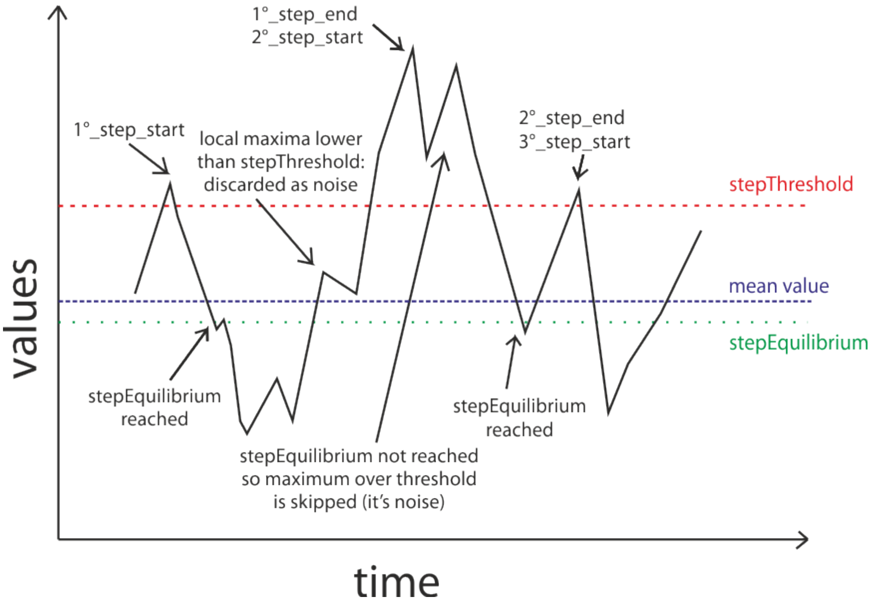


Figure 5.8. Example of Step Segmentation Algorithm.

1. compute `stepEquilibrium` as the most frequent value lower than the average of the signal values; it is used to avoid an erroneous segmentation caused by noise; `stepEquilibrium` will be used in the third phase;
2. compute `stepThreshold` as the $k-th$ highest relative maximum of the signal, where

k is the number of steps of the walk (it is worth reminding that the system knows this value for enrolled walks); similarly to `stepEquilibrium`, it will be used in the third phase;

3. divide the signal into steps using the variables described before and works as follows:
   (a) look for the first relative maximum in the signal, which will be the starting point of the first step;
   (b) scan the following walk data searching for a value that is lower than `stepEquilibrium`;
   (c) look for the next relative maximum greater than `stepThreshold`; it will be the ending point of the current step and the starting point for the next one;
   (d) repeat from (b) till to the end of the entire walk signal.
   If a local maximum is found after another without passing through a value lower than `stepEquilibrium`, this maximum will be discarded, classifying it as noise; Figure 5.8 shows a graphic representation of the third phase of the step segmentation algorithm;
4. look for possible outlier steps using the following strategy:
   (a) compute the DTW distances between the steps identified for the walk;
   (b) for each step, compute the average distance from all the others;
   (c) compute the average of average distances, denoted as μ, and the standard deviation of average distances, denoted as σ;
   (d) discard the steps whose average distance from the other steps is greater than $\mu + \sigma$;
   The procedure does not discard steps whose average distance from the others is lower than $\mu - \sigma$ because, using distances, this kind of steps are instead very good for recognition because they are in some way more uniform to all the others.

When a probe is processed using this procedure, the first two steps are replaced by a fitting procedure. In this case, the parameters for *stepEquilibrium* and *stepThreshold* are taken from the gallery template intended to compare. In this way, the step segmentation results are more accurate when the probe belongs to the same identity of the template to compare. Moreover, no previous knowledge of the number of steps making up the probe is needed because the *stepThreshold* is already computed.

The introduction of the *stepEquilibrium* allows to avoid possible errors due to noise, as it can happen with the previous version of the step segmentation algorithm. In fact, we notice that the old version produces poor results with external datasets, with an average decrease of about 20% of accuracy. Of course, this procedure does not solve the constraint of the knowledge of the number of steps. However, this limitation is now applied only to the templates making up the gallery, where a more constrained acquisition setting can be allowed. Moreover, a further study [118] (see below) shows the possibility to remove this limitation in general thanks to gait stabilization properties over time. In any case, possible solutions are mentioned to either count the actual steps or maintaining their number (almost) fixed. The first is to exploit a Kalman filter-based step counter algorithm. This is demonstrated in [119] to be quite accurate. Our experiments confirm that a good estimation of the number of steps can be provided by exploiting this strategy. A second possibility is to trigger start and stop of walking data acquisitions by using external sources, such

Dataset: BWR (in-house)

Verification Single Template (EER)

	BWR	RONG 2007	DERAWI 2010	Differences	
WALK	0.1836			NA	
BS	0.3064	0.4891	0.3430	-0.1827	-0.0366
BSvsA	0.2825	0.3100	0.3202	-0.0275	-0.0377
ASvsA	0.2019	0.5117	0.3155	-0.3098	-0.1136
SSW	0.2158	0.2754	0.1942	-0.0596	0.0216

Verification Multiple Template (EER)

	BWR	RONG 2007	DERAWI 2010	Differences	
WALK	0.1477			NA	
BS	0.3356	0.37495	0.33	-0.0394	0.0056
BSvsA	0.297	0.34685	0.369	-0.0499	-0.0720
ASvsA	0.19	0.35435	0.32	-0.1644	-0.1300
SSW	0.22	0.3373	0.2	-0.1173	0.0200

Open Set Identification (EER)

	BWR	RONG 2007	DERAWI 2010	Differences	
WALK	0.32445			NA	
BS	0.6383	0.734	0.623	-0.0957	0.0153
BSvsA	0.6702	0.7553	0.7021	-0.0851	-0.0319
ASvsA	0.4468	0.5426	0.6064	-0.0958	-0.1596
SSW	0.5426	0.7128	0.4149	-0.1702	0.1277

Closed Set Identification (RR)

	BWR	RONG 2007	DERAWI 2010	Differences	
WALK	0.8936			NA	
BS	0.4149	0.3511	0.4042	0.0638	0.0107
BSvsA	0.4362	0.4255	0.3404	0.0107	0.0958
ASvsA	0.6489	0.5638	0.4574	0.0851	0.1915
SSW	0.5851	0.383	0.7553	0.2021	-0.1702

Table 5.2. Results achieved on BWR (in-house) dataset in [76].

as Bluetooth beacons. In this way, if the path is not too long, the number of steps by different users will not be too much different.

The effectiveness of the improvements reported above is also testified by a comparison of the proposed segmentation strategy with two other state-of-the-art procedures, namely the ones presented in [63] and [59], already described in Section 4.4.1.2 and Section 4.4.1.1, respectively. As already mentioned, the experiments are not limited to the in-house dataset but are also carried out on the two largest freely available datasets, namely ZJU-gaitacc and OU-ISIR, already described in Section 4.2. Of course, only data from the accelerometers in a position compatible with a carried smartphone were used, namely on the hip. Moreover, no "purification" of the signal is carried out, since there is no knowledge of the capture triggering conditions. The system has been tested in CSI, VER, VER_MULTI and OSI. Tables 5.2, 5.3, 5.4 summarize the achieved results. The cells in green report cases in which our step segmentation algorithm provides better results than the others, while the red ones represent a worsening.

General considerations about algorithms complexity: This chapter also presents a brief discussion regarding the computational effort required in order to execute a single comparison of the proposed algorithms. WALK is pure DTW and its cost, in the basic formulation, is $O(n \times m)$, where n and m are the number of signal points (or samples) of the two time series (in the signals of our dataset there

Dataset: ZJU-gaitacc

Verification Single Template (EER)

	BWR	RONG 2007	DERAWI 2010	Differences	
WALK	0.3269			NA	
BS	0.3402	0.481	0.4783	-0.1408	-0.1381
BSvsA	0.3702	0.479	0.4778	-0.1088	-0.1076
ASvsA	0.3476	0.481	0.4773	-0.1334	-0.1297
SSW	0.3383	0.482	0.4812	-0.1437	-0.1429

Verification Multiple Template (EER)

	BWR	RONG 2007	DERAWI 2010	Differences	
WALK	0.0926			NA	
BS	0.328	0.302	0.2847	0.0260	0.0433
BSvsA	0.4104	0.335	0.3311	0.0754	0.0793
ASvsA	0.3625	0.275	0.2704	0.0875	0.0921
SSW	0.1025	0.302	0.2923	-0.1995	-0.1898

Open Set Identification (EER)

	BWR	RONG 2007	DERAWI 2010	Differences	
WALK	0.3233			NA	
BS	0.4682	0.568	0.484	-0.0998	-0.0158
BSvsA	0.5726	0.644	0.5394	-0.0714	0.0332
ASvsA	0.5397	0.467	0.4231	0.0727	0.1166
SSW	0.4162	0.579	0.5608	-0.1628	-0.1446

Closed Set (RR) Identification

	BWR	RONG 2007	DERAWI 2010	Differences	
WALK	0.9282			NA	
BS	0.8274	0.5673	0.6824	0.2601	0.1450
BSvsA	0.6668	0.5058	0.6487	0.1610	0.0181
ASvsA	0.714	0.7196	0.8102	-0.0056	-0.0962
SSW	0.7671	0.5366	0.5563	0.2305	0.2108

Table 5.3. Results achieved on ZJU-gaitacc dataset in [76].

Dataset: OU-ISIR

Verification Single Template (EER)

	BWR	RONG 2007	DERAWI 2010	Differences	
WALK	0.3661			NA	
BS	0.4405	0.4243	0.4115	0.0162	0.0290
BSvsA	0.4535	0.4333	0.4288	0.0202	0.0247
ASvsA	0.4472	0.4044	0.414	0.0428	0.0332
SSW	0.3675	0.3678	0.3625	-0.0003	0.0050

Verification Multiple Template (EER)

	BWR	RONG 2007	DERAWI 2010	Differences	
WALK	0.2723			NA	
BS	0.4116	0.379	0.3575	0.0326	0.0541
BSvsA	0.3942	0.3553	0.3382	0.0389	0.0560
ASvsA	0.396	0.3054	0.3356	0.0906	0.0604
SSW	0.2722	0.2734	0.2714	-0.0012	0.0008

Open Set Identification (EER)

	BWR	RONG 2007	DERAWI 2010	Differences	
WALK	0.7962			NA	
BS	0.821	0.9022	0.7942	-0.0812	0.0268
BSvsA	0.8372	0.9437	0.7689	-0.1065	0.0683
ASvsA	0.798	0.7958	0.7496	0.0022	0.0484
SSW	0.8003	0.8008	0.7976	-0.0005	0.0027

Closed Set (RR) Identification

	BWR	RONG 2007	DERAWI 2010	Differences	
WALK	0.2381			NA	
BS	0.2422	0.1111	0.2173	0.1311	0.0249
BSvsA	0.2386	0.1408	0.2673	0.0978	-0.0287
ASvsA	0.2750	0.2407	0.2715	0.0343	0.0035
SSW	0.2355	0.2320	0.2376	0.0036	-0.0021

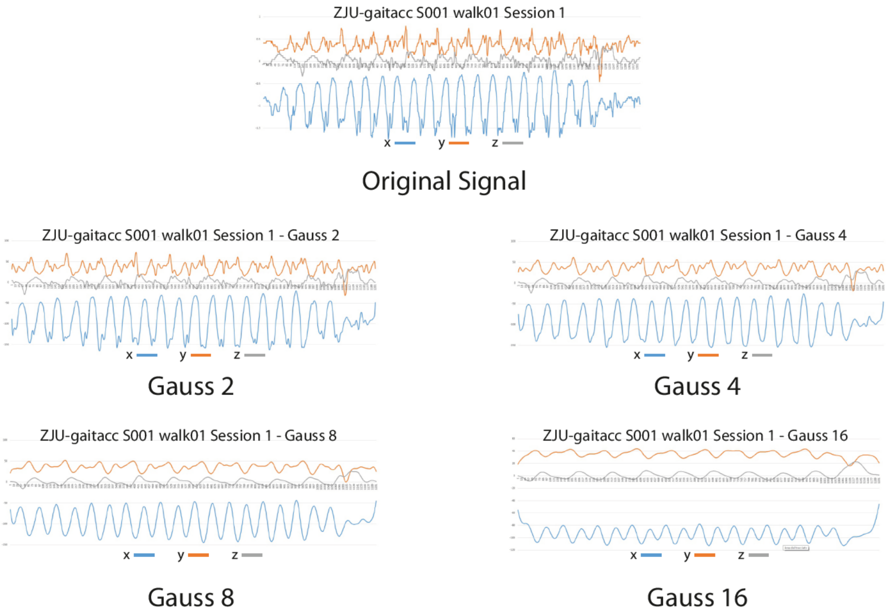
Table 5.4. Results achieved on OU-ISIR dataset in [76].

are about 1250 points). Of course, this procedure must be repeated one time per each axis. Comparison algorithms entailing the elimination of outliers as well as the identification of the centroid step from the probe template, require a main cycle that is $O(k^2p^2)$, where k is the average number of signal points in a single step (about 100 in our dataset signals) and p is the average number of steps (about 10 in our dataset). Further required computations have a generally lower complexity. However, n is about $k \times p$, therefore, up to a multiplicative constant, this preliminary step has a complexity comparable to WALK. Then the comparison for each single step is $O(k^2)$, repeated for the number of steps to compare: only one comparison for BEST STEP, p^2 comparisons on the average for ALL STEPS, taking again to $O(n^2)$. STEP SLIDING WINDOW complexity follows a different pattern. Given two walks to compare (probe and gallery) of n_1 and n_2 points respectively (corresponding to p_1 and p_2 steps), and assuming n_2 as the shortest signal, we have an overall complexity of $(p_1 - p_2 + 1) \times O(n_2^2)$. The larger the difference in signal length, the higher the number of comparisons, yet between shorter fragments. Given a signal of length $\bar{n}$ with $\bar{p}$ steps, the opposite situations are: a signal to compare with a single step of length $\bar{k}$, taking to $\bar{p} \times O(\bar{k}^2)$ comparisons, and two signals that are equals or differ for a single step, both taking $O(\bar{n}^2)$ up to a multiplicative constant. Notice that the role of the two signals in the comparisons, i.e., main stream and sliding window, depends on their relative lengths. Again, being $\bar{n}$ the length of the longest signal in the pair to compare, $O(\bar{n}^2)$ up to a multiplicative constant is an upper bound for the comparison of a single pair of signals. In conclusion, the asymptotic algorithmic complexity is the same for all methods, a part for the multiplicative constant. However, what changes from WALK to STEP SLIDING WINDOW is a progressive release of constraints, yet accompanied by a decrease in performance.

5.2.2 Study on the Benefits of Gaussian Kernel Convolution

The work in [120], presented at the *BIOSIG 2018* conference, presents the benefits of the application of the convolution of gait signals with Gaussian kernels. We decide to directly test the proposed strategy on the already mentioned ZJU-gaitacc dataset using only the hip subset. As for the previous works, it is not feasible to carry out a preprocessing step to discard the first and the last points in the signals, which are usually either noise or unstable information. The proposal includes a further alternative version of the step segmentation procedure aiming at increasing its flexibility, especially when data included in the tests possibly comes from different devices. It computes the *stepEquilibrium* parameter as $\mu - \sigma$ (computed over the signal as usual). This is experimentally demonstrated to provide slightly better results with respect to the previous version. The *stepThreshold* can be determined either by estimating the number of steps with Kalman filter-based step counter (and following the same strategy presented above) or by exploiting Bluetooth sources in order to limit the walk to have more or less the same number of steps. Moreover, as anticipated, our investigations highlight that even if the number of steps passed as parameter (k) for the segmentation procedure is not accurate, the algorithm still provides acceptable results because of the cyclic regularity of the gait signals. This is probably due to the way k is used, and to the fact that after a certain number of steps, if no exceptional event happens, the gait pattern tends to stabilize [98].

For instance, in the presented experiments k has been set to 10 for all walks in all datasets. However, the single walks in ZJU-gaitacc probably contain more than 10 steps (they are about 20 meters long), but the same value of k has been successfully used. Another modification regards the probe fitting procedure. We notice that using the *stepEquilibrium* and *stepThreshold* parameters taken from the in-analysis gallery signal in order to segment the current probe possibly raises errors. This generally only happens when the data are acquired from a different device. Of course, this problem can be reduced or avoided by applying a data normalization procedure (see below in Section 5.4), but it requires to compute the normalization parameters of the involved device(s) and this is not always possible (e.g., when an external dataset is used). The outlier removal procedure is the same exploited in the previous proposals. The proposed strategy entails the convolution of the raw gait signals with a Gaussian kernel. The experiments test 4 different values for the σ parameter (i.e., the kernel width), namely 2, 4, 8, and 16. When the step segmentation procedure is required (for ASvsA), it is applied on the raw signals as well. In fact, we notice that after convolution the signals lose the relevant peaks exploited in the segmentation algorithm. This is due to the nature of Gaussian kernel convolution that acts as a smoothing function (and it increases its effect proportionally with the growth of σ). Figure 5.9 shows an example of the effects of this convolution with different values of σ. After the convolution, we test the system exploiting WALK and ASvsA recognition algorithms.



Original Signal

Gauss 2

Gauss 4

Gauss 8

Gauss 16

Figure 5.9. An example of the effects of Gaussian kernel convolution (with different σ) on the same raw gait signal.

The experiments also test the possibility of a score-level fusion between 2 or more results. This fusion is computed from the distance values obtained by comparing the differently convolved gait data, by either picking up the best result or by summing them up. In summary:

1) the signals are possibly divided into steps;

2) different Gaussian kernels are used for convolution with the original signal;
3) distances are computed according to either WALK or ASvsA;
4) the results are (possibly) fused by taking either the best score among those returned by the kernels in the combinations (Combined BEST or C_BEST) or by summing up all these scores (Combined SUM or C_SUM).

Table 5.5 summarizes the results achieved with different Gaussian kernels or their combinations. Combinations differ for both the number of kernels involved, and for the computation of the final result. WALK, that compares the entire gait signals, confirms itself as better than ALL STEPS VS. ALL (ASvsA), that rather exploits step segmentation. C_SUM always achieves better identification results than single kernels in Closed Set Identification (CSI), independently from the chosen combination and from the recognition strategy (with or without segmentation). Identification results in CSI obtained by C_BEST are generally worse than those obtained by single kernels. In Verification (VER) mode, WALK achieves an EER from 0.334 to 0.348, depending on the kernel/combination, with the best value obtained in different settings, that include both a single kernel or a different combination. ASvsA achieves an EER from 0.354 to 0.3674, with a single best value obtained by Gaussian kernel with $\sigma = 2$. In this modality, C_SUM generally achieves worse results, while C_BEST overcomes single kernels. As expected, a significant improvement of performance is achieved by Verification Multiple Template (VER_MULTI) with respect to VER (in practice, an order of magnitude). WALK achieves an EER between 0.036 and 0.046, while ASvsA reports an EER from 0.0395 to 0.061, which reveals a higher dependence on the chosen kernel/combination. As for WALK, C_BEST and C_SUM achieve comparable results also with single kernels. On the contrary, ASvsA achieves generally worse results with single kernels, while C_BEST seems to be a little bit better than C_SUM. Finally, in Open Set Identification (OSI), which is the hardest modality, C_SUM obtains the best result both with WALK and ASvsA. In summary, it is possible to observe that C_SUM is the best option for both CSI and OSI. C_BEST seems to be to prefer for both VER and VER_MULTI. In general, combinations work better than single kernels. Table 5.5 also reports the results of compared works. The values achieved in [76] for WALK are RR=0.9282 for CSI, EER=0.3269 for VER, EER=0.0926 for VER_MULTI, and EER=0.3233 for OSI. There is therefore an improvement, except for VER. As for ASvsA, RR=0.714 for CSI, EER=0.3476 for VER, EER=0.3625 for VER_MULTI, and EER=0.5397 for OSI. In this case, the improvement is even greater and generalized. The results in [55] for the right hip are RR=0.734 (CSI) and EER=0.089 (VER). While identification results are significantly increased, improved verification is obtained only when considering a gallery with more templates per user.

5.2.3 Study on the Impact of Gait Stabilization

The following work in [121], presented at the *SITIS 2018* conference, describes a study on the impact of the gait stabilization effect (already mentioned Section 3.1) and on how to exploit it for gait recognition.

This work takes into account the interesting results presented in [98], already discussed in Section 4.6.1. For reader convenience, we remind that such work investigates both the impact of the number of cycles involved in the comparison and

Table 5.5. Results with different single Gaussian kernels or combinations. The bold values are the best result(s) for each sub-category (recognition modality - kernel(s)), the green background identifies the best result(s) for the modality. The last two rows report performance of the compared works.

	WALK					ALL STEPS VS. ALL				
Gaussian Kernel	Closed Set Identification	Verification Single	Verification Multi	Open Set Identification		Closed Set Identification	Verification Single	Verification Multi	Open Set Identification	
				ERR	DIR(1, t)				ERR	DIR(1, t)
Single Gaussian										
2	0.9286	0.343	**0.039**	0.249	0.7512	**0.8581**	**0.3540**	0.0610	0.3240	0.6840
4	**0.9641**	0.337	0.046	0.226	0.7745	0.8559	0.3577	0.0550	0.2953	0.6818
8	0.9613	**0.334**	**0.039**	**0.209**	0.7908	0.8575	0.3674	0.0485	**0.2877**	0.7407
16	0.9341	0.355	**0.039**	0.248	0.7522	0.8302	0.3669	**0.0397**	0.2918	0.6665
Combined BEST - C_BEST										
2-4	**0.9641**	**0.334**	0.046	0.226	0.7740	0.8553	**0.3567**	0.0532	0.3103	0.7129
2-8	0.9613	0.339	**0.039**	0.209	0.7908	**0.8575**	0.3587	0.0469	**0.2737**	0.7249
2-16	0.9341	0.342	**0.039**	0.248	0.7522	0.8302	0.3581	**0.0395**	0.2950	0.6954
4-8	0.9613	0.341	**0.039**	0.209	0.7908	0.8570	0.3603	0.0476	0.2811	0.7325
4-16	0.9341	**0.334**	**0.039**	0.248	0.7522	0.8308	0.3602	0.0397	0.2975	0.7069
8-16	0.9346	0.35	0.04	0.248	0.7522	0.8297	0.3630	0.0407	0.3032	0.7134
2-4-8	0.9619	0.338	**0.039**	**0.208**	0.7908	0.8570	0.3592	0.0472	0.2740	0.7249
2-4-16	0.9341	0.342	0.04	0.248	0.7522	0.8308	0.3592	0.0397	0.2950	0.6954
2-8-16	0.9346	0.343	0.04	0.248	0.7522	0.8297	0.3596	0.0401	0.2956	0.6954
4-8-16	0.9346	0.344	0.04	0.248	0.7522	0.8297	0.3612	0.0401	0.2983	0.7063
ALL	0.9346	0.343	0.04	0.248	0.7522	0.8297	0.3600	0.0401	0.2956	0.6954
Combined SUM - C_SUM										
2-4	0.9662	**0.334**	0.046	0.232	0.7669	0.8652	0.3593	0.0581	0.3092	0.7074
2-8	0.9711	0.338	0.043	0.208	0.7919	0.8843	**0.3589**	0.0496	0.2729	0.7456
2-16	**0.9728**	0.343	0.042	0.199	0.8007	**0.9001**	0.3640	0.0426	0.2535	0.7544
4-8	0.9641	0.34	0.042	0.208	0.7919	0.8723	0.3629	0.0509	0.2606	0.7183
4-16	0.9657	0.345	0.038	0.197	0.8028	0.8919	0.3622	0.0427	**0.2364**	0.7325
8-16	0.9602	0.348	**0.036**	0.2	0.8001	0.8739	0.3669	**0.0411**	0.2680	0.7484
2-4-8	0.9679	0.338	0.044	0.21	0.7898	0.8783	**0.3589**	0.0491	0.2860	0.7369
2-4-16	0.9722	0.341	0.042	0.203	0.7963	0.8930	0.3635	0.0445	0.2680	0.7636
2-8-16	0.9711	0.344	0.039	0.199	0.8045	0.8925	0.3617	0.0436	0.2489	0.7571
4-8-16	0.9673	0.345	0.039	**0.195**	0.8001	0.8843	0.3637	0.0439	0.2448	0.7369
ALL	**0.9728**	0.342	0.041	0.2	0.7996	0.8936	0.3605	0.0474	0.2615	0.7642
[76]	0.9282	0.3269	0.0926	0.3233	-	0.714	0.3476	0.3625	0.5397	-
[55]	Identification: RR=0.734					Verification: EER=0.089				

their discriminative power in relation with the temporal positions they have in the original gait acquisition. The reported results highlight that at the growing of the number of steps involved in the comparison (k), the accuracy tends to increase, even if the gain is not that high. However, the comparing cost is directly proportional to such k, requiring much less effort for lower ks. Moreover, another interesting outcome from this study is that the cycles belonging to the last part of the walk (the "End" group of their experiments) provide better performance with respect to the ones in the start and the middle of the signal, even though they present a lower accuracy with respect to the use the entire walk during the comparison. This evidence highlights the effect of the so called "gait stabilization", i.e., the tendency of a better regularity of the gait pattern after some cycles. In other words, the longer the walk, the more regular are the last cycles in the gait signal. From their study, in conclusion, it seems possible to reduce the computational effort of a comparison by selecting only the more representative cycles in the last part of the walking signal (corresponding to their End group), with just a little decrease in terms of accuracy. The most computational demanding part of the proposed strategy is the selection of the k more representative cycles because it requires an all against all cycle comparison (each one using DTW). However, this effort is independent from the k itself, since the only relevant factor is the total number of cycles composing the walk (or, in other words, its length).

In our proposal, we attempted to exploit this phenomenon to increase the accuracy of our BWR system. At the same time, a further goal is to avoid heavy computational effort, such as for the selection of the k more representative cycles. The recognition strategy follows the protocol presented in [76]. As for our previous work, no preprocessing, denoising and/or signal enhancing techniques are applied on the signals. In fact, even if the use of such preprocessing generally increases the quality of time series, we notice that the resulting signals lose some important characteristics exploited in the proposed segmentation and recognition strategies (see also Section 5.3).

Gait signals are divided into steps exploiting the last version of the step segmentation procedure presented in our previous work in [120], which experimentally demonstrates to work better with respect to our previous segmentation strategies. In order to exploit the gait stabilization effect, the step segmentation procedure only aims at identifying the start and the end points of the steps and not to actually segment the signal. These points are then exploited to extract the last and more stable part of the signal. For the comparing phase, a modified version of the recognition method WALK presented in [76] is proposed. More in details: let's assume that s_1, s_2, ..., s_n is the ordered sequence of steps making up the walk, and s_i^s and s_i^e are the start and the end indexes of the step i respectively. The proposed comparison strategy requires to compare, using the well-known DTW algorithm, the portion of the walk signal between $s_{\kappa-1}^s$ and s_{n-1}^e where κ ranges from 2 to 5 in the presented experiments. The choice of removing the very last step is due to the fact that it generally presents more noise caused by the acquisition setup. In fact, during the acquisition, the subject knows that has to stop and this somehow modifies its unconscious gait strategy. In general, considering the use of the extracted start/end points of the steps during the recognition, it is also possible to conduct the step segmentation in the reverse order, starting from the end of the last step, and to stop

Table 5.6. Results achieved at the variations of κ.

	Closed Set Identification RR	Verification Single Template EER	Verification with Multiple Templates EER	Open Set Identification EER
$\kappa = 2$	95.32%	39.25%	8.74%	26.50%
$\kappa = 3$	95.65%	38.55%	8.54%	25.90%
$\kappa = 4$	96.15%	38.25%	8.54%	25.50%
$\kappa = 5$	96.49%	37.25%	8.24%	24.50%
[76]	92.82%	32.69%	9.26%	32.33%

it once the (desired) κ-th step is found. However, the computational effort for the complete step segmentation is negligible, especially with relatively short signals as the ones in analysis.

The recognition strategy has been tested on a subset of the already mentioned ZJU-gaitacc dataset (see Section 4.2). Among the datasets controller positions, only the data from the pelvis zone is taken into account, as in [76]. Experimental tests are presented in Closed Set Identification (CSI), Verification with a Single Template (VER), Verification with MultipleTemplates (VER_MULTI), and Open Set Identification (OSI).

For CSI, each walk template belonging to the first 50 subjects is taken in turn as a probe. It is compared against all the other walk signals of the ZJU-gaitacc dataset. In other words, the gallery is made up by the templates for all of the 175 subjects, in order to increase variability and possible errors. The results are reported in terms of Recognition Rate (RR). In order to better sketch the system behavior, we also report results at Rank 5 (rate of correct recognition within the fifth position), and Rank 10 (the same for tenth position). In an analogous way, for VER, each walk template belonging to the first 50 subjects is taken in turn as a probe as well. In this case, for each probe, we assume in turn the full set of possible claimed identities, that ranges over all out of the 175 possible ones. Each test is conducted by using each single gallery template in turn for the claimed and gallery identities. In summary, each probe is considered once as genuine and 174 times as impostor, and the possible errors are averaged over the possible cases. For VER_MULTI, all templates belonging to the same user are used in group and only the best comparison is taken into account for the acceptance/rejection choice. The rest of the protocol is the same, with of course a different total number of tests. In both cases results are reported in terms of Equal Error Rate (EER). For OSI, a similar protocol to CSI is used, of course taken the reject option into account. Results for this modality are presented in terms of EER.

As already mentioned before, the proposed strategy has been tested with 4 different values for κ, namely 2, 3, 4, and 5. In this way, it is possible to sketch an overall view on how the number of steps influences the accuracy.

Table 5.6 shows that the accuracy of the system increases with κ for all recognition modalities.

For CSI, the Recognition Rate ranges from a 95.32% with $\kappa=2$ to 96.49% with

$\kappa=5$. This is an improvement with respect to the best results of our previous work, where WALK achieves a RR=92.82%.

For VER, the results are really bad, with and Equal Error Rate (EER) that ranges from 39.25% with $\kappa=2$ to 37.25% with $\kappa=5$. These results are worst with respect to our previous work, that achieves an EER=32.69%.

In the case of VER_MULTI, the results are slightly better of our previous EER=9.26%. In fact, the achieved EER ranges from 8.74% with $\kappa=2$ to 8.24% with $\kappa=5$. It is possible also to observe that in this case, the results achieved by using $\kappa=3$ and $\kappa=4$ are equal. The results from single and multiple template verification, confirm that recognizing a user relying on a single template is much more difficult with respect to the use of multiple templates.

A high improvement is reported for Open Set Identification, the most challenging test scenario. In this case, the achieved EER ranges from a 26.50% with $\kappa=2$ to 24.50% with $\kappa=5$, against the previous EER=32.33%. This is still a not a optimal result, but it is worth noticing that gait recognition is a soft biometrics and generally it is used in conjunction with other strong biometrics.

As shown in Table 5.7, independently from the value of κ, the achieved results converge on the same recognition rate considering the correct identification within the rank 5 and the rank 10, respectively. This somehow confirms the stableness of this proposals. In fact, this proves that even for the few cases in which the system gives a wrong answer, the correct identity is still returned within the first ranks. The surprising factor is that even two steps, as long as they are taken in the last positions of the walk, seem to be enough to achieve a 99% of correct match within the first 5 ranks.

Figure 5.10 presents the CMC curves until the reaching of 100% of accuracy, at the varying of κ. The two vertical lines represent the rank 5 and the rank 10, respectively. The zoomed version of the CMC curves is also presented, in order to give a detailed view of the scores until rank 10.

Table 5.7. Results achieved in Closed Set Identification at rank 1, 5, and 10.

	R=1	R=5	R=10
$\kappa = 2$	95.32%	99.00%	99.33%
$\kappa = 3$	95.65%	99.00%	99.33%
$\kappa = 4$	96.15%	99.00%	99.33%
$\kappa = 5$	96.49%	99.00%	99.33%

In summary, it is possible to sketch some interesting overall considerations. On one hand, the proposed strategy reports a trend similar to the work in [98]. In fact, the best accuracy results are always achieved with $\kappa=5$ and the performances decrease at the decreasing of κ. On the other hand, it is possible to notice that taking only the last steps produces better results with respect to using the entire signal, differently from what happens with the proposal in [98]. These discordant results are probably due to a different length of gait signals involved in the tests (20 meters against 40 meters) and/or to the different signal quality (the low quality of those acquired by Wii Remote accelerometer against the medium/high quality of

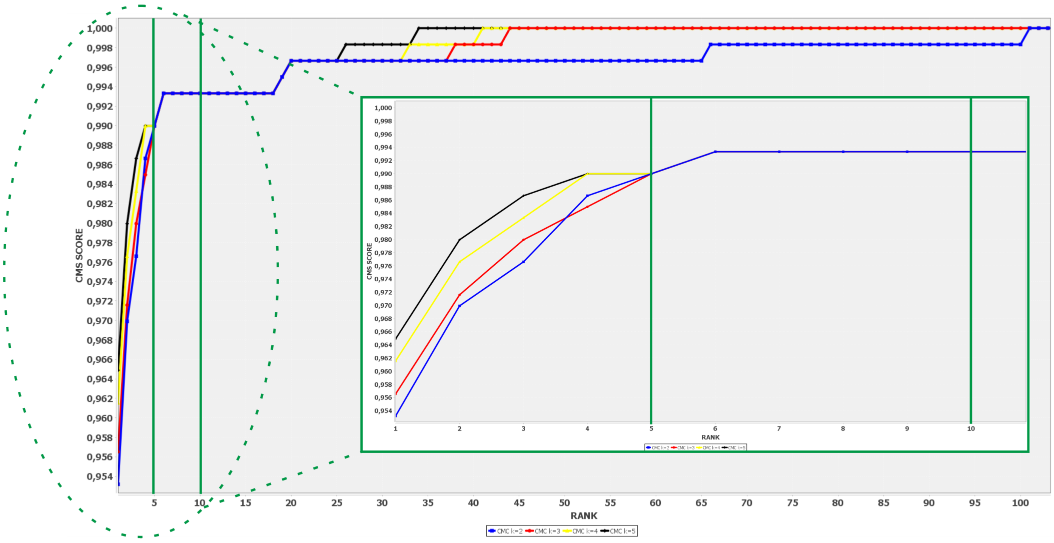


Figure 5.10. The CMC curves relative to the 4 different values of κ. The two vertical lines highlight the rank 5 and 10 respectively. A zoomed vision of CMC curves until the rank 10 is also presented.

those captured by the accelerometer embedded into the Samsung Galaxy Note 4). Moreover, the number of involved subjects is doubled in our case.

In general, the presented strategy shows some advantages. First of all, it compares a reduced portion of signal, decreasing the computational burden of the DTW algorithm. Second, it does not require to search for the more representative steps, as it happens for [98], avoiding an additional computational burden. As a third advantage, it is possible to notice an increase in the accuracy with respect to the comparing of the entire signal.

5.2.4 The Use of Beacon Technology in a Gait Recognition Scenario

The book chapter *Gait Recognition: The Wearable Solution* [118], presents a guide aiming at explaining the basic concepts of gait recognition, and the possible ways to approach it. It is also linked to the Biometric Walk Recognizer project because it firstly sketches an architecture schema for an automatic gait recognition system based on the signal acquisition triggered by Bluetooth 4.0 beacons. As already mentioned before in Section 5.1, beacons are low energy small transmitters that are able to connect to Bluetooth-enabled devices like smartphones. The use of these beacons can also be considered as a strategy for energy consumption reduction. In the presented scenario, in fact, the problem addressed is the possible continuous background execution of the biometric recognition app, since the goal is, for example, to automatically open a door at the end of a hallway only for authorized persons. The envisaged scenario is summarized as follows. The beacons configured at the beginning and at the end of the path continuously send their IDs. The designed mobile app remains idle until the Bluetooth interface receives a specific ID. After that, it generates a wake-up event. In practice, when the mobile phone is within the

beacon transmission range, if the received ID corresponds to a beacon registered as a start element, the app automatically starts the data acquisition. When the smartphone later reaches a beacon corresponding to a stop event, it stops data acquisition and transmits the data to the recognition server.
The advantages of the use of beacons are twofold. Bluetooth 4.0 has a low impact on energy consumption, and in any case it could be required by other applications too. Moreover, a beacon has a battery with an estimated lifetime of about 2 years even constantly working. Such lifetime can be further increased using a shut-down procedure for the possible day periods when the system is not required to work.
A dedicated app for gait acquisition, following the above protocol, has been developed to automatize the system. Moreover, the app includes a real time step counter algorithm and a data normalization procedure, better discussed in Section 5.4. The mobile app is divided into two parts, one allowing the user enrolment and one dedicated to data collection and transfer. The app captures data when required (a start beacon sends its ID) and then sends the collected gait data to the recognition server (after receiving an ID corresponding to a stop beacon). At the moment, no attempt to recognize the wearer of the smartphone (against the owner identity) is performed directly on the smartphone, where only verification applications could be implemented (for obvious security and privacy issues). Of course, this function can be easily included in the app, adding the biometric recognition module that usually works on the server, and is sufficiently lightweight to run on a mobile device too.

An improvement of the architectural schema sketched in [118] is presented in the already mentioned [121]. It describes a possible ubiquitous application of gait recognition by wearable sensors. The modification of the protocol entails the use of more pairs of beacons. The motivation of using more start/stop pairs of beacons is to allow capturing data between more pairs of specific points. In this way, the positions of beacons can be also used to choose locations in which data acquisition can possibly be more accurate (for example along a straight hallway). Moreover, it is possible to obtain the best compromise between signal quality, battery consumption and ubiquitousness.

Figure 5.11 shows an example of a floor with two secure doors at the end of two distinct hallways. Each start/stop pair of beacons triggers the acquisition depending on the proximity. After the acquisition is completed, the smartphone sends its ID and the acquired template to a server, and the biometric system performs the recognition task. Other examples of the beacon usage are shown in Figure 5.12.

5.2.5 Feature Based Gait Recognition

The work in [122], presented at the *ICPRAM 2018* conference, differs from the previous ones in the basic approach. It investigates the possible application of Machine Learning procedures in order to extract aggregate features from the signals, and to select the most relevant ones among them to build a biometric template. The aim is twofold: from one side, to discard less robust or less informative features, i.e., those more influenced by distortions, or that present quite flat values across signals; from the other side, achieving the goal of a lighter though accurate recognition procedure would be better suited to mobile settings.

In order to evaluate the possible influence of specific feature selection choices, 4

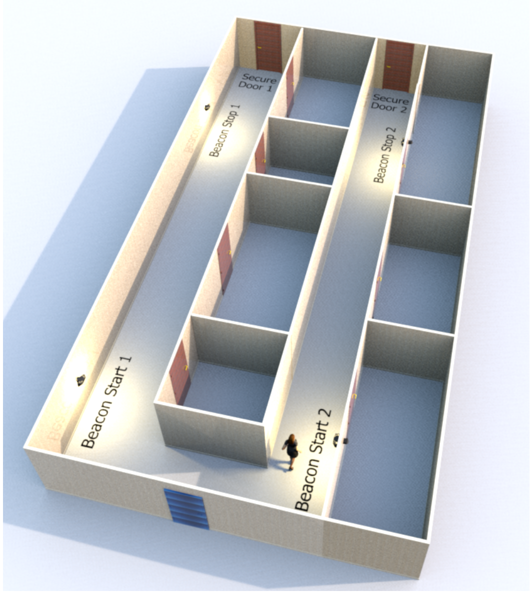
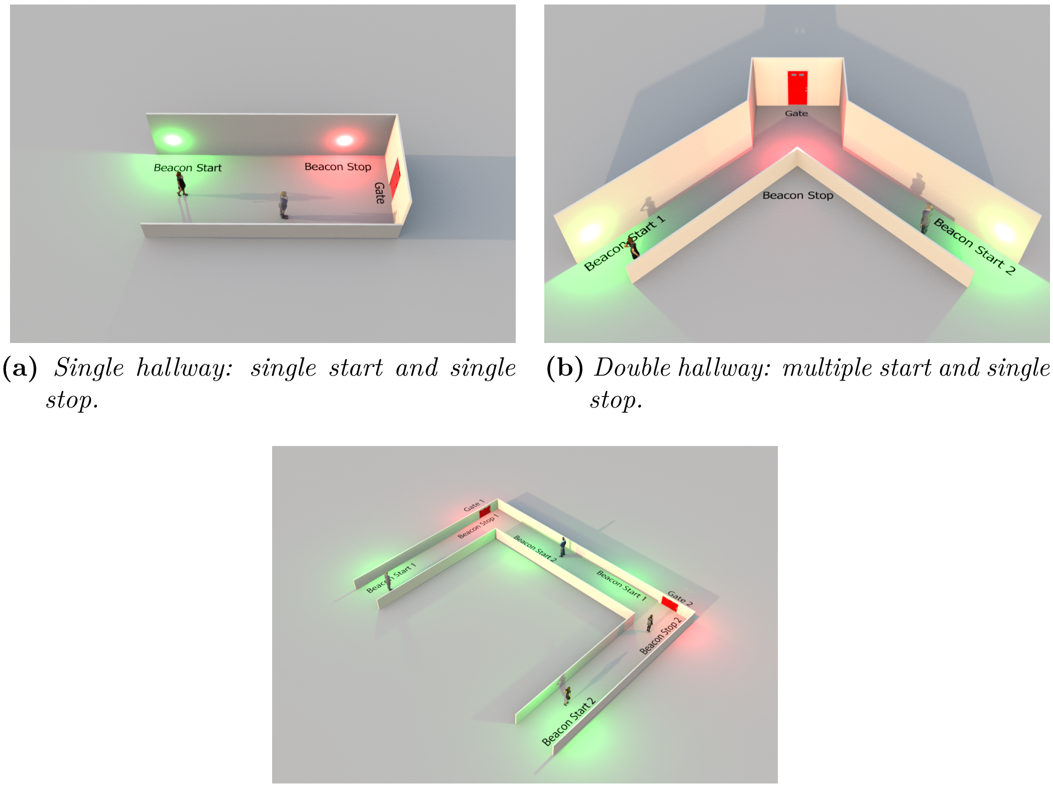


Figure 5.11. An example of a floor with two secure doors at the end for two distinct hallways. A beacon is positioned at the start and at the end of each hallway. In the example, a woman is approaching the Beacon Start 2, and this event will trigger data acquisition. Self-produced image, presented in [121].



(a) *Single hallway: single start and single stop.*

(b) *Double hallway: multiple start and single stop.*

(c) *Double gate: multiple start and stop.*

Figure 5.12. Three example scenarios. Self-produced images.

different test scenarios have been configured, with different characteristics regarding both the use of training information and the way to exploit them.

In all the test scenarios considered for the experiments presented below, the same Python libraries have been used for feature extraction and analysis. Tsfresh[2] library is used to automatically extract a large number of features from temporal series. It is usually exploited together with Pandas[3] for data analysis, and with Scikit-learn[4] library for Machine Learning. The extracted features can be later exploited to create regression or classification models, and to cluster or to compare time series.

Tsfresh includes library functions to extract a huge number of features (222) from a time series. Of course not all of them were taken into account for our experiments. Some examples follow, but it is not possible to provide the complete list of them.

- **abs_energy**: returns the absolute energy of the time series: $E = \sum_{i=1}^{n} x_i^2$ with n=number of points in the time series
- **absolute_sum_of_change**: returns the sum of absolute values of subsequent variations in the series: $E = \sum_{i=1}^{n} |x_{i+1} - x_i|$
- **approximate_entropy**: returns the approximate entropy of the signal
- **ar_coefficient**: returns the coefficient of the Auto Regressive (AR) process for a given configuration passed as parameter
- **augmented_dickey_fuller**: returns the result of Dickey-Fuller test
- **autocorrelation**: returns autocorrelation given a certain lag
- **count_above_mean**: returns the number of values in the time series higher than its mean
- **count_below_mean**: returns the number of values in the time series lower than its mean
- **cwt_coefficients**: computes the wavelet transform using this formula $\frac{2}{\sqrt{3a}\pi^{\frac{1}{4}}}(1-\frac{x^2}{a^2})exp(-\frac{x^2}{2a^2}))$
- **fft_coefficient**: computes the Fourier coefficients applying Fourier Tranform
- **mean**: returns the mean of the signal
- **mean_abs_change**: returns the mean of absolute values of consecutive changes in the time series $\sum_{i=1}^{n-1} |x_{i+1} - x_i|$
- **standard_deviation**: returns standard deviation
- **variance**: returns the variance
- **median**: returns the median value

[2] https://tsfresh.readthedocs.io/en/latest/index.html

[3] http://pandas.pydata.org/index.html

[4] http://scikit-learn.org/stable/

- **skewness**: returns the skewness (computed with the Fisher-Person standardized coefficient)

- **kurtosis**: returns the kurtosis (computed with the Fisher-Person standardized coefficient)

It is worth noticing that features are separately extracted from the signals produced on the three accelerometer axes, and the difference between test scenarios also regards the way to take their possible correlation into account. Test Scenario 1 does not use any training phase, while the others do. Considering the different domains and scale values of extracted features, a standardization procedure is exploited to build homogeneous vectors, using the well-known Gaussian normalization formula. For each feature, the average μ and the standard deviation σ are computed over gallery templates and then, for each value x, the resulting standardized value z is obtained by the formula:

$$z = \frac{(x - \mu)}{\sigma} \tag{5.1}$$

The μ and σ values are then stored, in order to normalize the further incoming probes used for testing with the same gallery. Each of the galleries that are used in turn for the experiments contains a number of templates (more than 450) that allows considering these parameters stable enough to avoid recomputing them for each probe. All test scenarios entail recognition in Verification with Multiple Template modality: each subject has more than one template, each of them is compared against the incoming probe, and the best comparison among the gallery and the probe is returned as the verification result. A probe set vs. gallery set distance matrix is produced in order to evaluate the performances. For each scenario, distances are computed between pairs of vectors built according to the scenario setting. Experiments are carried out using both Manhattan and Euclidean Distance as alternative metrics. In the following, the 4 test scenarios are described in more details.

Test Scenario 1 (T1) is used as baseline for the benchmark. In fact, it entails the use of template vectors made up by 666 features, i.e., all the features extracted by Tsfresh (222 per axis). As mentioned before, no training phase is required for this method. In this case, in order to get a fair benchmark, the templates that are used in the training set for the other test scenarios are not used during testing.

Test Scenario 2 (T2) aims at selecting and keeping only the most relevant features. For each axis, only the features that have a probability of at least 80% of changing across vectors are taken into account. In other words, only features presenting the highest variance are maintained. This analysis is carried out by Scikit-learn library. The next selection step entails a further pruning, that discards features that do not present this property for all axes, i.e., those that are informative enough but only for a subset of axes. This provides a total of 55 features per axis, summing up to 165. This selection is carried out in the training phase, so that in testing only the identified features are taken into account for both gallery and probe template sets.

Test Scenario 3 (T3) uses the same variance-based pruning exploited in T2. As a second step of feature selection, the complement of the features identified in T2 is maintained. After discarding the ones that show a too high homogeneity of values across the training set, only the features that are relevant for a strict subset of

Table 5.8. Results in term of EER for the 4 test scenarios.

TEST	EUCLIDEAN DISTANCE	MANHATTAN DISTANCE
T1	24.6%	22.4%
T2	22.5%	20.2%
T3	31.5%	30.6%
T4	19.6%	18.7%

axes (1 or 2) are maintained. This set contains 24 features, 9 from the x axis, 10 from the y axis, and 5 from z axis. Even in this case, the selection is carried out during training, and the features identified are then extracted from gallery and probe samples in the testing step.

Test Scenario 4 (T4) uses a totally different approach for feature selection. The choice of features to be kept is based on the Principal Feature Analysis (PFA)[123]. It uses the same principles of the well-known Principal Component Analysis (PCA)[124], also exploited, e.g., in face recognition for feature space reduction. The same PCA criteria are applied to select a subset of dimension q of the most representative features from the complete original set. During training, we test different dimensions, and the best results are obtained with q=60 and q=62.

The dataset exploited for the experiments of this work is the ZJU-gaitacc. The dataset has been divided into training and testing sets. The training set contains the first 3 walks of each session, while the testing set contains the remaining 3 ones. As for probe and gallery partitions of the testing set, in order to get more results, the walks from the first and from the second session have been used in turn as either probe or gallery. The obtained results have been averaged to produce the final performance measures.

Performances are reported in terms of Equal Error Rate (EER). T1 achieves EER=24.6% with Euclidean Distance (ED) and a slightly better EER= 22.6% with Manhattan Distance (MD). T2 shows an improvement, achieving EER=22.5% with ED and EER=20.2% with MD. This seems to demonstrate that the selection of the features that provide the highest information for all axes improves recognition performance. On the contrary, T3 achieves worse performance than T1, namely EER=31.5% with ED and EER=30.6% with MD. This is probably due to the too low number of features (24) and possibly to the uneven distribution across axes. T4 achieves the best results, represented by EER=19.6% with ED and EER=18.7% with MD. Overall, the best results are always obtained by MD, independently from the test scenario. The performance of PFA demonstrates that reduction techniques exploiting data correlation are effective with this kind of temporal series, obtaining an improvement of about 7% over T2. This is not dramatically significant, but encourages continuing investigating along this line. Table 5.8 summarizes the obtained results. Figure 5.13 shows FAR and FRR curves for all scenarios with both the exploited metrics.

It is interesting to make a comparison with results reported in our previous work [76], obtained on the same dataset using pure DTW for the same Verification

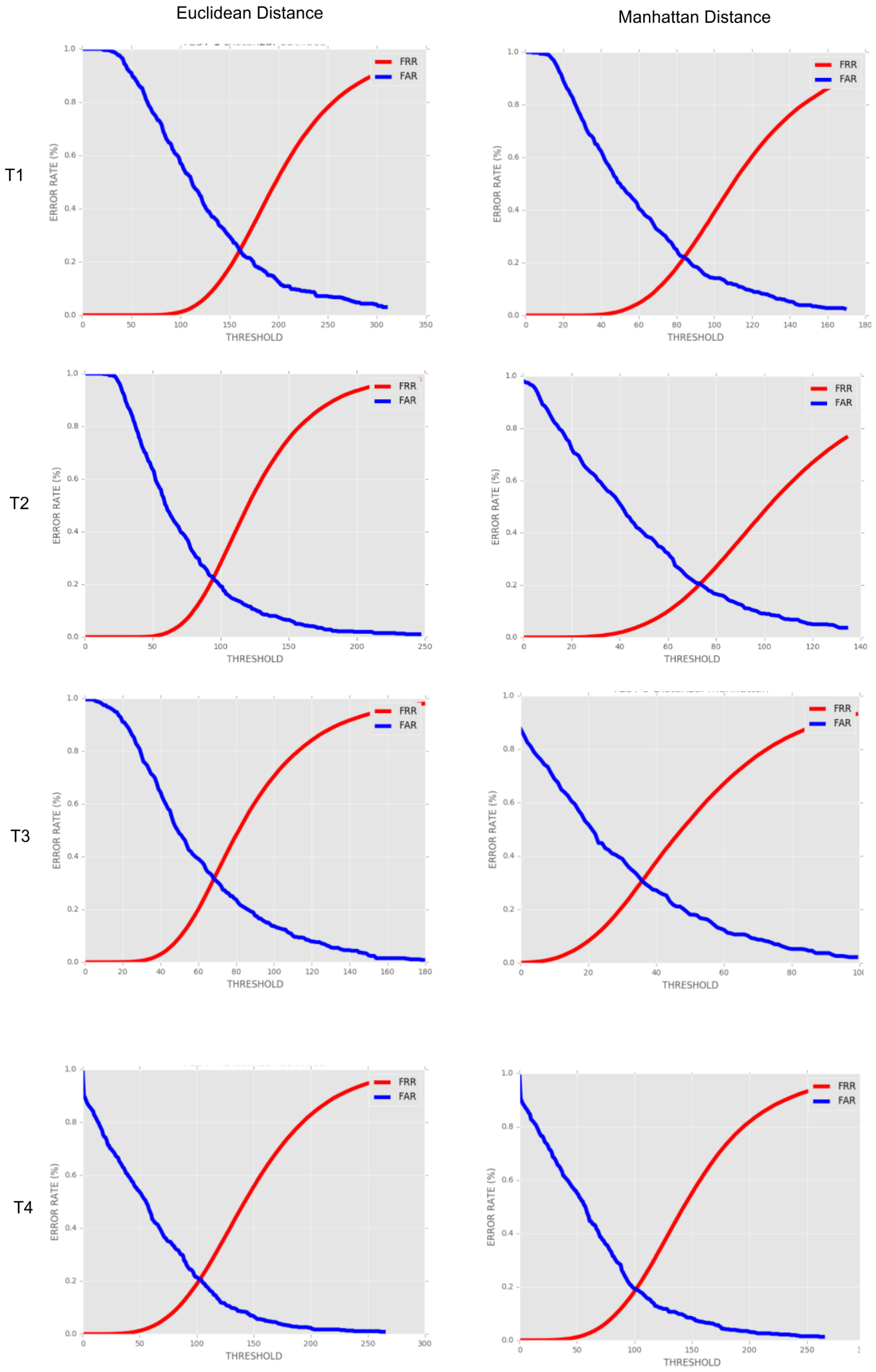


Figure 5.13. FAR and FRR for all test scenarios using both Euclidean Distance and Manhattan Distance.

with Multiple Template modality. On one hand, the algorithm comparing the whole signal (EER=9.2%), as well as the best of those exploiting step segmentation (EER=10.25%), got better results than the approaches presented here. These two algorithms are the slowest and most computational demanding in [76]. Moreover, the first one requires signals to be not dramatically different in terms of length. On the other hand, the other three proposals in [76] based on step segmentation compute distance with a comparing strategy comparable in terms of computational costs and speed to those proposed here, and allow releasing the constraint of a similar number of steps. However, they show lower performances (EER of 32.8%, 41.04%, and 0.3625% respectively). Approaches based on feature extraction work on a kind of aggregated information that does not depend on the signal length, given that it is long enough to extract sufficient information. The above comparison seems to suggest a possible compromise between different application constraints, that deserves more investigation. As a further comparison, the results in [55] are reported too. We considered only the single accelerometer scenario (i.e., entailing the same setting of our experiments). That work achieves an EER that ranges from 8.6% to 13% (depending on the chosen body location). It is generally better than our approach, but it is worth noticing that their procedure requires about 0.3 second to run on a powerful pc, while our approach is devised to work on smartphones that have lower computational power. It is further worth underlining that making more comparisons with other state of the art methods is not possible at present, due to the different (and generally much smaller) datasets used.

5.3 Other Attempts

During this PhD, we also tried other different approaches for preprocessing, step segmentation and signal processing. Some interesting ones are reported in the following. It is also worth underlining that these strategies do not produce a global[5] improvement in the performances and some of them even decrease the system accuracy (the benchmarks are generally our previous works). As a general protocol, we decided to test the different strategies starting from our in-house dataset or on a portion (the first 50 subjects with the first 4 walks) of the ZJU-gaitacc dataset. Then we tried to apply the same method on complete ZJU-gaitacc dataset only if the approach seemed to be promising. This protocol is mainly chosen due to the fact that a complete test with such dataset generally requires even more than one week on a medium level notebook. For the preprocessing and step segmentation strategies, the comparison methods are the ones presented in 5.2 and the evaluation of the results is based on verification/recognition performances. This is because a pure evaluation of the accuracy of a step segmentation algorithm, for example, would require one or more human-annotated datasets, that are not available at the moment.

[5] There are cases in which the same strategies reports a slight improvement in some scenarios and with some recognition methods but significantly decrease the results when used with other recognition strategies and/or datasets.

5.3.1 Preprocessing Strategies

We firstly tried to apply the state-of-the-art preprocessing strategies in order to enhance the gait signal quality, such as the already mentioned interpolation, noise reduction, filters and signal enhancing/smoothing strategies.

We applied linear interpolation with different interval lengths. It is worth underlining that interpolation has not been tested on ZJU-gaitacc dataset because it presents walking signals that are already interpolated. We noticed that some of the resulting signals in our dataset lose peaks and other informative characteristics and this also negatively impacts on the performances. This is probably due to the high sampling rate of the exploited accelerometer(s). In fact, when the highest sampling rate has been tested (namely 1000Hz), the results were more or less the same than the ones obtained with the original signals although the execution time is considerably higher. We also tried a modified version of interpolation that adds missing points equal to the previous computed point. We called this strategy Plateau Interpolation because it tends to create some plateau in the resulting signal. Unfortunately, this interpolation strategy seems to produce lower results with respect to the linear version. Finally, we also tried non-linear interpolation approaches, using the spline version, with even worse results.

We tested different noise reduction strategies, applying either moving average, weighted moving average, or wavelet denoising. For all of these families of algorithms, we tested different parameters (and mother wavelets in the third case). The resulting signals, even if they shown a more regular appearance, seem to lose some of their discriminative characteristics, producing a degradation in terms of performances during the biometric recognition phase. Similar results have been achieved using filters. In this case we tested high pass and low pass filters. For the last category, we also tested the Butterworth filter, used in some literature works. We tried to apply Gaussian normalization too, but the results stayed unchanged though slightly increasing computational time. Finally, we tested a modified version of Gaussian normalization, that computes two pairs of μ and σ namely μ_{under}, σ_{under} and μ_{over}, σ_{over}. Such pairs are computed as usual but taking into account only points under the mean of the original signal and over it, respectively. Unfortunately, this strategy did not provide a global increase in term of performances and worked better only in some cases.

For the signal smoothing strategies, we also developed and tested a modified version of the symmetric moving average. Given a window of length k, for each point i, it computes the average μ of its $\frac{k}{2}$ neighbour (taken symmetrically). During the computation, it substitutes the current $i-th$ point with μ if its value is lower than μ or let it unchanged otherwise.

We also tried an in-house signal enhancing strategy. Given a window of length k, for each point i it computes the sums of its $\frac{k}{2}$ previous and following neighbors respectively. Then, for each point i, if the absolute distance is greater than an suitable threshold (we test different values for it), the procedure substitutes the i-th point with a multiplied version of it (we test $i*1.5$, $i*1.75$, and $i*2$). This procedure, even if effectively enhances peaks in the signal, did not produce an improvement in the comparison phase, probably because it also increases the weight of noisy points.

Another tested attempt exploits the standard magnitude vector, often used in literature. Even if this strategy lowered the computation time of the comparison phase with respect to repeat the comparison one time per axis, it generally produced lower results too. This is probably due to the dominance of one axis with respect to the others (see Section 4.3) that is totally lost with aggregative preprocessing.

We also tried different combinations of the mentioned approaches, but none of them seemed to provide a global and significant improvement, probably for the reasons already mentioned above. Therefore, in our works, we decided to use raw data without any preprocessing (except for the ZJU-gaitacc dataset, that is published already interpolated by the authors). The only preprocessing strategy that seemed to provide a global improvement with our system is the already mentioned Gaussian Kernel Convolution (see 5.2.2).

5.3.2 Step Segmentation Procedures

We tried to improve our step segmentation procedure in different ways. Some of such modified versions provided good results and have been already described in Section 5.2. Other tested variations generally provide a less accurate segmentation. A first group of these attempts exploits the noise reduction and the interpolation strategies mentioned above in order to clean the signal only for segmentation sake. One problem of the application of these algorithms is the slight shifting of local maxima and minima. In fact, when these preprocessing strategies are applied, the resulting signal presents a more or less similar shape, but the relevant points are shifted with respect to the original ones. For this reason, these versions require a further phase to realign the points. Notwithstanding this, the segmentation accuracy did not report a global improvement. A second group of attempts exploits the Persistence 1D algorithm[6]. This algorithm aims at finding "local extrema and their persistence in one-dimensional data". One of the most interesting factor is that this procedure does not modify the original signal and the found local maxima and minima are in 1 to 1 correspondence with it. Unfortunately, it generally provides more points with respect to the ones corresponding to starts and stops of the steps. We tried different strategies in order to collapse such points, in order to keep only the more relevant ones, but all the produced segmentations provided lower performances during the recognition tests compared with the step segmentation strategy proposed in [75] and its evolutions.

5.3.3 Strategies for the Comparison Subsystem

When developing strategies for the comparison subsystem during the project course in which the Biometric Walk Recognizer approach takes the first shape, we decided to use a standard and robust 1D signal comparison method: the basic formulation of the DTW algorithm. This choice was mainly due to its easy implementation and to the fact that, considering the average number of points making up a single walk (about 1500 samples), its computational complexity ($O(n^2)$) is reasonable. Considering the good results achieved by this "simple" method, we tried to apply different state-of-the-art adaptations/modifications of it.

[6] https://www.csc.kth.se/~weinkauf/notes/persistence1d.html

A first attempt is the 2D DTW version. Such version uses the timestamps for the horizontal axis and acceleration values for the vertical axis. Of course, this version provides identical results to the one of the basic formulation if the data are interpolated. This is not the case for our in-house dataset and, in general, it is not the case of the raw data coming from an Android device (as mentioned in 4.1). Unfortunately, this attempt did not improve accuracy results, probably due to the fact that smartphone acquired data, even if not with a fixed sampling rate, are collected with an high-enough frequency. We also tried this 2D version with magnitude vectors, achieving similar results. It is also worth noticing that all the public freely available datasets collect data without any rotation, so that the main advantage of the use of magnitude vector, i.e. the rotation invariant property, is not actually exploited.

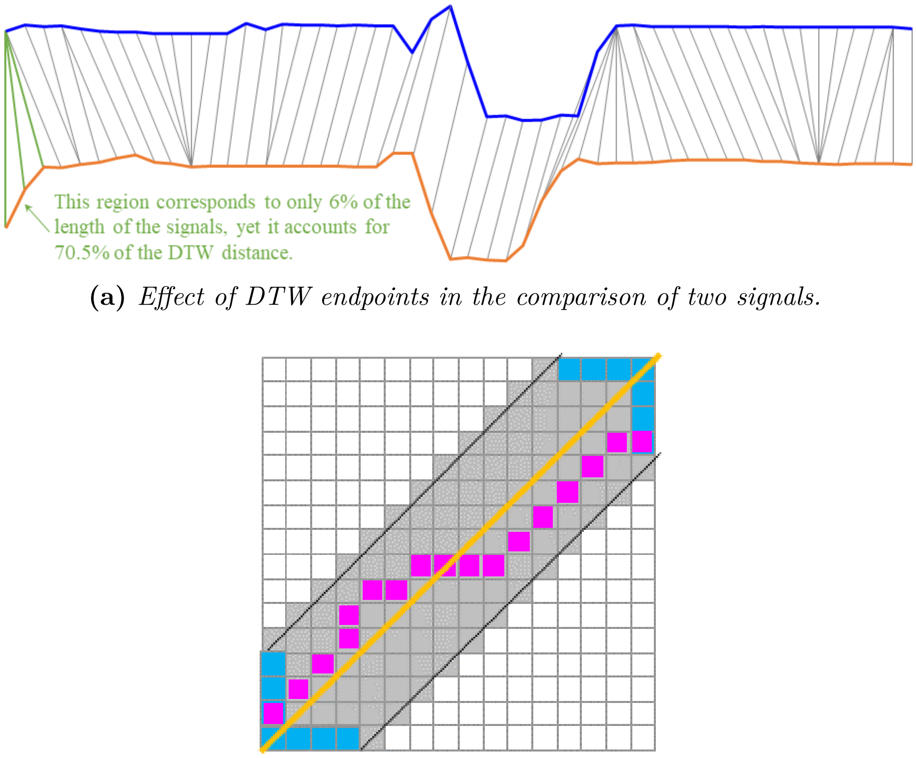
In our DTW-based recognition strategies, each axis is taken as a "totally" independent source. This means that there is no correlation with each out of the three DTW comparisons. Of course, this means that DTW can possibly choose different warping paths in order to find the best correspondence on each axis. An attempt we tried has been to find the warping path (and related distance) from the dominant axis and to force the same path on the other two axes. Unfortunately, this strategy seemed to produce lower results.

Another attempt we tested is to apply the DTW modifications proposed by Abdullah Mueen and Eamonn Keogh[7]. These modifications allow a faster comparison, and take into account the endpoints and values misalignments. In fact, as highlighted in their guide, it can be possible that a not negligible part of the final distance score provided by DTW is due to the comparison of first and last points or to a vertical shifting (DTW only align values in time). This strategy, from the theoretical point of view, presents some interesting properties. First of all, using a window of length w, the search space for the best warping path is limited and this decreases the computational time. As a second advantage, the management of endpoints produces more accurate distances. The endpoints are the firsts and the lasts points in the time series. In order to manage them, we set the number of endpoints as w and we allow the warping path to start and stop from any of such endpoints. Figure 5.14 shows an example of the effect of endpoints during the comparison of two signals and their visualization on the DTW matrix.

As third advantage, the z-normalization allows to better compare the shapes of gait signals. Unfortunately, this strategy produced only a slight improvement for WALK while the performances for the methods exploiting step segmentation was identical or even lower. This is due to the low number of samples making up the steps (about 100). In these cases, the reduction of the warping path joined with the use of endpoints seems to squeeze the distance distribution, rising errors. We also tried to apply each of the techniques separately, without any improvement.

Another attempt entails the comparison of signal shapes, only considering the directions and not the values. In order to do this, we cleaned the signal with a Gaussian kernel convolution (using $\sigma = 16$, see also 5.2.2). Then we converted the signal in a series of 1, 0 and -1 representing when it is in increasing, stationary or decreasing phase, respectively. Finally, we compared them with DTW algorithm,

[7] `www.cs.unm.edu/~mueen/DTW.pdf`


This region corresponds to only 6% of the length of the signals, yet it accounts for 70.5% of the DTW distance.

(a) *Effect of DTW endpoints in the comparison of two signals.*

(b) *DTW endpoints (in blue) in the DTW matrix.*

Figure 5.14. Example of DTW endpoints.

with poor results.

We also tried to exploit histogram similarity, dividing signal values into 10 bins and using classical histogram comparison strategies. We tried this strategy both with entire signal and with single steps. In both cases, the results reported worst performances with respect to our benchmark.

Finally, we exploit wavelet convolution, in order to extract feature from gait signals. We tested different mother wavelets with different parameters but we reported lower results even in this case.

5.4 Accelerometer Data Normalization and Application to Gait Recognition

There are few works in literature facing the problem of cross-device comparisons. In a real scenario, especially in identification settings, e.g., to grant access to authorized users without an explicit authentication, a possible requirement can be the recognition of a subject even if he/she is using a different smartphone from the one used during the enrolment. In general, the comparison of cross-device signals is expected to achieve lower accuracy, unless using a preprocessing procedure that tries to realign the signals in terms of recorded values. As discussed in Section 4.1.1.1, even two identical accelerometers are likely to produce different values in identical positions and conditions. To have a proof of this fact, it is possible to place two smartphones over a table and have a look at the values red by the sensor with any App capable to show them. This calibration problem is common for all devices and, if on one hand, for common tasks (such as screen rotation or gaming) it is negligible, on the other hand, for biometric purposes, it is not. A possible solution to this problem is discussed in our approach in [52], presented at *BIOSIG 2016* conference. It entails a high level procedure that does not require to manipulate the sensor itself, even because, being embedded into another device, this will not be a feasible task. The procedure is very simple, can be easily repeated whenever it is necessary, and the normalization parameters are recorded in a text file stored on the device itself. Therefore, they can be used by any applications that may benefit from them. Moreover, the procedure can be executed once and for all[8].

Let's assume that the *Offset* (at $0g$) and the value at $1g$ (*Ref_ Value* in the following) have been already computed. The general equation 5.2 synthesizes the proposed normalization formula. It is derived from the well-known Min-Max formula, that is used to map a given *Value* onto a *New_ Value* in the $[0, 1]$ interval. It has the general formulation: $New\_Value = (Value - Min(Value)/Max(Value) - Min(Value))$. The New_Value in $[0, 1]$ can be further translated into a new interval $[New\_Min, New\_Max]$ by $New\_Value*(New\_Max-New\_Min)+New\_Min$. The equation 5.2 represents a variation of this schema. It does not take as reference values the minimum and maximum measured by the accelerometer, that are not easily identifiable for each device, but rather the two reference values measured at $0g$ and $1g$ for each axis (that may fall in different points in the accelerometer range). The general equation is then specialized over the three axes:

$$New\_Value = \frac{Value - Offset}{Ref\_Value - Offset} \tag{5.2}$$

In this way, a rescaling in $[0g, 1g]$ is obtained. This allows normalizing the range of values from different accelerometers. Since the accelerometer produces a linear response within the measure range, each movement is translated into a discrete value that is directly proportional to the physical acceleration. By aligning the results with respect to $[0g, 1g]$ it is also possible to achieve a correct alignment of the values originally not included in the same interval.

The goal of the procedure is therefore to compute $Offset$ and Ref_Value for each axis. The procedure requires a series of six simple tests. Each of them actually provides a value which is the average of samples taken over a continuous interval of 15 seconds. The device positions to obtain the required measures are shown in Table 5.9, where they are referred to the smartphone layout. These positions are chosen because it is possible to know in advance the expected (and correct) values to get.

To better explain, let suppose to have the device put on a plain surface with the screen up. In this position, it is possible to measure the offset values at $0g$ for X and Y axes and the reference value (*Ref_ Value*) at $1g$ for Z axis. This is done by acquiring data for 15 seconds, and then computing the average value for each axis. Therefore, from the measurements in this position we obtain three values:

[8] Having to do with a physical sensor, as suggested in datasheets, it is however preferable to re-calibrate the sensor periodically, because real output data can even differ over time

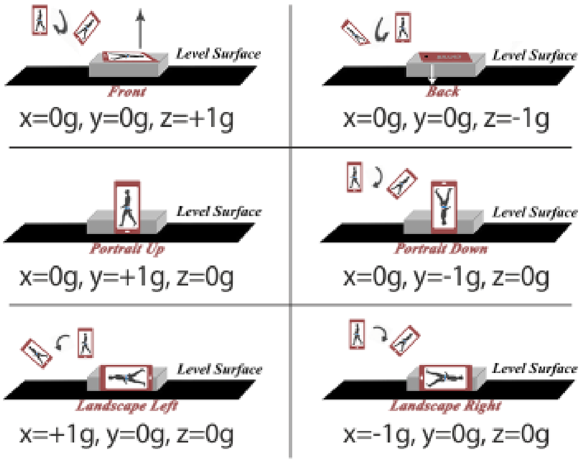


Table 5.9. Ideal accelerometer values with respect to the smartphone positions.

Ref_Z_Value and two values that will be used to compute the final offsets for X and Y, i.e., *X_Front* and *Y_Front*. For each axis, the final value for *Offset* will be given by the average of 4 values measured at $0g$ in different positions. For instance, as for X axis, such values will be the average of *X_Front*, *X_Back*, *X_PortraitUp*, and *X_PortraitDown*. Equations 5.3, 5.4, and 5.5 show the required computation to obtain the 3 *Offset* values.

$$X\_Offset = \frac{X\_Front + X\_Back + X\_PortraitUp + X\_PortraitDown}{4} \tag{5.3}$$

$$Y\_Offset = \frac{Y\_Front + Y\_Back + Y\_LandscapeLeft + Y\_LandscapeRight}{4} \tag{5.4}$$

$$Z\_Offset = \frac{Z\_LandscapeLeft + Z\_LandscapeRight + Z\_PortraitUp + Z\_PortraitDown}{4} \tag{5.5}$$

As a further advantage of this formulation, it is possible to apply this procedure even after the data acquisition, provided that the capture device is available. The results achieved after this preprocessing seems to significantly exceed the ones obtained using the original signal especially in cross device setting. Moreover, even considering the data from the same device, the performances have a slight increase (see Tables 5.12 and 5.13).

Summarizing, the advantages of this procedure are: it requires few minutes, it is one-shot, and it provides a file that can be red by all other applications in order to normalize the accelerometer data from time to time. As a drawback, it requires to compute the normalization parameters on the device and such values cannot be transferred from a device to another. This means that, for example, it is not possible to apply it on external datasets due to the lack of the required information.

In order to test our procedure, we collected a new dataset with data coming from different devices. In fact, at the moment, there are no publicly available datasets with this characteristic. The exploited dataset is the BWR-MultiDevice, already described in Section 4.2.

In order to test the advantages of using our normalization procedure, the 5 recognition algorithms proposed in our previous work [76] are used. Experiments are designed to focus exclusively on normalization effects, so we decided not to apply any kind of denoising and/or interpolation preprocessing. Results are computed for all 3 recognition modalities, namely Closed Set Identification (CSI), Verification (VER), and Open Set Identification (OSI). Half of these tests are performed using the walk signals without normalization and the other half using the same data after the application of the proposed normalization procedure. The full set of analyzed test scenarios (*ts*s) is created combining the following conditions:

- **AllSessions** = probe and gallery sets belong to both sessions.
- **Session_vs_Session** = one session in turn is used as probe set and the other as gallery set.
- **SameSession** = probe and gallery sets belong to the same session (this *ts* is not realistic but can be used as reference point).
- **AllDevices** = both probe and gallery sets belong to all devices.
- **Device_vs_Device** = one device is used in turn as probe and one as gallery source.
- **SameDevice** = both probe and gallery set come from the same device.
  The only missing combination is SameSession with Device_vs_Device, because it is not much realistic.

To summarize, the test scenarios (*ts*s) are the following: AllSessions combined with AllDevices (1 *ts*), with Device_vs_Device (6 *ts*s), and with SameDevice (3 *ts*s) generates a total of 10 possible *ts*s; Session_vs_Session combined with AllDevices (2 *ts*s), with Device_vs_Device (12 *ts*s), and with SameDevice (6 *ts*s), generates a total of 20 *ts*s; finally SameSession combined with AllDevices (2 *ts*s), and with SameDevice (6 *ts*s) (in this case the combinations with Device_vs_Device are skipped because it is not much realistic), generates a total of 8 *ts*s. Summing them up, there is a total of 38 *ts*s, each one enacted with each of the 5 recognition algorithms in each of the 3 different recognition modalities, for a total of 570 tests. Moreover, each test is repeated with and without normalization, for a total of 1140 tests, out of which we report those carried out with the recognition algorithms achieving the best results. An all-against-all comparison is always carried out, where, given the chosen scenario, each template in the probe set is used in turn and compared with all the gallery.

All tests in all scenarios demonstrate a generally significant performance improvement, that reaches up to 225%. Even for tests involving a single device there is an extremely high improvement, though unexpected given that they involve exactly the same smartphone. Analyzing the results reported in Table 5.12 for WALK algorithm and in Table 5.13 for ALL STEPS VS. ALL, it is possible to observe that also using a single device, normalization increases accuracy. In fact, we got up to a 89.74% relative improvement in CSI, achieved using Samsung with which the results pass from 39% to 74% of RR; the improvement in VER reaches 157.14%, with the performances passing from 36% to 14% of EER, and a 75.68% improvement in OSI, with performances passing from 70% to 30% of EER, in the case of SameDevice combined with SameSession using OnePlus with ALL STEPS VS ALL.

Table 5.10. Compact view of results achieved by WALK.

Closed Set Identification - WALK			
Test	RR - O.D.	RR - N.D.	Improv.
AllDevices	52.0%	54.5%	4.81%
Device_vs_Device	35.3%	49.3%	39.62%
SameDevice	50.3%	52.0%	3.31%

Verification - WALK			
Test	ERR - O.D.	ERR - N.D.	Improv.
AllDevices	31.8%	29.6%	7.43%
Device_vs_Device	31.4%	29.5%	6.23%
SameDevice	28.8%	29.0%	-0.69%

Open Set Identification - WALK			
Test	ERR - O.D.	ERR - N.D.	Improv.
AllDevices	31.8%	29.6%	7.43%
Device_vs_Device	79.2%	72.3%	9.45%
SameDevice	25.2%	25.0%	0.67%

Table 5.11. Compact view of results achieved by ALL STEPS VS. ALL.

Closed Set Identification - ALL STEPS VS. ALL			
Test	RR - O.D.	RR - N.D.	Improv.
AllDevices	22.5%	34.5%	53.33%
Device_vs_Device	15.2%	26.7%	47.06%
SameDevice	22.0%	29.0%	106.67%

Verification - ALL STEPS VS. ALL			
Test	ERR - O.D.	ERR - N.D.	Improv.
AllDevices	50.0%	50.0%	0.00%
Device_vs_Device	47.2%	42.5%	11.17%
SameDevice	44.9%	43.1%	4.26%

Open Set Identification - ALL STEPS VS. ALL			
Test	ERR - O.D.	ERR - N.D.	Improv.
AllDevices	83.3%	71.3%	16.83%
Device_vs_Device	92.2%	88.7%	3.95%
SameDevice	65.8%	47.7%	38.11%

In the cases of Device_vs_Device *ts*s, we got the following improvements: up to 131.25% in CSI, when combined with AllSessions using templates from Samsung as probes and templates from Sony as gallery, with the ALL STEPS VS. ALL, with the result increasing from 32% to 74% of RR; we got up to 34.5% improvement in VER when combined with AllSessions using Sony templates as probes and OnePlus templates as gallery with the ALL STEPS VS. ALL, with results improving from 44.8% to 33.3% of EER; finally, we got up to 64.71% relative improvement in OSI when combined with AllSessions using Samsung templates as probes, and Sony templates as gallery with the WALK recognition method, with results passing from 56% to 34% of EER. Notwithstanding the peaks due to worse starting values, the improvements in cross-device comparison are higher on average than those obtained when using the same device.

Tables 5.10 and 5.11 show a compact view of the performances achieved by WALK and ALL STEPS VS. ALL respectively. The results are averaged for the *ts*s involving AllDevices, Device vs Device, and SameDevice in turn.

Table 5.12 shows the complete set of results for the *ts*s with the recognition method WALK, that experimentally achieved the best results among the recognition methods in [75], even without normalization. For symmetric conditions, e.g. the pair of Session_vs_Session, we just report the average results. With WALK, the global performances stay the same or increase. The *ts*s involving different devices, especially in Device_vs_Device setting, always achieve an improvement in all three recognition modalities (except for few *ts*s in VER and in OSI). The improvements are up to 70.37% for CSI, up to 60,3% for VER, and up to 64,7% in OSI. It is worth pointing out that the major benefits of normalization, as expected, are in Device_vs_Device *ts*s.

Table 5.13 shows the complete set of results for *ts*s with the recognition method ALL STEPS VS. ALL, that is the one that achieved the best benefits from normalization. Even in this case, for symmetric conditions we just report the average results. The baseline performances are lower than the previous ones. This comparison algorithm is free from the limitation to require about the same number of steps/segments in the signals to compare, and therefore this result was expected due to possible greater inaccuracy. In fact, this algorithm uses step segmentation, and compares single steps having a limited number of signal points (about 100). Therefore, it is more affected by signal distortions, such as systematic errors. In fact, in the original (not normalized) dataset, it achieves quite poor performances. However, it is interesting to notice that normalization provides an even higher improvement with respect to WALK, especially in Device_vs_Device *ts*s, confirming the benefits of produced by a normalization procedure.

In conclusion, the contribution of this part of the research is not a new gait comparison strategy but rather to demonstrate that a normalization procedure can significantly improve the accuracy achieved with gait signals from different embedded devices. Solutions that aims at improving accelerometer signal quality for gait recognition, at the best of our knowledge, do not tackle extensively cross-device signal comparisons. This proposal reports an effective procedure for signal normalization.

Unfortunately, it was not possible to test it on other existing datasets, due to the lack of the required measures, but it can be easily implemented for a brand new system gallery, either during acquisition, or even afterwards, given that the requested measures can be (possibly) computed later. Experimental results demonstrate that normalization also positively affects the comparisons of data from the same device, though being especially beneficial in the cross-device ones. This confirms that it is not possible to export normalization parameters from one sensor to another, though of the same model.

5.5 Other Works

This Section presents other works related to gait recognition and possible application using either the combination with other biometrics or a different architecture, such as cloud services.

The approach presented in [125] sketches the possible fusion between gait and spatial handwriting recognition. The goal is to provide a continuous and transparent user

Table 5.12. Results with WALK. The green cells (lighter) report improvement while the red ones (darker) decrements.

Recognition Method: WALK												
Test Scenario		Device		Closed Set Identification			Verification			Open Set Identification		
				RR O.D.	RR N.D.	Improv.	EER O.D.	EER N.D.	Improv.	EER O.D.	EER N.D.	Improv.
AS	AD			95.00%	97.00%	2.11%	28.00%	24.10%	16.18%	26.30%	23.70%	10.97%
	D_vs_D	OP	Sams	90.00%	92.00%	2.22%	20.70%	21.00%	-1.45%	38.00%	35.00%	8.57%
		OP	Sony	65.00%	87.00%	33.85%	29.50%	24.70%	19.43%	58.00%	43.00%	34.88%
		Sony	Sams	74.00%	88.00%	18.92%	27.30%	24.80%	10.08%	53.00%	35.50%	49.30%
		Sony	OP	76.00%	87.00%	14.47%	29.50%	24.70%	19.43%	57.00%	39.00%	46.15%
		Sams	OP	97.00%	99.00%	2.06%	20.70%	21.00%	-1.45%	30.00%	25.50%	17.65%
		Sams	Sony	70.00%	93.00%	32.86%	27.30%	24.80%	10.08%	56.00%	34.00%	64.71%
	SD	OP		94.00%	94.00%	0.00%	24.60%	23.50%	4.68%	24.60%	23.60%	4.24%
		Sony		90.00%	90.00%	0.00%	28.00%	28.00%	0.00%	28.00%	28.40%	-1.43%
		Sams		96.00%	96.00%	0.00%	24.40%	24.00%	1.67%	24.40%	24.00%	1.67%
S_vs_S	AD			52.00%	54.50%	4.81%	31.80%	29.60%	7.43%	31.80%	29.60%	7.43%
	D_vs_D	OP	Sams	39.00%	49.00%	25.64%	28.08%	28.45%	-1.34%	75.00%	70.00%	7.14%
		OP	Sony	32.00%	48.00%	50.00%	32.45%	26.43%	22.80%	82.00%	73.00%	12.33%
		Sony	Sams	27.00%	46.00%	70.37%	33.50%	31.55%	6.18%	82.00%	78.00%	5.13%
		Sony	OP	27.00%	42.00%	55.56%	32.50%	31.75%	2.36%	85.00%	76.00%	11.84%
		Sams	OP	50.00%	55.00%	10.00%	28.08%	30.53%	-8.73%	70.00%	70.00%	0.00%
		Sams	Sony	37.00%	56.00%	51.35%	33.50%	28.38%	18.06%	81.00%	67.00%	20.90%
	SD	OP		56.00%	57.00%	1.79%	27.00%	26.80%	0.75%	27.00%	26.80%	0.75%
		Sony		43.00%	45.00%	4.65%	31.20%	31.20%	0.00%	31.20%	31.20%	0.00%
		Sams		52.00%	54.00%	3.85%	28.20%	29.00%	-2.84%	28.20%	29.00%	-2.84%
SS	AD			94.50%	95.50%	1.06%	21.40%	13.35%	60.30%	23.00%	22.10%	4.07%
	SD	OP		91.00%	91.00%	0.00%	8.00%	8.00%	0.00%	25.00%	25.00%	0.00%
		Sony		94.00%	95.00%	1.06%	6.00%	6.00%	0.00%	20.00%	20.00%	0.00%
		Sams		72.00%	72.00%	0.00%	16.00%	15.95%	0.31%	30.50%	30.00%	1.67%

Legend of acronyms:

AD=AllDevices
AS=AllSessions
D_vs_D=Device_vs_Device

EER=Equal Error Rate
N.D.=Normalized Dataset
O.D.=Original Dataset

OP=OnePlus
RR=Recognition Rate
S_vs_S=Session_vs_Session

Sams=Samsung
SD=SameDevice
SS=SameSession

Table 5.13. Results with ALL STEPS VS. ALL. The green cells (lighter) report improvement while the red ones (darker) decrements.

Recognition Method: ALL STEPS VS. ALL												
Test Scenario		Device		Closed Set Identification			Verification			Open Set Identification		
				RR O.D.	RR N.D.	Improv.	EER O.D.	EER N.D.	Improv.	EER O.D.	EER N.D.	Improv.
AS	AD			51.00%	78.00%	52.94%	50.00%	50.00%	0.00%	57.70%	38.60%	49.48%
	D_vs_D	OP	Sams	45.00%	67.00%	48.89%	50.00%	50.00%	0.00%	70.50%	48.50%	45.36%
		OP	Sony	32.00%	52.00%	62.50%	44.00%	37.70%	16.71%	79.00%	61.50%	28.46%
		Sony	Sams	28.00%	57.00%	103.57%	50.00%	50.00%	0.00%	84.00%	57.00%	47.37%
		Sony	OP	38.00%	56.00%	47.37%	44.80%	33.30%	34.53%	74.00%	58.50%	26.50%
		Sams	OP	54.00%	76.00%	40.74%	50.00%	50.00%	0.00%	67.00%	48.00%	39.58%
		Sams	Sony	32.00%	74.00%	131.25%	50.00%	50.00%	0.00%	67.00%	48.00%	39.58%
	SD	OP		43.00%	73.00%	69.77%	40.30%	32.70%	23.24%	40.30%	32.60%	23.62%
		Sony		52.00%	51.00%	-1.96%	40.30%	38.80%	3.87%	40.30%	38.80%	3.87%
		Sams		39.00%	74.00%	89.74%	50.00%	50.00%	0.00%	50.00%	50.00%	0.00%
S_vs_S	AD			22.50%	34.50%	53.33%	50.00%	50.00%	0.00%	83.30%	71.30%	16.83%
	D_vs_D	OP	Sams	17.00%	25.00%	47.06%	45.00%	38.95%	15.53%	92.50%	85.00%	8.82%
		OP	Sony	15.00%	31.00%	106.67%	44.60%	38.05%	17.21%	90.00%	83.50%	7.78%
		Sony	Sams	13.00%	23.00%	76.92%	47.50%	39.65%	19.80%	94.50%	91.00%	3.85%
		Sony	OP	14.00%	22.00%	57.14%	47.65%	46.10%	3.36%	92.00%	91.50%	0.55%
		Sams	OP	16.00%	30.00%	87.50%	50.00%	50.00%	0.00%	93.50%	91.00%	2.75%
		Sams	Sony	16.00%	29.00%	81.25%	48.50%	42.05%	15.34%	90.50%	90.00%	0.56%
	SD	OP		18.00%	34.00%	88.89%	42.85%	36.55%	17.24%	90.00%	80.00%	12.50%
		Sony		25.00%	22.00%	-13.64%	41.85%	42.75%	-2.15%	84.00%	88.00%	-4.76%
		Sams		23.00%	31.00%	34.78%	50.00%	49.90%	0.20%	89.50%	82.00%	9.15%
SS	AD			54.00%	79.00%	46.30%	44.05%	35.00%	25.86%	56.50%	38.35%	47.33%
	SD	OP		40.00%	71.00%	77.50%	36.00%	14.00%	157.14%	65.00%	37.00%	75.68%
		Sony		54.00%	54.00%	0.00%	27.60%	26.25%	5.14%	58.50%	57.50%	1.74%
		Sams		40.00%	75.00%	87.50%	40.40%	32.00%	26.25%	74.00%	48.50%	52.58%

Legend of acronyms:

AD=AllDevices	EER=Equal Error Rate	OP=OnePlus	Sams=Samsung
AS=AllSessions	N.D.=Normalized Dataset	RR=Recognition Rate	SD=SameDevice
D_vs_D=Device_vs_Device	O.D.=Original Dataset	S_vs_S=Session_vs_Session	SS=SameSession

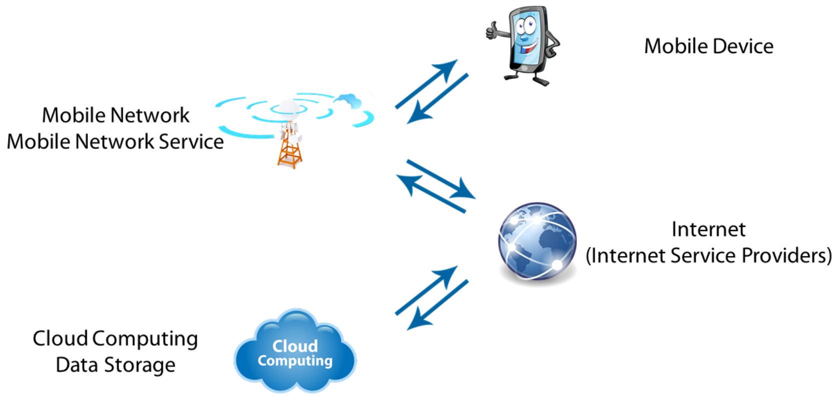


Figure 5.15. General purpose Mobile+Cloud architecture.

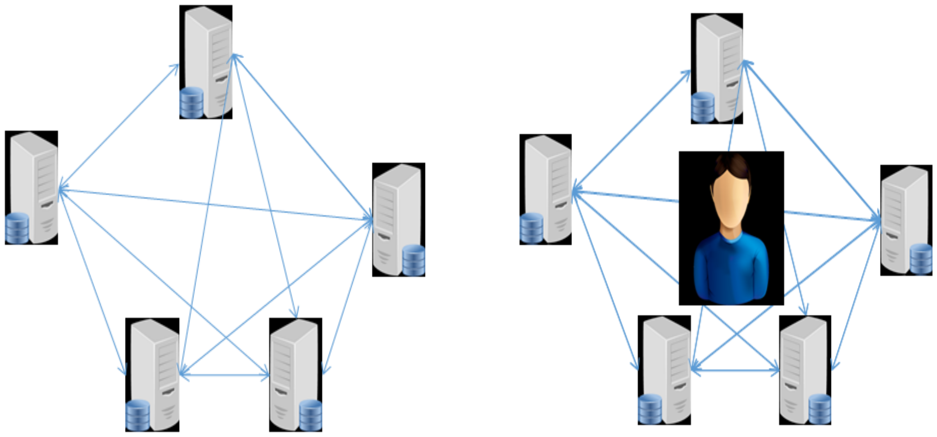
authentication. For the spatial handwriting, in collaboration with the Shizuoka University, the recognition is performed on Kanji language written in the air. In fact, in Japan, the writing system is based on two alphabets and a set of over 2000 Kanji characters. Moreover, Japanese names are typically written in Kanji and there are different ways to write the same name. On many occasions, Japanese would illustrate their names by "handwriting in the air". This is a natural communication paradigm widely accepted in Japan, and for this reason it can be considered as a kind of natural interaction. In addition, a preliminary study [126] shows that the tracking of the spatial movement can possibly be used as a (soft) biometric trait. A prototype of recognition system based on the fusion of these two biometrics can require a walk in an hallway followed by the handwriting of the user name. The system has the aim to verify if the claimed identity (given by the air handwritten name, possibly captured by a smartwatch) is the one declared by the user exploiting both the gait signal (possibly captured by a smartphone located in the hip zone or in the trouser pocket) and the handwriting dynamics.

The work in [127] draws the outline of the architecture and the possible techniques for a gait recognition system that collects data from smartphones' accelerometer and process them on the cloud, in order to reduce the computational burden on the smartphone itself. The architecture follows the mobile+cloud scheme (see Figure 5.15). From the communication security point of view, the new standards, such as HTTPS with TLS 1.2[9] or the new 1.3 version, are increasing more and more their encryption/protection capabilities, allowing a secure data transfer between a mobile device and the recognition server/cloud service. Moreover, it is possible to include in the acquisition application the requirement for a specific "fingerprint" on the Certification Authority (CA) TLS certificate, effectively blocking rogue CA, possibly used in Man in the Middle (MITM) attacks. This kind of architecture can also be adapted to other biometric traits. However, the only one that can be captured via mobile without any user explicit action is gait, and this makes related features particularly appealing.

The use of cloud services possibly introduces the need to address some security

[9] https://tools.ietf.org/html/rfc5246

issues. For example, it is important to assure that an infected/malicious server in the cloud cannot access to biometric templates. A way to face this problem can be the popular multiparty computational approach. In such model, the server(s) and client do not know each other's data. It also requires some form of encryption (e.g., homomorphic encryption) for communications. Moreover, each pair of parties is generally connected by a secure channel, and the communication is assumed to be synchronous. With this conditions it is possible to use either a server-centred or a user-centred model (see Figure 5.16).



(a) *Server Centred Model.* **(b)** *User Centred Model.*

Figure 5.16. Server and User Centred Models. The black box in the figure indicates data sequestered from the other parties.

In conclusion, the cloud offers the possibility to use an external and somehow independent authenticator that, except in the case of particular attacks, provides an optimal compromise between efficiency and security. The cloud can be replaced by a local server, but this can not be always possible due to lack of resources.

5.6 List of Research Contributions

This Section presents the list of the main contributions of this research:

- **A novel step segmentation procedure.** A novel step segmentation procedure has been developed and tested in [76]. Further improvements are also presented in [52, 120] in order to increase the flexibility, especially in a multi-device set up.

- **The investigation of the best preprocessing strategies for wearable sensor-based gait recognition.** Different literature preprocessing strategies have been tested during the PhD research, including experiments on large benchmark datasets (see 5.3), even if with low results. The work in [120] presents the use of the Gaussian kernel convolution and the benefits in terms of recognition accuracy achievable by exploits this kind of preprocessing. The achieved results are better than those reported in the paper introducing the ZJU-gaitacc dataset. It is not possible to establish further comparison with other works that used much smaller and in-house collected datasets. However, it will be possible in the future to compare the results reported here with those achieved by other methods exploiting the same public dataset.

- **The design of novel signal comparison strategies.** The work in [76] presents 5 recognition strategies based on DTW algorithm. The work in [122] describes a recognition strategy based on feature extraction and the results of experiments with different novel schemas for feature selection.

- **The proposal of possible real world applications.** The presented research investigates possible application scenarios and proposes a prototypical acquisition set up to reduce the battery draining while maintaining a quasi-ubiquitous recognition [118, 121]. It also investigates the possibility of exploiting cloud services for subjects recognition [127].

- **A possible solution to interoperability problems and a data normalization procedure.**. The work in [52] presents a study related to the decrease of performances achieved when comparing gait signals captured by different smartphones. It presents a novel data normalization procedure to reduce this problem. Such a procedure is general-purpose and can be also used for other applications requiring more accurate data from the accelerometer.

- **A new dataset with gait signals acquired by multiple devices.** The work in [52] presents a new freely available dataset containing walk signals from 25 subjects collected with 3 different acquisition devices (three smartphones of different brands) in two sessions.

- **An up-to-date literature review.** As mentioned before, the content of the Chapter 4 has been extracted from an extensive survey work still under review.

Chapter 6

Conclusions and Future Work

As pointed out in Section 4.4.6, most of the state-of the-art approaches exploit gait recognition in verification modality only. The implicit claim of the identity is represented by the fact that the owner of the device, e.g., a smartphone, is enrolled as the only subject in the system gallery. Therefore, the application only compares the corresponding template(s) with the probe acquired from time to time. This is different from what happens with computer vision-based approaches, which are mostly used in video surveillance, where there is no identity claim possibility, and biometric sample acquisition might also be covert or the user might not be (continuously) aware of it. The aim, in this case, is to identify an unclaimed identity, comparing it against an either white (allow transit) or black (rise an alarm) list, or to continuously re-identify (different) subjects appearing in videos. However, it is also possible to hypothesize a wider use of wearable sensor-based approaches. Low-consumption radio devices might remotely trigger the capture of the gait signal of an approaching subject. The signal might be sent to a local or remote server (or to a cloud service) controlling the access to a protected area, that might automatically open a secured door to authorized subjects only. The kind of interaction entails a mixture of awareness (the user has to voluntarily install a suitable application) and unaware capture (the user has nothing to explicitly do, so that the capture process can be maintained transparent). Moreover, for sake of defenders of privacy issues, it is to say that, while face and silhouette can be acquired in a completely covert way, this is not possible for the accelerometer signal. A concealed signal capture would require generally illicit techniques (an accelerometer hidden on an unaware subject, or injection of a hidden capture application) or to modify the production chain of, e.g., a smartphone (to set up a default/on demand remote transmission of the accelerometer signal not necessarily triggered by the user). On the other hand, especially when transmitting the complete accelerometer template, it is worth applying some encryption techniques to protect it from steal/spoof. In this kind of scenario, possible limitations of this biometric trait can be compensated for by using a multibiometric approach. For instance, it is possible to add face recognition from images acquired by a camera at a reasonable distance and with a frontal perspective on the person approaching a protected gate. This would be a common setting after all, that would not require a special behavior from the user, but walking normally: it is to consider that an average walking pace is highly probable while entering a protected zone.

It is worth devoting a special discussion to the use of the accelerometers and other sensors embedded in smartwatches. According to the works mentioned in Section 4.5.1, these kinds of sensors are suitable to distinguish walking state from standing, and for activity detection and possible recognition. However, the signal from the smartwatch accelerometer presents some disadvantages for gait recognition. Contrarily to the smartphone, the watch is positioned in a fixed position and does not change its orientation relative to the wrist. However, the orientation of the signal and its distribution over the axes changes sharply with (possibly involuntary) arm/hand movements. These get merged with the walk general stereotypic pattern, that may involve arm swinging too but gets disrupted by unrelated movements. Even the use of the magnitude vector to identify signal peaks, useful for segmentation, is hindered by possible wrist rotations. A reliable walking detection is critical, but also a preliminary action recognition can improve the subject recognition. The datasets used for gait recognition by wrist-worn sensors exploit a controlled capture with a limited number of possible action states, and further studies must investigate the true potential of using these devices that are recently widely spreading.

As for now, it is possible to summarize a list of the main open problems: 1) wearable sensor-based gait recognition suffers from variable device orientation, only partially addressed by aggregate values; 2) variable speed and ongoing actions modify natural gait kinematics when they are captured by accelerometers and similar sensors; 3) a different ground slope or the kind of shoes, especially in the case of high heels, can affect long term accuracy; 4) it is not obvious how to choose the best strategy for a comparison subsystem in relation with the walk length, given that the walk pattern stabilizes over a medium-long time; 5) annotated datasets with exhaustive demographic and context variations are not available yet; this would allow study on gender recognition, age estimation and/or ethnicity prediction as can be possible with other biometrics such as face.

The fact that automatic systems could recognize a subject from the walking pattern encourages to continue searching for features characterizing this trait. Different approaches attempted in the last years testify the wide range of techniques able to process gait signals and use them for recognition. The performances are generally interesting, notwithstanding the fact that gait is a soft biometrics. Therefore, gait signals can be considered as a good candidate for supporting strong biometrics. Especially for the wearable-based version, they can be combined with other traits with no interference, since the acquisition device is carried by the user. When combined with gait recorded by other sources (e.g., floor sensors or video), the synchronization of data could benefit all the involved systems. When used for identification, a mobile phone or smartwatch itself can be used as an "ID token". In fact, the user can be requested to "provide" the walk associated with the right identity, and also to carry the right device, as for a bank card and its pin code. Another interesting factor is the robustness of gait against impersonating attack, even by well-trained attackers. This can be a valuable support for a stronger biometric trait possibly being less robust to presentation attacks (a.k.a. spoofing). As it is happening with other biometrics, especially those based on video/image sources, proposals exploiting CNNs are a new research trend. The major problem with this kind of approach is the present absence of large enough datasets to best exploit the capability of deep architectures. A possible solution for this lack of data can be transfer learning, but to the best of our knowledge it has not been yet tested.

Even with all the limitations discussed in this thesis, gait recognition is a promising field of research. It can both become a reliable support for strong biometrics, and be used autonomously.

List of Publications

[52] Maria De Marsico, Daniele De Pasquale, and Alessio Mecca. "Embedded Accelerometer Signal Normalization for Cross-Device Gait Recognition". In: *15th Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*. IEEE. 2016, pp. 1–5.

[75] Maria De Marsico and Alessio Mecca. "Biometric Walk Recognizer". In: *International Conference on Image Analysis and Processing*. Springer. 2015, pp. 19–26.

[76] Maria De Marsico and Alessio Mecca. "Biometric Walk Recognizer - Gait recognition by a single smartphone accelerometer". In: *Multimedia Tools and Applications* (2016). DOI: 10.1007/s11042-016-3654-1.

[118] Maria De Marsico and Alessio Mecca. "Gait Recognition: The Wearable Solution". In: *Human Recognition in Unconstrained Environments: Using Computer Vision, Pattern Recognition and Machine Learning Methods for Biometrics* (2017), pp. 177–195.

[120] De Marsico, Maria and Alessio Mecca. "Benefits of Gaussian Convolution in Gait Recognition". In: *17th Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*. in-press, 2018.

[121] Alessio Mecca. "Impact of Gait Stabilization: a Study on How to Exploit it for User Recognition". In: *SITIS 2018 - International Conference on Signal Image Technology and Internet Based Systems*. IEEE, 2018.

[122] Maria De Marsico, Eduard Gabriel Fartade, and Alessio Mecca. "Feature-based Analysis of Gait Signals for Biometric Recognition". In: *ICPRAM 2018 - 7th International Conference on Pattern Recognition Applications and Methods*. SCITEPRESS, 2018, pp. 630–637.

[125] Kamen Kanev et al. "Mobiles and Wearables: Owner Biometrics and Authentication". In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM. 2016, pp. 318–319.

[127] Aniello Castiglione et al. "Walking on the Cloud: Gait Recognition, a Wearable Solution". In: *Network and System Security, 2018 12th International Conference on*. in-press, 2018.

List of Figures

List of Tables

Bibliography

[1] Roger Clarke. “Human identification in information systems: Management challenges and public policy issues”. In: *Information Technology & People* 7.4 (1994), pp. 6–37.

[2] Shirley Gaw and Edward W Felten. “Password management strategies for online accounts”. In: *Proceedings of the second Symposium on Usable Privacy and Security*. ACM. 2006, pp. 44–55.

[3] Dinei Florencio and Cormac Herley. “A large-scale study of web password habits”. In: *Proceedings of the 16th international conference on World Wide Web*. ACM. 2007, pp. 657–666.

[4] Eiji Hayashi and Jason Hong. “A diary study of password usage in daily life”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2011, pp. 2627–2630.

[5] Anil K Jain, Arun A Ross, and Karthik Nandakumar. *Introduction to biometrics*. Springer Science & Business Media, 2011.

[6] ISO. *ISO/IEC 2382-37:2017(E)*. https://standards.iso.org/ittf/PubliclyAvailableStandards/c066693_ISO_IEC_2382-37_2017.zip. [Online; accessed 21-December-2018]. 2017.

[7] ISO. *ISO/IEC 19795-1*. https://www.iso.org/obp/ui/#iso:std:iso-iec:19795:-1:ed-1:v1:en. [Online; accessed 12-January-2019]. 2006.

[8] Christopher L Vaughan, Brian L Davis, CO Jeremy, et al. “Dynamics of human gait”. In: (1999).

[9] Roger M Enoka. *Neuromechanical basis of kinesiology*. ERIC, 1988.

[10] HJ Ralston, V Inman, and F Todd. “Human walking”. In: *Baltimore: Williams and Wilkins* (1981).

[11] N Alberto Borghese, L Bianchi, and F Lacquaniti. “Kinematic determinants of human locomotion.” In: *The Journal of physiology* 494.3 (1996), pp. 863–879.

[12] Lewis M Nashner and Gin McCollum. “The organization of human postural movements: a formal basis and experimental synthesis”. In: *Behavioral and brain sciences* 8.1 (1985), pp. 135–150.

[13] V Dietz, J Quintern, and M Sillem. “Stumbling reactions in man: significance of proprioceptive and pre-programmed mechanisms.” In: *The Journal of Physiology* 386.1 (1987), pp. 149–163.

[14] T Pozzo, A Berthoz, and L Lefort. “Head stabilization during various locomotor tasks in humans”. In: *Experimental Brain Research* 82.1 (1990), pp. 97–106.

[15] Antonio Pedotti. “A study of motor coordination and neuromuscular activities in human locomotion”. In: *Biological cybernetics* 26.1 (1977), pp. 53–62.

[16] A Pedotti, VV Krishnan, and L Stark. “Optimization of muscle-force sequencing in human locomotion”. In: *Mathematical Biosciences* 38.1-2 (1978), pp. 57–76.

[17] X André-Thoma. “Equilibre et équilibration”. In: *Editeurs Masson & Cie, Paris* (1940).

[18] L Bianchi, D Angelini, and F Lacquaniti. “Individual characteristics of human walking mechanics”. In: *Pflügers Archiv* 436.3 (1998), pp. 343–356.

[19] Davrondzhon Gafurov and Einar Snekkenes. “Towards understanding the uniqueness of gait biometric”. In: *Automatic Face & Gesture Recognition, 2008. FG’08. 8th IEEE International Conference on.* IEEE. 2008, pp. 1–8.

[20] Davrondzhon Gafurov, Einar Snekkenes, and Patrick Bours. “Improved gait recognition performance using cycle matching”. In: *Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on.* IEEE. 2010, pp. 836–841.

[21] Davrondzhon Gafurov. “A survey of biometric gait recognition: Approaches, security and challenges”. In: *Annual Norwegian Computer Science Conference.* Citeseer. 2007, pp. 19–21.

[22] Zhaojun Xue et al. “Infrared gait recognition based on wavelet transform and support vector machine”. In: *Pattern recognition* 43.8 (2010), pp. 2904–2910.

[23] Joarder Kamruzzaman and Rezaul K Begg. “Support vector machines and other pattern recognition approaches to the diagnosis of cerebral palsy gait”. In: *IEEE Transactions on Biomedical Engineering* 53.12 (2006), pp. 2479–2490.

[24] Jochen Klucken et al. “Unbiased and mobile gait analysis detects motor impairment in Parkinson’s disease”. In: *PloS one* 8.2 (2013), e56956.

[25] Pankaj Wasnik et al. “Fusing biometric scores using subjective logic for gait recognition on smartphone”. In: *BIOSIG 2017* (2017).

[26] Rui Pu and Yunhong Wang. “2-D Structure-Based Gait Recognition in Video Using Incremental GMM-HMM”. In: *Computer Vision-ACCV 2014 Workshops.* Springer. 2014, pp. 58–70.

[27] Ju Han and Bir Bhanu. “Individual recognition using gait energy image”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (2006), pp. 316–322.

[28] Tracey KM Lee, Mohammed Belkhatir, and Saeid Sanei. “A comprehensive review of past and present vision-based techniques for gait recognition”. In: *Multimedia tools and applications* 72.3 (2014), pp. 2833–2869.

[29] Chandra Prakash, Rajesh Kumar, and Namita Mittal. “Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges”. In: *Artificial Intelligence Review* 49.1 (2018), pp. 1–40.

[30] V. Joseph Raj and S Balamurugan. “Survey of Current Trends in Human Gait Recognition Approaches”. In: *International Journal of Advanced Research* 5 (Oct. 2017), pp. 1851–1864.

[31] Suvarna Shirke, SS Pawar, and Kamal Shah. “Literature review: Model free human gait recognition”. In: *Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on*. IEEE. 2014, pp. 891–895.

[32] Robert J Orr and Gregory D Abowd. “The smart floor: a mechanism for natural user identification and tracking”. In: *CHI’00 extended abstracts on Human factors in computing systems*. ACM. 2000, pp. 275–276.

[33] Jaakko Suutala and Juha Röning. “Towards the adaptive identification of walkers: Automated feature selection of footsteps using distinction sensitive LVQ”. In: *Int. Workshop on Processing Sensory Information for Proactive Systems (PSIPS 2004)*. 2004, pp. 14–15.

[34] Lee Middleton et al. “A floor sensor system for gait recognition”. In: *Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on*. IEEE. 2005, pp. 171–176.

[35] Mohammad Omar Derawi. “Accelerometer-based gait analysis, a survey”. In: *Nor Informasjonssikkerhetskonferanse NISK* (2010).

[36] Alvaro Muro-De-La-Herran, Begonya Garcia-Zapirain, and Amaia Mendez-Zorrilla. “Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications”. In: *Sensors* 14.2 (2014), pp. 3362–3394.

[37] Jorge Blasco et al. “A survey of wearable biometric recognition systems”. In: *ACM Computing Surveys (CSUR)* 49.3 (2016), p. 43.

[38] Aakarsh Malhotra et al. “Fingerphoto Authentication Using Smartphone Camera Captured Under Varying Environmental Conditions”. In: *Human Recognition in Unconstrained Environments: Using Computer Vision, Pattern Recognition and Machine Learning Methods for Biometrics* (2017), pp. 119–144.

[39] Silvio Barra et al. “A Hand-based Biometric System in Visible Light for Mobile Environments”. In: *Information Sciences* (2018).

[40] Maria De Marsico et al. “Insights into the results of MICHE I-Mobile Iris CHallenge Evaluation”. In: *Pattern Recognition* 74 (2018), pp. 286–304.

[41] Maria De Marsico, Michele Nappi, and Hugo Proença. “Results from MICHE II–Mobile Iris CHallenge Evaluation II”. In: *Pattern Recognition Letters* 91 (2017), pp. 3–10.

[42] Maria De Marsico et al. “Firme: Face and iris recognition for mobile engagement”. In: *Image and Vision Computing* 32.12 (2014), pp. 1161–1172.

[43] Seyed Amir Hoseini-Tabatabaei, Alexander Gluhak, and Rahim Tafazolli. "A survey on smartphone-based systems for opportunistic user context recognition". In: *ACM Computing Surveys (CSUR)* 45.3 (2013), p. 27.

[44] Hiroaki Sakoe and Seibi Chiba. "Dynamic programming algorithm optimization for spoken word recognition". In: *IEEE transactions on acoustics, speech, and signal processing* 26.1 (1978), pp. 43–49.

[45] Sebastijan Sprager and Matjaz B Juric. "Inertial sensor-based gait recognition: a review". In: *Sensors* 15.9 (2015), pp. 22089–22127.

[46] Pavan Turaga et al. "Machine recognition of human activities: A survey". In: *IEEE Transactions on Circuits and Systems for Video technology* 18.11 (2008), pp. 1473–1488.

[47] Akin Avci et al. "Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey". In: *Architecture of computing systems (ARCS), 2010 23rd international conference on*. VDE. 2010, pp. 1–10.

[48] Ronald Poppe. "A survey on vision-based human action recognition". In: *Image and vision computing* 28.6 (2010), pp. 976–990.

[49] Chen Chen, Roozbeh Jafari, and Nasser Kehtarnavaz. "A survey of depth and inertial sensor fusion for human action recognition". In: *Multimedia Tools and Applications* 76.3 (2017), pp. 4405–4425.

[50] Analog Devices. *3-Axis,±2 g/±4 g/±8 g/±16g Digital Accelerometer*. http://www.analog.com/media/en/technical-documentation/data-sheets/ADXL345.pdf/. 2013.

[51] Mohamed Gad-el Hak. *The MEMS handbook*. CRC press, 2001.

[52] Maria De Marsico, Daniele De Pasquale, and Alessio Mecca. "Embedded Accelerometer Signal Normalization for Cross-Device Gait Recognition". In: *15th Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*. IEEE. 2016, pp. 1–5.

[53] R. Vera-Rodriguez et al. "Comparative Analysis and Fusion of Spatiotemporal Information for Footstep Recognition". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.4 (2013), pp. 823–834. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2012.164.

[54] Thanh Trung Ngo et al. "The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication". In: *Pattern Recognition* 47.1 (2014), pp. 228–237.

[55] Yuting Zhang et al. "Accelerometer-based gait recognition by sparse representation of signature points with clusters". In: *IEEE transactions on cybernetics* 45.9 (2015), pp. 1864–1875.

[56] Rob J Hyndman. "Moving averages". In: *International encyclopedia of statistical science*. Springer, 2011, pp. 866–869.

[57] Mohammad O Derawi et al. "Unobtrusive user-authentication on mobile phones using biometric gait recognition". In: *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on*. IEEE. 2010, pp. 306–311.

[58] Stéphane Mallat. *A wavelet tour of signal processing.* Academic press, 1999. Chap. 11.

[59] Liu Rong et al. "Identification of individual walking patterns using gait acceleration". In: *2007 1st International Conference on Bioinformatics and Biomedical Engineering.* IEEE. 2007, pp. 543–546.

[60] Liu Rong et al. "A wearable acceleration sensor system for gait recognition". In: *2007 2nd IEEE Conference on Industrial Electronics and Applications.* IEEE. 2007, pp. 2654–2659.

[61] Claudia Nickel et al. "Using hidden markov models for accelerometer-based biometric gait recognition". In: *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on.* IEEE. 2011, pp. 58–63.

[62] Claudia Nickel, Holger Brandt, and Christoph Busch. "Classification of Acceleration Data for Biometric Gait Recognition on Mobile Devices." In: *SIG* 11 (2011), pp. 57–66.

[63] Mohammad O Derawi, Patrick Bours, and Kjetil Holien. "Improved cycle detection for accelerometer based gait authentication". In: *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on.* IEEE. 2010, pp. 312–317.

[64] Claudia Nickel et al. "Scenario Test of Accelerometer-Based Biometric Gait Recognition". In: *Security and Communication Networks (IWSCN), 2011 Third International Workshop on.* IEEE. 2011, pp. 15–21.

[65] Claudia Nickel, Tobias Wirtl, and Christoph Busch. "Authentication of smartphone users based on the way they walk using k-NN algorithm". In: *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on.* IEEE. 2012, pp. 16–20.

[66] Felix Juefei-Xu et al. "Gait-id on the move: Pace independent human identification using cell phone accelerometer dynamics". In: *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on.* IEEE. 2012, pp. 8–15.

[67] Richard G Lyons. *Understanding Digital Signal Processing, 3/E.* Pearson Education India, 2011.

[68] Davrondzhon Gafurov, Kirsi Helkala, and Torkjel Søndrol. "Biometric gait authentication using accelerometer sensor". In: *Journal of computers* 1.7 (2006), pp. 51–59.

[69] Davrondzhon Gafurov, Kirsi Helkala, and Torkjel Søndrol. "Gait recognition using acceleration from MEMS". In: *First International Conference on Availability, Reliability and Security (ARES'06).* IEEE. 2006, 6–pp.

[70] Heikki J Ailisto et al. "Identifying people from gait pattern with accelerometers". In: *Defense and Security.* Int. Society for Optics and Photonics. 2005, pp. 7–14.

[71] Jani Mäntyjärvi et al. “Identifying users of portable devices from gait pattern with accelerometers”. In: *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on.* Vol. 2. IEEE. 2005, pp. ii–973.

[72] Ngo Thanh Trung et al. “Phase registration in a gallery improving gait authentication”. In: *Biometrics (IJCB), Int. Joint Conf. on.* IEEE. 2011, pp. 1–7.

[73] Yasushi Makihara et al. “Phase registration of a single quasi-periodic signal using self dynamic time warping”. In: *Asian Conference on Computer Vision.* Springer. 2010, pp. 667–678.

[74] Gang Pan, Ye Zhang, and Zhisheng Wu. “Accelerometer-based gait recognition via voting by signature points”. In: *Electronics letters* 45.22 (2009), pp. 1116–1118.

[75] Maria De Marsico and Alessio Mecca. “Biometric Walk Recognizer”. In: *International Conference on Image Analysis and Processing.* Springer. 2015, pp. 19–26.

[76] Maria De Marsico and Alessio Mecca. “Biometric Walk Recognizer - Gait recognition by a single smartphone accelerometer”. In: *Multimedia Tools and Applications* (2016). DOI: 10.1007/s11042-016-3654-1.

[77] Michael Fitzgerald Nowlan. “Human Identification via Gait Recognition Using Accelerometer Gyro Forces”. In: *Yale Computer Science. http://www. cs. yale. edu/homes/mfn3/pub/mfn_gait_id. pdf (accessed November 12, 2013)* (2009).

[78] Matteo Gadaleta and Michele Rossi. “Idnet: Smartphone-based gait recognition with convolutional neural networks”. In: *arXiv preprint arXiv:1606.03238* (2016).

[79] Giacomo Giorgi et al. “Try Walking in My Shoes, if You Can: Accurate Gait Recognition Through Deep Learning”. In: *International Conference on Computer Safety, Reliability, and Security.* Springer. 2017, pp. 384–395.

[80] Pablo Fernandez-Lopez et al. “Gait recognition using smartphone”. In: *Security Technology (ICCST), 2016 IEEE International Carnahan Conference on.* IEEE. 2016, pp. 1–7.

[81] Pavel Senin. “Dynamic time warping algorithm review”. In: *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA* 855 (2008), pp. 1–23.

[82] Rhonda Wilson. “Filter Topologies”. In: *Audio Engineering Society Conference: UK 7th Conference: Digital Signal Processing (DSP).* 1992. URL: http://www.aes.org/e-lib/browse.cfm?elib=6169.

[83] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. “Cell phone-based biometric identification”. In: *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on.* IEEE. 2010, pp. 1–7.

[84] Gary M Weiss and Haym Hirsh. “Learning to Predict Rare Events in Event Sequences.” In: *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining.* 1998, pp. 359–363.

[85] Hong Lu et al. “Unobtrusive gait verification for mobile phones”. In: *Proceedings of the 2014 ACM international symposium on wearable computers*. ACM. 2014, pp. 91–98.

[86] William J Hardcastle and Nigel Hewlett. *Coarticulation: Theory, data and techniques*. Cambridge Univ. Press, 2006.

[87] M. Günther et al. “Unconstrained Face Detection and Open-Set Face Recognition Challenge”. In: *2017 IEEE International Joint Conference on Biometrics (IJCB)*. 2017, pp. 697–706.

[88] Dayong Wang, Charles Otto, and Anil K Jain. “Face search at scale”. In: *IEEE transactions on pattern analysis and machine intelligence* 39.6 (2017), pp. 1122–1136.

[89] Daniele Riboni and Claudio Bettini. “Context-aware activity recognition through a combination of ontological and statistical reasoning”. In: *International Conference on Ubiquitous Intelligence and Computing*. Springer. 2009, pp. 39–53.

[90] Guglielmo Cola et al. “Gait-based authentication using a wrist-worn device”. In: *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. ACM. 2016, pp. 208–217.

[91] Andrew H Johnston and Gary M Weiss. “Smartwatch-based biometric gait recognition”. In: *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*. IEEE. 2015, pp. 1–6.

[92] Weitao Xu et al. “Gait-watch: A context-aware authentication system for smart watch based on gait recognition”. In: *Proceedings of the Second International Conference on Internet-of-Things Design and Implementation*. ACM. 2017, pp. 59–70.

[93] Yiran Shen et al. “Face recognition on smartphones via optimised sparse representation classification”. In: *Proceedings of the 13th international symposium on Information processing in sensor networks*. IEEE Press. 2014, pp. 237–248.

[94] Anshul Rai et al. “Zee: Zero-effort crowdsourcing for indoor localization”. In: *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM. 2012, pp. 293–304.

[95] Neamah Al-Naffakh. “A Comprehensive Evaluation of Feature Selection for Gait Recognition Using Smartwatches”. In: *International Journal for Information Security Research* 6 (3 2017), pp. 691–700.

[96] Yu Zhong and Yunbin Deng. “Sensor orientation invariant mobile gait biometrics”. In: *Biometrics (IJCB), 2014 IEEE International Joint Conference on*. IEEE. 2014, pp. 1–8.

[97] Trung Thanh Ngo et al. “Orientation-compensative signal registration for owner authentication using an accelerometer”. In: *IEICE TRANSACTIONS on Information and Systems* 97.3 (2014), pp. 541–553.

[98] Pablo Fernandez-Lopez et al. “Optimizing resources on smartphone gait recognition”. In: *Biometrics (IJCB), 2017 IEEE Int. Joint Conference on.* IEEE. 2017, pp. 31–36.

[99] Muhammad Muaaz and René Mayrhofer. “Smartphone-based gait recognition: from authentication to imitation”. In: *IEEE Transactions on Mobile Computing* 16.11 (2017), pp. 3209–3221.

[100] Brendan F Klare et al. “Face recognition performance: Role of demographic information”. In: *IEEE Transactions on Information Forensics and Security* 7.6 (2012), pp. 1789–1801.

[101] Maria De Marsico et al. “Demographics versus biometric automatic interoperability”. In: *International Conference on Image Analysis and Processing.* Springer. 2013, pp. 472–481.

[102] Maria De Marsico et al. “Leveraging implicit demographic information for face recognition using a multi-expert system”. In: *Multimedia Tools and Applications* 76.22 (2017), pp. 23383–23411.

[103] Ankita Jain and Vivek Kanhangad. “Gender classification in smartphones using gait information”. In: *Expert Systems with Applications* 93 (2018), pp. 257–266.

[104] Navneet Dalal and Bill Triggs. “Histograms of oriented gradients for human detection”. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.* Vol. 1. IEEE. 2005, pp. 886–893.

[105] Xianta Jiang et al. “A Wearable Gait Phase Detection System Based on Force Myography Techniques”. In: *Sensors* 18.4 (2018), p. 1279.

[106] Pallavi Meharia and Dharma P Agrawal. “The human key: Identification and authentication in wearable devices using gait”. In: *Journal of Information Privacy and Security* 11.2 (2015), pp. 80–96.

[107] Jeffrey M Hausdorff et al. “Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations”. In: *Journal of applied physiology* 80.5 (1996), pp. 1448–1457.

[108] Todd C Pataky et al. “Gait recognition: highly unique dynamic plantar pressure patterns among 104 individuals”. In: *Journal of The Royal Society Interface* 9.69 (2012), pp. 790–800.

[109] Takahiro Takeda et al. “Biometric personal authentication by one step foot pressure distribution change by load distribution sensor”. In: *Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on.* IEEE. 2009, pp. 906–910.

[110] Kuo-Hui Yeh et al. “I Walk, Therefore I Am: Continuous User Authentication with Plantar Biometrics”. In: *IEEE Communications Magazine* 56.2 (2018), pp. 150–157.

[111] Thiago Teixeira et al. “PEM-ID: Identifying people by gait-matching using cameras and wearable accelerometers”. In: *Distributed Smart Cameras, 2009. ICDSC 2009. Third ACM/IEEE International Conference on.* IEEE. 2009, pp. 1–8.

[112] Qin Zou et al. “Robust gait recognition by integrating inertial and RGBD sensors”. In: *IEEE transactions on cybernetics* 48.4 (2018), pp. 1136–1150.

[113] Elena Vildjiounaite et al. “Unobtrusive multimodal biometrics for ensuring privacy and information security with personal devices”. In: *International Conference on Pervasive Computing*. Springer. 2006, pp. 187–201.

[114] Davrondzhon Gafurov, Einar Snekkenes, and Tor Erik Buvarp. “Robustness of biometric gait authentication against impersonation attack”. In: *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer. 2006, pp. 479–488.

[115] Davrondzhon Gafurov, Einar Snekkenes, and Patrick Bours. “Spoof attacks on gait authentication system”. In: *IEEE Transactions on Information Forensics and Security* 2.3 (2007), pp. 491–502.

[116] Davrondzhon Gafurov. “Security analysis of impostor attempts with respect to gender in gait biometrics”. In: *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*. IEEE. 2007, pp. 1–6.

[117] Abdenour Hadid et al. “Can gait biometrics be spoofed?” In: *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE. 2012, pp. 3280–3283.

[118] Maria De Marsico and Alessio Mecca. “Gait Recognition: The Wearable Solution”. In: *Human Recognition in Unconstrained Environments: Using Computer Vision, Pattern Recognition and Machine Learning Methods for Biometrics* (2017), pp. 177–195.

[119] Kinh Tran, Tu Le, and Tien Dinh. “A high-accuracy step counting algorithm for iPhones using accelerometer”. In: *Signal Processing and Information Technology (ISSPIT), 2012 IEEE International Symposium on*. IEEE. 2012, pp. 000213–000217.

[120] De Marsico, Maria and Alessio Mecca. “Benefits of Gaussian Convolution in Gait Recognition”. In: *17th Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*. in-press, 2018.

[121] Alessio Mecca. “Impact of Gait Stabilization: a Study on How to Exploit it for User Recognition”. In: *SITIS 2018 - International Conference on Signal Image Technology and Internet Based Systems*. IEEE, 2018.

[122] Maria De Marsico, Eduard Gabriel Fartade, and Alessio Mecca. “Feature-based Analysis of Gait Signals for Biometric Recognition”. In: *ICPRAM 2018 - 7th International Conference on Pattern Recognition Applications and Methods*. SCITEPRESS, 2018, pp. 630–637.

[123] Yijuan Lu et al. “Feature selection using principal feature analysis”. In: *Proceedings of the 15th ACM international conference on Multimedia*. ACM. 2007, pp. 301–304.

[124] Ian Jolliffe. “Principal component analysis”. In: *International encyclopedia of statistical science*. Springer, 2011, pp. 1094–1096.

[125] Kamen Kanev et al. “Mobiles and Wearables: Owner Biometrics and Authentication”. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM. 2016, pp. 318–319.

[126] Kamen Kanev, Maria De Marsico, Paolo Bottoni, et al. “A human computer interactions framework for biometric user identification”. In: *JJAP Conference Proceedings*. Vol. 4. The Japan Society of Applied Physics. 2015.

[127] Aniello Castiglione et al. “Walking on the Cloud: Gait Recognition, a Wearable Solution”. In: *Network and System Security, 2018 12th International Conference on*. in-press, 2018.