THÈSE DE DOCTORAT DE L'ÉTABLISSEMENT UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ

PRÉPARÉE À L'UNIVERSITÉ DE BOURGOGNE

École doctorale n°37

Sciences Pour l'Ingénieur et Microtechniques

Doctorat d'Informatique

par

MARGARITA KHOKHLOVA

Évaluation clinique de la démarche à partir de données 3D

Thèse présentée et soutenue à Dijon, le 19 Novembre 2018

Composition du Jury :

Franck MARZANI	Professeur à l'Université de Bourgogne Franche-Comté, Le2i	Examinateur
Laurent MASCARILLA	Professeur à l'Université de La Rochelle	Rapporteur
Franck MULTON	Professeur à l'Université l'université de Rennes Chef de l'équipe conjointe Inria MimeTIC	Rapporteur
Frédéric LERASLE	Professeur à l'université Paul Sabatier, LAAS-CNRS	Examinateur
Saïda BOUAKAZ	Professeur à Université Claude Bernard Lyon 1	Examinateur
Briac COLOBERT	Responsable Recherche et Developpement chez Proteor	Invité
Albert DIPANDA	Professeur à l'Université de Bourgogne Franche-Comté	Directeur de thèse
Cyrille MIGNIOT	Maître de Conférences à l'Université de Bourgogne Franche-Comté, Le2i	Co-encadrant de thèse

N° | X | X | X |

SPIM
école doctorale sciences pour l'ingénieur et microtechniques

Title: Évaluation clinique de la démarche à partir de données 3D

Keywords: 3D descriptors, Gait Analysis, Abnormal Gait Dectection

Abstract:
Clinical Gait analysis is traditionally subjective, being performed by clinicians observing patients gait. A common alternative to such analysis is markers-based systems and ground-force platforms based systems. However, this standard gait analysis requires specialized locomotion laboratories, expensive equipment, and lengthy setup and post-processing times. Researchers made numerous attempts to propose a computer vision based alternative for clinical gait analysis. With the appearance of commercial 3D cameras, the problem of qualitative gait assessment was reviewed. Researchers realized the potential of depth-sensing devices for motion analysis applications. However, despite much encouraging progress in 3D sensing technologies, their real use in clinical application remains scarce.
In this dissertation, we develop models and techniques for movement assessment using a Microsoft Kinect sensor. In particular, we study the possibility to use different data provided by an RGBD camera for motion and posture analysis. The main contributions of this dissertation are the following. First, we executed a literature study to estimate the important gait parameters, the feasibility of different possible technical solutions and existing gait assessment methods. Second, we propose a 3D point cloud based posture descriptor. The designed descriptor can classify static human postures based on 3D data without the use of skeletonization algorithms. Third, we build an acquisition system to be used for gait analysis based on the Kinect v2 sensor. Fourth, we propose an abnormal gait detection approach based on the skeleton data. We demonstrate that our gait analysis tool works well on a collection of custom data and existing benchmarks. We show that our gait assessment approach advances the progress in the field, is ready to be used for gait assessment scenario and requires a minimum of the equipment.

Titre : Évaluation clinique de la démarche à partir de données 3D

Mots-clés : Descripteur 3D, Analyse de la démarche, Détection de la marche anormale

Résumé :
L'analyse de la démarche clinique est généralement subjective, étant effectuée par des cliniciens observant la démarche des patients. Des alternatives à une telle analyse sont les systèmes basés sur les marqueurs et les systèmes basés sur les plates-formes au sol. Cependant, cette analyse standard de la marche nécessite des laboratoires spécialisés, des équipements coûteux et de longs délais d'installation et de post-traitement. Il y a eu de nombreuses tentatives dans la recherche pour proposer une alternative basée sur la vision par ordinateur pour l'analyse de la demarche. Avec l'apparition de caméras 3D bon marché, le problème de l'évaluation qualitative de la démarche a été re-examiné. Les chercheurs ont réalisé le potentiel des dispositifs de caméras 3D pour les applications d'analyse de mouvement. Cependant, malgré des progrès très encourageants dans les technologies de détection 3D, leur utilisation réelle dans l'application clinique reste rare.
Cette thèse propose des modèles et des techniques pour l'évaluation du mouvement à l'aide d'un capteur Microsoft Kinect. En particulier, nous étudions la possibilité d'utiliser différentes données fournies par une caméra RGBD pour l'analyse du mouvement et de la posture. Les principales contributions sont les suivantes. Nous avons réalisé une étude de l'état de l'art pour estimer les paramètres importants de la démarche, la faisabilité de différentes solutions techniques et les méthodes d'évaluation de la démarche existantes. Ensuite, nous proposons un descripteur de posture basé sur un nuage de points 3D. Le descripteur conçu peut classer les postures humaines statiques a partir des données 3D. Nous construisons un système d'acquisition pour l'analyse de la marche basée sur les données acquises par un capteur Kinect v2. Enfin, nous proposons une approche de détection de démarche anormale basée sur les données du squelette. Nous démontrons que notre outil d'analyse de la marche fonctionne bien sur une collection de données que nous avons générée ainsi que sur des données de l'état de l'art. Notre méthode d'évaluation de la démarche montre des avancées significatives dans le domaine. Notre approche nécessite un équipement limité et est prête à être utilisée pour l'évaluation de la démarche en conditions réelles.

UBFC UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ
Université Bourgogne Franche-Comté
32, avenue de l'Observatoire
25000 Besanon, France

ACKNOWLEDGEMENTS

First of all, I want to thank the Region of Burgundy who financed this JCE thesis and company Proteor who gave us clinical insights and necessary equipment to run our experiments, and especially Briac Colobert and Vincent Carre.

There are many people I must thank for contributing to the three wonderful years of my experience as a PhD student. This was a unique experience and I am very happy to have this three years in my life, which would have never be the same without the people around me. I want to thank my PhD advisor Albert Dipanda, who was always able to find some time for me even being extremely busy with his role of a director of ESIREM. I want to thank my second Advisor, Cyrille Migniot, with whom we worked closely during these three years, for his patience and ability to listen to all my ideas and give his insights and advises. I express my gratitude to Mathieu Gueugnon for the possibility perform some experiments at CHU Dijon.

Finally, I want to thank the people who have been with me during this work or encouraged me to take it and accept the challenge, to change my life and to go and study abroad: my parents Nikolay and Valentina, my sister Kate, Sergey, Ludwig Lenart, Matthieu, Marie, Zara, Ali, Ani, Daria, Lesha, Polina, Nastya, Anya, Polina, Marie and Masha.

I will always be grateful to Maurits Diephuis for his support and time.

It was great to share the PhD experience with fellow researchers in my last year: Axel, Romain, Pierre, Richard, Yoan, Serge, Alex, Anthony, Fan, El-Bay, Sergey. I have to thank many people who have contributed to my day-to-day life in Dijon: Kais, Camille, Nejmi, Haris, Jessica, Faustine and Stephanie.

UNION EUROPEENNE

RÉGION BOURGOGNE FRANCHE COMTÉ

avec le Fonds européen de développement régional (FEDER)

Contents

1

Introduction

Human movement analysis is the observation and definition of body locomotion. From the early days of the emergence of Computer Vision, human motion has been studied. Firstly with 2D data and later extended to three dimensions. The 3D data can be provided by depth cameras, which capture the structure of the scene and deliver so-called Point Clouds: the 3D coordinates of the scene points. Human movement analysis is currently one of the actively researched application domains in Computer Vision.

This thesis is a continuation of the RaHa project, in which previously the following studies were performed. First, the enhancing filters for point cloud data were constructed [96]. Second, the 3D flow computational method was developed [95]. Third, the acquisition and segmentation of static point clouds were performed [174] along with the human detection.

This thesis is dedicated to human motion analysis using 3D sensors. The work is performed in collaboration with the company Proteor located in Dijon. Proteor develops two main activities: custom-made orthopedic fittings for patients and the design, manufacture and sales of components for prostheses, limb orthoses and spinal orthoses. Due to their professional activity, Proteor regularly performs analysis of patients' walks. We seek to study the possibilities to perform automatic gait analysis for Proteor patients with a prosthesis. Ideally, a prosthetic would behave in such a way that the user would feel like their limb was never disabled, and his gait should be close to normal. We pursue a qualitative gait analysis measure search. Traditionally, Proteor uses semi-subjective and ground-force platform based gait evaluation methods. The current work is the first step in the exploratory research in the direction of computer vision based pathological gait classification for our laboratory and Proteor. The latter provides clinical insight for the targeted problem, whilst the Le2i laboratory has substantial experience in the 3D vision domain.

Contents

In this thesis we evaluated the latest 3D descriptors [205, 249], built an acquisition system based on a Kinect v.2 sensor, proposed a new non-skeleton based 3D human posture descriptor [248] and a skeleton based gait unsupervised classification machine Learning model.

1.1/ Technology and Methodology used in Gait Analysis Studies

Human gait is the subject of many prior studies, advocating a wide array of hard and software solutions. We present an outlook on the different existing hardware setups and gait protocols, in order to familiarize the reader with the research subject.

1.1.1/ Hardware

The traditional measures used to analyze gait parameters in clinical conditions are semi-subjective. Medical specialists observe the quality of a patient's gait by making him/her walk, sometimes a specific distance or performing specific actions, such as standing up from a chair and walking or walking with an 8-shaped trajectory et cetera. More information about the different deployed walking tests is given in section 1.1.2 later in this work. Exercises are sometimes followed by a survey, in which the patient is asked to give a subjective evaluation of the quality of his/her gait. The disadvantage of these methods is the fact that they give subjective measurements, particularly concerning accuracy and precision, which has a negative effect on the diagnosis, follow-up, and treatment of the pathologies. The subjective nature of the traditional evaluation affects the accuracy, exactitude, repeatability, and reproducibility of the measurements.

Progress in technology has led to the development of a series of devices and techniques, which allow for objective gait evaluation, making measurements more efficient and effective, and providing specialists with reliable information. Gait analysis is widely accepted as a useful research tool but is yet to be automatized. Automatic gait analysis will allow for a more accurate assessment of gait deviations, than visual gait assessment [92].

The main approaches for gait analysis from the hardware point of view could be classified as machine vision based Non-Wearable Sensors (NWS), Floor Sensor-based and Wearable Sensor-based (WS). There are also solutions combining one or several approaches available. Conventional and RGB-D cameras correspond to the first category of methods. NWS takes objective measurements of the different parameters through digital image processing. These groups of methods allow estimating many important spatio-local and kinematic parameters of the gait. 2D camera-based systems commonly require the use of controlled research facilities, where sensors are statically positioned and capture gait data while the subject walks on a clearly marked walkway. Typically, the 2D camera gives only limited scene structure information. However, there are different methods to obtain the shape information from 2D images and video sequences, such as shape from shading and shape from motion. A multi-camera setup can be used to calculate the scene point clouds using the principle of stereo vision. Today, the 3D models obtained with multi-camera setups tend to give accurate human motion tracking results. Using the 3D motion analysis allows for analyzing more parameters than in 2D approaches. However, the cameras should be statically placed and calibrated, which limits the acquisition

volume and flexibility of the system.

3D cameras or depth cameras deliver images showing the distance to points in a scene from a specific origin. Usually, such cameras belong to the active sensing category and have an emitter of an IR light projecting a pattern on the scene and a sensor capturing the reflected IR light. An example of such an active sensing device is shown in Figure 1.1 a. 3D camera-based systems have fewer restrictions for the acquisition setup, provide real-valued data, conserve the object dimensions, and allow an easy segmentation of the patient from the background, but commonly have suffer from a small acquisition volume size due to their limited distance range. Nevertheless, a single depth camera can provide 3D data in real-time. Some examples of 3D optic sensors are laser range scanners (LRS), infrared sensors and Time-of-Flight (TOF) cameras.

Floor Sensor-based methods [185] use special platforms, installed on the floor, where gait information is measured through pressure sensors and ground reaction force sensors (GRF), which measure the force exerted by the subject's feet on the floor when he/she walks. The main problem of these systems is their limited size, making it impossible to collect much data successively from the same patient. The second problem is the impossibility to obtain kinematic gait parameters directly. Third, such sensors are expensive. An example of a GRF is shown in Figure 1.1 b.

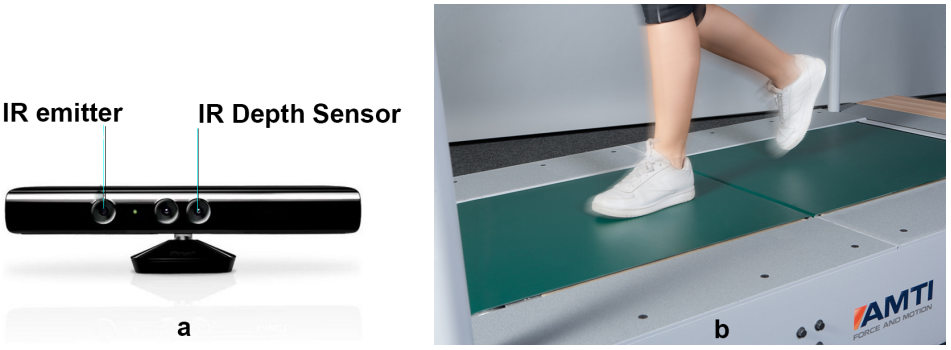


Figure 1.1: a) Depth camera Kinect with IR projector and IR sensor. b) A typical Ground Force plates setup. Image rights: GP Musculoskeletal System Modeling Lab, Connecticut.

WS are the systems which employ one or multiple sensors attached to a patient's body that measure an object's velocity, acceleration and gravitational forces using a combination of accelerometers, gyroscopes, strain gauges, inclinometers, goniometers, magnetometers et cetera. WS can also be a combination of IR markers and 3D cameras, which track the markers positions in the acquisition volume. In contrast to NWS and Floor Sensor-based systems, WS systems (except for marker+camera based setups) make it possible to analyze data outside the laboratory and capture information on the human gait during a person's everyday activities [150]. This group of sensors is very interesting and potentially affordable. At the moment, some mobile applications successfully track human activity with the use of the build-in sensors. However, there are still few commercialized models available which will be able to provided clinically relevant data. Further research is necessary to standardize the methods for defining kinetic variables to develop a more reliable process of analyzing gait in a clinical setting [109].

Depending on the type of the sensor, different groups of gait parameters can be obtained. Currently, the golden standard methods for gait assessment in clinical settings

are markers-cameras based motion capture systems such as Vicon (Oxford, UK) and Optotrac (NDI, Optotrak Certus, Waterloo, Canada). Figure 1.2 illustrates a typical motion capture system design. The number of cameras and other elements might vary, which obviously influences the attainable precision.

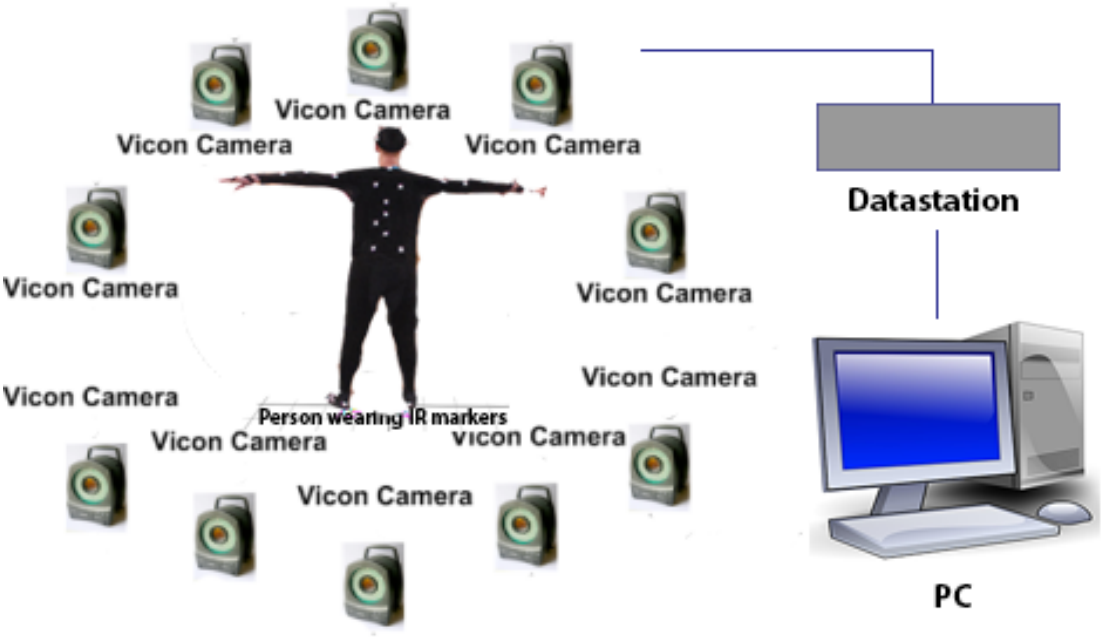


Figure 1.2: A typical motion capture setup composed of cameras tracking the IR markers attached to a patient's body.

Both solutions consist of several cameras viewing a certain volume, markers to place on the patient's body and a motion software application to recuperate the data. They give quite accurate motion estimation [238]. However, they carry a significant financial burden for clinics and hospitals. In addition, the setup of such systems for each patient is time costly[1]. These systems are often paired with a GRF sensor, which makes the overall acquisition setup even more expensive.

1.1.2/ Walking Tests

Commonly, the observed patient just walks freely in one direction, while the experienced clinical specialist evaluates his/her gait. For reliable gait data collection and evaluation, a sufficient amount of steps should be performed by a patient. Alternatively, there are many walking test protocols described in literature. Researchers used different walking distances from 5 to 82 meters [189]. In the case when the acquisition space is limited, a treadmill might be used.

There are several tests extensively used for gait evaluation. We studied the possibility to adapt an existing one for our gait assessment tool. The tests targeting a general walking scenario, i.e. not particular disease-oriented, are considered. The most used tests are:

[1] According to the clinical staff of the CHU Dijon, the time to place Vicon markers and calibrate the system takes at least 15 minutes per patient.

- One of the widely used for mobility assessments in physical therapy is the Timed-up and Go (TUG) test [19]. The test consists of 5 components: stand up, walk 3 meters, turn around, walk 3 meters back towards a chair, and sit down.
- The "Figure of 8 Walk" (F8W) is another technique to assess walking skills of older adults [69]. A typical F8W walking path consists of both straight and curved paths, designed to represent everyday walking skills.
- The most popular amongst all the tests is probably the standard clinical 10MWT assessment. Ten meters is enough to collect a sufficient number of gait cycles (repetitive part of the gait) of a patient for the following analysis [175].
- 25-Foot Walk (T25-FW) [36] is a clinical tool that evaluates patients for quantitative mobility and leg function performance in a timed, 25-foot walk.
- Tinetti Performance-Oriented Mobility Assessment is a test consisting of several tasks such as waking, standing up from a chair, turning to 360 degrees etc. Tinetti Performance-Oriented Mobility Assessment doesn't impose the walking distance but assumes that it is big enough to capture gait parameters. In the instructions, a patient is asked to walk down the hallway and back. The researchers propose the task-oriented outcome measure that assesses gait and balance ability; it is composed of a 9-item gait portion (POMA-G) and 7-item balance portion (POMA-B)[14].

For indoor studies using depth sensors, the short walking path or a treadmill suit the most. The range of most modern consumer RGB-D cameras is limited to 5 meters. Therefore, unless using a multi-device setup, only the TUG [19] test is applicable among the widely practiced walking tests. However, the fact that the patients have to stand up and sit down, next to walking a small distance, does no give a sufficient impression of a patients natural walking abilities.

Taking into account the limits of the acquisition volume, some researchers envisage the possibility to not use a standard clinical test, but a treadmill instead, to capture more gait cycles of patients using a small setup space. In the next subsection, a literature study is presented in order to validate the possibility to use a treadmill for gait assessment task.

1.1.3/ A TREADMILL FOR GAIT ANALYSIS

Traditionally, gait analysis data was collected during normal (overground) walking using wearable or non-wearable sensors. The distance is usually not a problem for WS, except for markers and camera-based systems. However, they provide limited information on the gait. NWS are commonly used in lab conditions, where space is limited due to the environment and the specifics of sensors used.

Using an instrumented treadmill for clinical gait analysis may help to overcome the limitations of overground testing. Thanks to the treadmill, consecutive cycles with gait data can be recorded in a short period of time and in a limited space, thus increasing data collection efficiency and reducing the cost of the analysis.

The other advantages are the possibility to collect multiple steps in controlled conditions, which increases data reliability and allows for variability analysis over time. The benefits

of deploying a treadmill not withstanding, research thus indicate that it may influence the natural gait of a patient, making its application anything but a panacea.

A recent paper [239] provides a summary of the differences in the estimated parameters with and without treadmill. Researchers came to the conclusion that the bio-mechanics of walking on a treadmill versus overground are comparable, with the majority of minor differences appearing in the kinetic measures. It was reported that people walk with a higher cadence, shorter stance time and reduced preferred walking speed on a treadmill, compared to overground walking [207]. Furthermore, treadmill walking artificially reduces natural variability and complexity, thereby creating a more stable and predictable gait pattern [31].

Kinematic and kinetic patterns are very similar in general, between treadmill and overground walking, only exhibiting some differences in amplitude [51]. For instance, Riley et al. calculated the coefficient of repeatability for healthy, able-bodied individuals walking overground and on a treadmill and found that, in all cases, the mean kinematic differences between overground and treadmill were less than this coefficient. The only difference in kinetics that exceeded it were the knee extension moment, the anterior-posterior maximum and the mediallateral minimum ground reaction forces. Researchers [51][239] conclude that differences between a treadmill and overground walking are small and typically not clinically relevant

Some research was done to examine the feasibility of treadmill usage for patients with abnormal gait, such as children with Cerebral Palsy [190], stroke patients [83] and transtibial amputees [102]. Authors state that there are systematic differences between treadmill and overground walking. However, in general, gait deviations seem to be more pronounced in treadmill walking. In particular, subjects with unpaired controls decreased step length. Patients' step width variability has increased. Transverse and frontal plane range of motion of the pelvis and trunk were decreased. [102].

Summarizing, treadmill walking in gait analysis can be considered a valid method to detect motor control deficits. The main thing to take into consideration is the fact that it affects kinematic parameters such as knee flexion, flat foot contact, asymmetry of stance et cetera, slightly, so the evaluation procedures should be adapted accordingly.

Overall, the use of a treadmill allows to create a gait analysis tool in a controlled environment within a small capture volume. It gives real-time feedback on gait parameters, so the ability of patients to adapt their gait can be assessed and provides information on compensatory mechanisms. The use of a treadmill thus expands the possibilities of gait analysis, allowing for a more functional and, likely, more sensitive assessment.

We considered the recommendations given by [239]. In particular, a self-paced mode can be used, in which the speed of the treadmill is controlled automatically by the walking speed of the subject to allow for natural walking speed variations. Most people are not as familiar with treadmill walking as with overground walking. Therefore, a major limitation of most of the above-mentioned studies is that they included only a short period (two minutes) of familiarization with the treadmill. It has been shown that after six minutes of treadmill walking, spatiotemporal parameters and knee kinematics are no longer different from overground walking [28]. However, a very long preparation time can be tiresome for some patients, so it can be adjusted accordingly.

Overall, the treadmill is a feasible solution for clinical gait analysis, which allows for maximum data acquisition in a minimum volume. It can be used for the gait analysis tool

proposed in this work.

1.1.4/ Conclusion on Technology and Methodology used in Gait Studies

For clinical analysis, the reliability and accessibility of the obtained data are the most critical factors.

There is a great need for a low-cost reliable subjective gait assessment tool and 3D sensors are very well suited for this goal, as they are affordable and provide real-time 3D data about gait. They have the potential to give 3D information about human motion, simplify the segmentation step and do not require a long setup time and calibration. 3D sensors have a potential to deliver the same data as the widely adopted motion capture systems at a significantly reduced cost and time burden.

There is a variety of TOF IR 3D sensors presented on the market, such as Microsoft Kinect v.2, Creative Senz3D by Intel, Basler cameras and others. For automatic gait assessment, a setup with a 3D sensor and a treadmill has great potential. It allows to perform the acquisition of long gait sequences that are similar enough to real-life conditions. Lastly, it negates the need for a large acquisition volume or a specially tailored clinical space.

1.2/ Problematic Statement

Gait is a manner of walking on a solid substrate. Observation of gait can provide early diagnostic clues for a number of movement disorders. Nowadays, orthopedic hospitals and clinics visually evaluate patients movements and gait in order to diagnose pathologies, design surgical operations and plan treatments for individuals with conditions affecting their ability to walk. Automatic gait calssification and analysis may enhance clinical practice.

Our partner in this project is the company Proteor [8], which specializes in developing custom-made orthopedic fitting for patients. The company seeks an automatic gait analysis tool, which will classify normal and abnormal gait and locate the source of pathology. The requirements for such a tool are summarized in Figure 1.3. The role of the company in this work was the following. We conducted numerous meetings with Proteor to discuss the research direction and obtain the insights from the clinical point of view. Part of the hardware used in our experiments was provided by the Proteor. We also performed several test acquisitions in their premises while working on the system setup.

Human movement capture in 3D is a field of active research. Not surprisingly, many methods from 2D motion tracking are extended to 3D approaches. 3D motion capture could be performed via a multi-camera setup or active depth-sensor technology, with the use of special body markers. There are several advantages of 3D over 2D, which makes 3D information so valuable for a wide range of applications. It gives the opportunity to accurately estimate the exact position and orientation of the object relative to the sensor. 3D object recognition tends to be more robust, when objects are occluded. 3D data allows for an easy segmentation of the scene and allows to evaluate the 3D human movement. Lastly, 3D information also helps in cases when it is desired to prevent the collision or

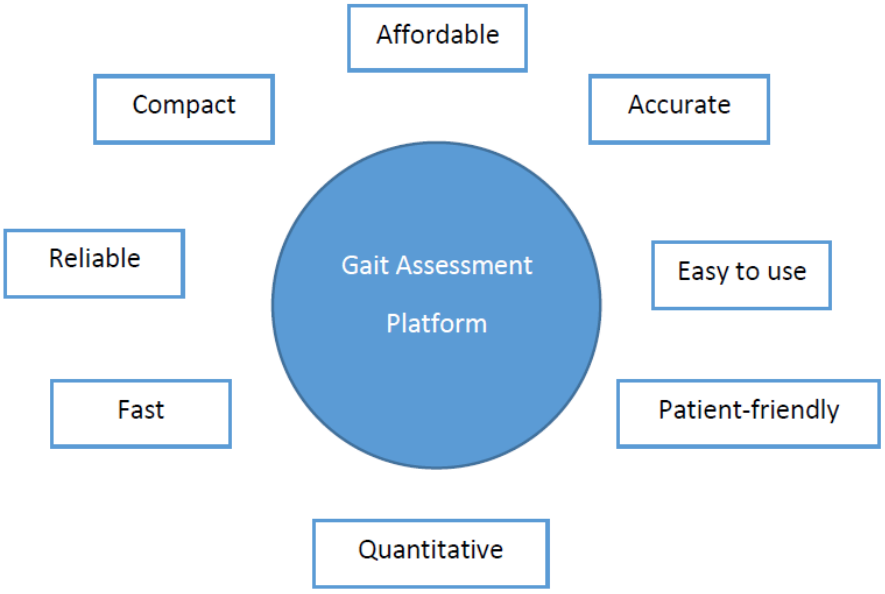


Figure 1.3: The desired qualities of a gait assessment platform.

similar tasks, better representing the real world.

This project aims to use an RGB-D camera based acquisition setup for 3D data human gait classification and analysis. Such a camera captures the structure of the scene, using an active depth sensor, which measures the distance to the object. Using the resulting depth image and camera calibration data, the 3D coordinates of the scene points, commonly referred to as a point cloud, may be obtained.

Figure 1.4: The complete point cloud of a scene where a subject walks on a treadmill. Bright green points show the estimated skeleton joint position.

Fig. 1.4 shows a colored point cloud of a subject walking on a treadmill acquired on the

Proteor premises. Based on the depth data, the skeleton representation of the human figure could be distilled. Since the appearances of affordable 3D sensors such as the Kinect, researchers have started to intensively use them for various applications related to human motion analysis, including gait analysis.

Modern RGB-D cameras may provide a low-cost gait analysis method in clinical settings. An automatic motion analysis tool will help to standardize the monitoring procedure and provide the numerical data for later analysis.

Further in this work we will concentrate on a more specific analysis of the available 3D sensors and build a gait acquisition system based on them.

Figure 1.5 shows the main difficulties to consider, questions to answer and decisions to be made in order to complete the thesis work.

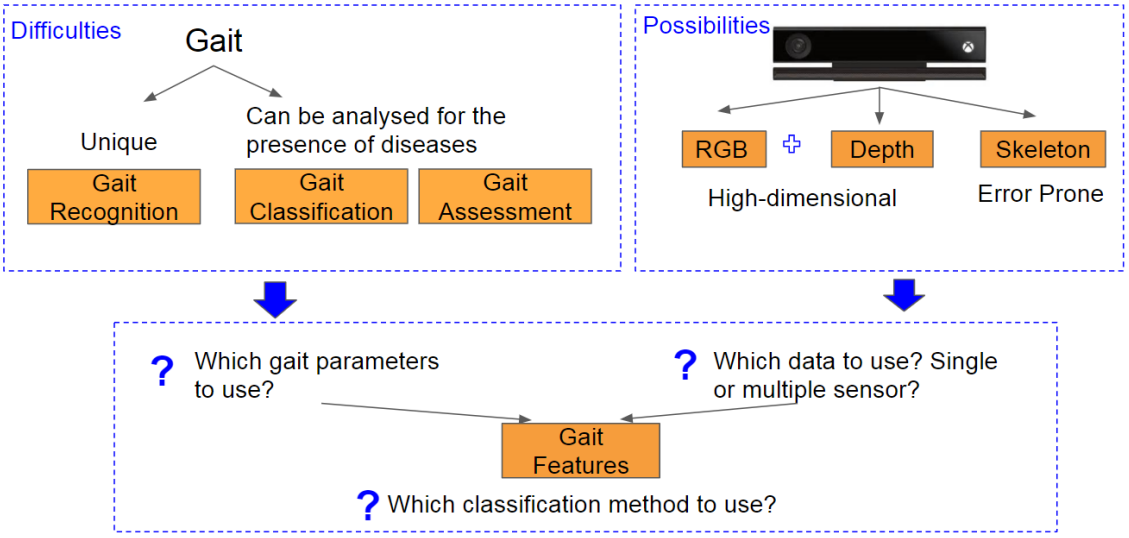


Figure 1.5: The overview of the thesis problematic on a block scheme and main questions to be answered in this work.

The goal of this Ph.D. thesis is to obtain in-depth knowledge of human gait mechanisms and functions. We seek a new tool for clinical gait assessment, which will allow to automatically extract all the parameters of the gait and perform classification of normal and abnormal gait patterns.

The next research question is to assign qualitative meaning to the obtained gait information. The desired method should give an objective gait evaluation for patients with a low-limb prosthesis, starting with a gait pathologies classification. In particularly, fully-automated 3D sensor based approaches will be researched and prototyped.

This PhD tackles the following sub-problems:

- A state-of-art research in 3D motion descriptors and gait analysis.
- The comparison between golden-standard motion capture systems and consumer RGB-D cameras in the context of gait analysis.
- The acquisition of the custom gait database.
- The design of human posture descriptors based on 3D point cloud data.

- Development of unsupervised human gait classification algorithms, based on skeleton data.

1.3/ THESIS PLAN

This thesis is composed of four major chapters. The literature review is performed in Chapter 2. Section 2.1 presents the gait parameters used in clinical studies, explains the periodic nature of gait. Gait cycles are explained in 2.1.1, while gait parameters are reviewed in 2.1.2. We provide an overview of what pathological gait is, in 2.1.3. Part 2.1.5 introduces the available gait datasets. Finally, the section is summarized in 2.4.

Section 2.2 is dedicated to 3D motion descriptors and is organized as follows. Section 2.2.2 presents the requirements and evaluation criteria for motion descriptors. First, we talk about classifying methods and then we introduce the most popular 3D datasets for motion descriptor evaluation. Further in parts 2.2.3, 2.2.4 and 2.2.5, we talk about the common trends and representative methods to describe the motion for three applications respectively: action recognition, gesture recognition and gait analysis. We summarize the main approaches and point out their interesting and novel parts. Section 2.2.6 concludes the section 3D motion descriptors review, where we highlight the main existing trends in motion description for point cloud sequences and propose our outlook for the future.

Next section 2.3 presents skeleton-based gait descriptors and three gait-related applications which commonly use them: Binary Gait Classification 2.3.1, Qualitative Gait Assessment 2.3.2 and Gait Recognition 2.3.3. The conclusions made for skeleton-based gait descriptors are summarized in part 2.3.4. Finally, the Conclusion for the Chapter 2 is given in 2.4.

Chapter 3 is dedicated to the search for an optimal 3D sensor-based gait acquisition platform. Here we look into detail at the available RGB-D cameras (part 3.1.1) in general and Kinect in particular (part 3.1.2). Section 3.1.3 introduces the skeleton. It is a popular human figure representation determined from a depth image of an RGB-D sensor. Then we overview the studies analyzing the performance of Kinect sensors in different computer vision applications 3.1.4. The possibility to use a multi-RGB-D setup is introduced in 3.1.5. Section 3.2 describes our contribution to the assessment of the validity of the Kinect v.2 for kinematic gait parameters estimation.

Section 3.3 describes the experiments performed with single 3.3.1 and multi-Kinect 3.3.2 acquisition setups. Finally, section 3.4 concludes the chapter.

Chapter 4 describes a novel human posture descriptor for point clouds. Section 4.1 presents the posture estimation problematic. Section 4.2 overviews existing methods for human posture recognition. Section 4.3 introduces the proposed descriptor and its parameters. Section 4.4 describes the data used in the experiments presented in section 4.5. Section 4.6 summarizes the results, proposes possible applications and outlines future work.

Chapter 5 is dedicated to gait assessment based on skeleton joint data. First in 5.1 we review the gait assessment task, highlighting the limitations of the earlier methods. A database acquired in our laboratory is described in 5.2. Section 5.3 presents a new covariance-based binary classification method for normal and abnormal gait. The novel descriptor designed is presented in 5.3.1. Experiments to validate our approach are

described in part 5.3.2. Then section 5.4 presents Long Short Term Memory networks (LSTM) and describes the selected gait model trained to classify gait in our new database. The specific glsen.LSTM cell structure and use cases are briefly presented in 5.4.1. The architecture developed for our experiment is described in part 5.4.2. Experiments and results for our proposed gait assessment method can be found in 5.4.3 and 5.4.4. Finally, Section 5.5 concludes the chapter, where we highlight the outcomes of the skeleton-based methods study and discuss future research.

Lastly, the thesis conclusions 6 finalize this work.

1.4/ CONTRIBUTIONS

The work on the current thesis is summarized by several scientific articles [249, 205, 248]. [249, 205] propose the analysis and review of the existing 3D motion descriptors. [248] presents the 3D posture recognition method, and the code is available in https://github.com/margokhokhlova/Occupancy_Decriptor. The final article summarizing the results of this thesis was submitted recently to a Elsevier journal "Artificial Intelligence in Medicine", and covariance descriptor from chapter 5 was described in the accepted paper for SITIS 2018. The final LSTM-based gait model is available for the reuse in https://github.com/margokhokhlova/LSTM_gait_model.

2

LITERATURE STUDY

This chapter provides the literature review for different aspects studied during this PhD thesis. This PhD is an exploratory work, hence we envisage different hardware solutions, acquisition setups, gait assessments tests and parameters.

First of all, we gather the information about the gait, normal and abnormal gait patterns and parameters in section 2.1.

We study the possibilities to use the data of different modalities from a 3D sensor, therefore we engage a literature study of the existing methods in two different directions: the 3D based motion descriptors and skeleton based descriptors.

3D movement based gait descriptors are not widely used, so we enlarge the literature study towards the action and gesture recognition domains. Section 2.2 is dedicated to 3D motion description. Skeleton based descriptors were mostly reviewed in the context of gait assessment and biometrics extraction. Skeleton-based gait research findings are summarized in section 2.3.

Contents

2.1/ GAIT PARAMETERS

In the last decades, interest in obtaining in-depth knowledge of human gait mechanisms and functions has increased dramatically. Gait is a manner of walking on a solid substrate. Walking is most basic method of transportation, but an inability to walk will drastically change a person's life. A reducing mobility pathology can create significant health problems over both short and long term. Observation of gait can provide early diagnostic clues for a number of movement disorders [2] such as Parkinson's disease, cerebral palsy, stroke, arthritis, chronic obstructive pulmonary disease and many others. It is stated in early medical studies that the gait is unique, if the twenty-four different components of human gait are considered [3].

Depending on the field of research, different gait parameters are evaluated. Broadly, the parameters of the gait could be divided into kinematic ones, such as knee or hip flexing angle and spatio-temporal ones, such as step length, step width, walking speed, cycle time et cetera. Parameters of the gait can be used for analysis in different scenarios. Figure 2.1 demonstrates various applications using gait parameters. One of such applications is clinical gait analysis and classification (i.e detecting a specific pathology).

The difficulty for clinical gait analysis is the fact that parameters of the gait are connected to the height, weight, general health and age of the person and also depend on the walking speed. Other factors are culture, motivation, prior sports activity and efforts. Moreover, some subtle differences in gait could be due to the person's mood or whether or not he has eating before the test assessment, along with the time of the year, time of the day and other criteria.

This section will touch on all relevant gait aspects. It provides an overview of the most used gait parameters and gives an overview of how they can be used in gait analysis. It presents the definition of pathological gait and describes deviations of abnormal gait patterns from normal ones. This section also presents some databases commonly used for gait analysis and classification. We address both pure medical research papers and Computer Science gait related research.

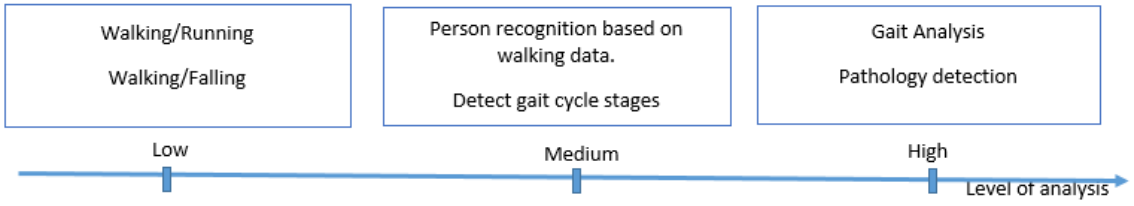


Figure 2.1: Gait related applications and levels of analysis, from simple action recognition to clinical analysis.

2.1.1/ GAIT CYCLE

Gait is periodic by nature, and is commonly divided into cycles, which in turn have 8 key events [186]. Gait cycle is the sequence of events or movements during forward locomotion from the moment in which one foot contacts the ground to the moment when that same foot contacts the ground again. The 8 gait cycle events are:

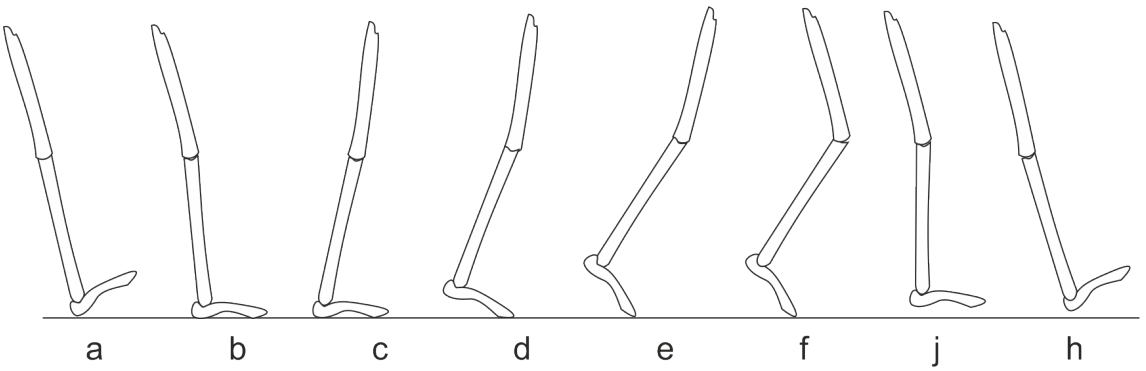


Figure 2.2: Main events of a gait cycle: a) initial contact b) loading response c) midstance d) terminal stance e) preswing f) inital swing j) midswing h) termimal swing. In research papers, the starting phase may vary.

1. Initial contact: A heel strike initiates the gait cycle and represents the point at which the body's centre of gravity is at its lowest position.
2. Loading response: foot-flat is the time when the planar surface of the foot touches the ground.
3. Midstance: occurs when the contralateral foot passes the stance foot.
4. Terminal stance: heel-off occurs as the heel loses contact with the ground and push off is initiated via the triceps muscles.
5. Preswing: toe-off terminates the stance phase as the foot leaves the ground.
6. Initial swing: acceleration begins as soon as the foot leaves the ground and the subject activates the hip flexor muscles to accelerate the leg forward.
7. Midswing: occurs when the foot passes directly beneath the body, coincidental with midstance for the other foot.
8. Terminal swing: deceleration describes the action of the muscles as they slow the leg and stabilize the foot in preparation for the next heel gait cycle.

Fig. 2.2 shows the main events of a gait cycle.

Generally, step cycles are defined as two successive heel strikes of the right and left feet. It is important to talk about gait cycles before listing other parameters because many of them are related to the gait cycle phases.

Temporal Gait Segmentation

Since gait is periodic, it is straight forward to use its repeating part for the gait analysis. Analysis of gait data often examines gait variables with reference to one or more of these gait events or phases. Therefore, temporal segmentation of the gait into cycles is a common and essential step in many gait-related works.

There are many different approaches to segment the gait, and we introduce several of them below. The segmentation is relatively easy to perform. However, the presence of noise can make this task more challenging. We overview the methods used in the context of a Kinect based gait assessment system.

Joint-based methods, when the skeleton is first extracted from the depth image, are the most used to segment gait temporally. A common method is to analyze intra body parts distances or their trajectory to find repetitive parts. The study by Pfister et al. [155] used the time from peak hip/knee flexion to peak hip/knee flexion of the same limb in order to define stride timing. In the work of Nguyen et al. [211] each gait cycle is estimated by a pair of two consecutive local maxima of a sequence of distances between two feet. A sliding window of fixed length is used to detect the maxima. Devanne et al. [199] propose to analyze the evolution of the limbs motion along the sequence. Authors consider that within a time interval corresponding to transition between two steps, the motion of limbs is low. Conversely, within a time interval corresponding to step performance, the motion of limbs is higher. To characterize the limb motion, researchers propose to analyze the shape variation of the two legs. From skeleton data provided by depth sensors, they use joints corresponding to legs in addition to the hip center to build a 3D curve by connecting these joints. This 3D curve is used to find repeating peaks and valleys and to segment the sequence.

It is also possible to divide the gait into cycles without the skeleton data as shown by Stone et al. [89]. Researchers used the projection of 3D point clouds for each frame on the ground. By analyzing the dynamics of the center of projections limited by a given height, the different steps are determined. Auvinet et al.[167] use the cyclic longitudinal distance between the knees to estimate heel strike and identify steps. The knee region is estimated by an empirical method based on human body proportions statistics and assumed knowledge about acquisition setup.

There is no particular common used algorithm to divide the gait in cycles. Researchers are mainly adopting the simplest ones, or the ones which are most robust to noise. The starting phase of the gait cycle can then vary depending on the selected segmentation method.

2.1.2/ GAIT CHARACTERISTICS

The objective of the gait assessment is achieving a quantitative objective measurement of the different parameters that characterize gait automatically. The list of gait parameters is exhaustive. We selected the twenty most widely used along with their estimation methods. Later in 2.3.2 we review skeleton-based gait parameters assessment approaches in detail.

Stance time: It is the amount of time that passes during the stance phase of one extremity in a gait cycle. It includes single support and double support. Usually, the stance time is about 60% of the cycle time with a normal speed. [82] train a unified multi-view pHMM model on the silhouette images to estimate the gait stances.

Time of heel strike and toe-off: The gait phase during which the forward leg that is traveling through space finds its initial contact with the ground. Clark et al. [115] proposes an unsupervised automated analysis algorithm to estimate the gait event time points for toe-off and ground contact.

Swing time: Swing phase is a non-weight-bearing phase of gait, i.e. when foot swings forward between one episode of ground contact and the next. Swing time is the amount of

time that passes during the swing phase of one extremity in a gait cycle. This parameter is quite hard to estimate without a ground-force platform.

Speed and velocity of the gait: These two parameters are relatively easy to estimate, especially when the average walking speed and velocity is to be calculated for the whole motion body. Speed is the rate at which someone moves, and velocity is the speed in the given direction. Gait speed is commonly both easy to determine using Computer Vision methods and a reliable parameter. Both speed and velocity could be calculated using motion flow estimation algorithms. Clark et al. [172] propose a kinect-based solution to assess the gait in people living with stroke. Amongst others, they estimate anterior displacement of the shoulder center through-out the field of view of the Kinect, to obtain mean and peak gait velocity.

Stride frequency or Cadence: Cadense corresponds to the number of steps per minute during normal gait. Stride frequency is one of the key parameters studied by Barak et al. [44] for elderly fall detection. Easy to calculate with any cycle segmentation method from the section 2.1.1 and a commonly used parameter.

Stride and step length: It is the distance covered in an average step, either from heel to heel or toe to toe. Stride length is the distance covered between successful points of the same foot. Stone and Skubic [89] computed the subject's centroid in depth images to deduce the step length and the lower part of the legs to detect gait cycle. In the study by Patterson et al. [72] a stage showing support on both feet describe the balance of a person while walking. It can also be used in order to determine some neurological diseases.

Walking base: It is side-to-side distance between the line of the two feet, taken between the midpoints of the heel. This can be easily estimated using the skeleton data as demonstrated by Mentiplay et al. [182] or a ground-force platform.

Step angle/Turning Angle: This parameter corresponds to the degree of toe out when measuring the angle formed by each foot's line of progression and the line intersecting the center of the heel and the second toe. Step angle can be easily estimated using the ground-force platform but is more challenging to accurately estimate by NWS methods (unless an upper-view is considered).

Knee, hip and ankle flexion angles: Flexion is a bending movement around a joint is a very important criteria of the gait, which can be characterized by the evolution of the angle between two limbs during a gait cycle. Commonly, the knee flexion angle is calculated using skeleton data provided by the Kinect sensor. With the use of the skeleton, the angle calculation is pretty straightforward. For example, Nguyen et al. [211] compute the low-limbs angles in the following way. Two planes are first computed with the three joints of each leg. Then, the angle of interest is determined as the angle between the two normal vectors of these two planes.

Existence of tremors: Tremors in walk are sudden movements or impairments, often due to neurological disorders. Usually tremors are connected to impairments of the gait, which can be detected by evaluating the symmetry of the gait. The estimation of tremors is straight forward, but can be challenging in the presence of noise.

Body posture: The overall body posture can often be described by the center of gravity, or the lean angle of the body, position of the shoulders and head. Pelvic rotation can also be evaluated. Sometimes, arm swings can also be measured. Usually researchers evaluate the body posture by the skeleton data provided by a Kinect or a MOCAP system.

Symmetry of the gait: It is an important criteria of a pathological gait. The following diseases could lead to a non-symmetrical gait: cerebral palsy, stroke, hip arthritis and leg length discrepancy [168]. Therefore, gait asymmetry can consequently be used to identify pathology and track recovery. The symmetry index based on the longitudinal spatial difference between lower-limb movements during the gait cycle was calculated using a point cloud data by Auvinet et al. [168].

Body segment orientations: These are commonly used data in the case of markers-based or WS based gait analysis. Body segment orientations are also provided by the Kinect sensor SDK.

Ground reaction forces: GRF is the force exerted by the ground on a body in contact with it. It is a commonly used gait feature, but a ground platform is required to obtain it.

Muscle Forces: These correspond to the force and torque reactions at various body joints. Two electromyographic (EMG) based techniques are often used to estimate muscle forces. They are inverse dynamics with static optimization and computer muscle control that uses forward dynamics to minimize tracking.

Next, there is a group of parameters related to the physical capabilities of an individual or based on the detection of particular events.

Traversed distance: The parameter is usually related to the maximum walking distance of the person, and is mostly measured with the use of WS only.

Gait anatomy: A group of different parameters related to endurance. The maximum time a person can walk, taking into account the number and duration of the stops, speed, et cetera. This parameter is hard to evaluate in a laboratory condition. The usage of a treadmill can possibly facilitate the task.

Falls: Falls are a major cause of injuries in the case of the elder population. In gait analysis, most attention was dedicated to the development of fall prediction systems. The methods used for fall predictions are various in the selected parameters and sensors and a detailed recent overview is [241].

Trajectory: This is the parameter related to the possibility of the patient to walk in a straight line, or the amplitude of the movements when he/she is performing a turn. Can be estimated by various WS and NWS methods.

Long-term monitoring of the gait: A characteristic related to repeatability of the gait. It is a good way to evaluate the gait and possibly degradation due to age or other factors. Rarely used in a real-time scenario, unless a patients rehabilitation to recovery is monitored systematically.

The in depth research was performed by Roberts and Mongeon [242] to identify the most used biomechanical gait parameters. According to the authors, spatio-temporal parameters were found to be the most often measured biomechanical parameters and reported by the greatest number of articles. They specifically include:

- walking velocity (50 articles)
- cadence (30 articles)
- stride length (23 articles)
- step length (21 articles)

If only a single parameter is evaluated in the gait study:

- walking velocity (50 articles)
- ankle angle (47 articles)

Figure 2.3 shows the parameters grouped together as categories.

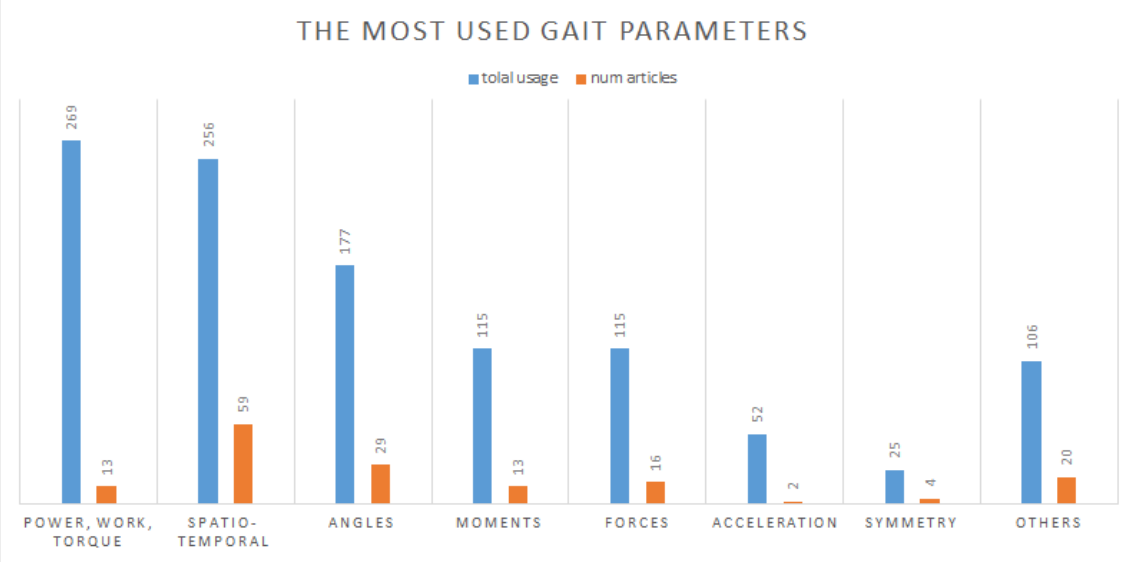


Figure 2.3: The most relevant groups of biomechanical parameters for gait analysis in the healthy adult population according to the research by [242] dated 2017. Power-related and Spatio-temporal parameters are the most widely used, however the spatio-temporal and kinematic are used more often when only a single parameter is considered.

2.1.3/ Pathological Gait

Abnormal or pathological gaits are the gaits which deviate from normal gait patterns. There could be many different reasons for such a deviation, the most common of which are different neurological diseases or age-related pathology.

Because walking patterns differ dramatically in impaired populations, neurological patients, and elders, it is difficult to develop an algorithm that works in a uniform fashion for gait patterns that are not homogeneous.

Gait assessment allows to detect pathology and neurological diseases:

- Neurological diseases such as multiple sclerosis or Parkinson's.
- Systemic diseases such as cardiopathies, in which gait is clearly affected.
- Alterations in the deambulation dynamic due to sequelae from stroke.
- Pathology due to lower-limb amputation.
- Diseases caused by aging, which affect a large percentage of the population.

There are numeral causes of abnormal gait. Here we provide with the list of several types of pathological gait caused by different reasons:

Antalgic: Is a pathology with a stance phase on the affected side shortened. This can be caused by pain or neurological disturbance.

Trendelberg gait (or gluteus medius lurch): Trendelberg gait is an abnormal gait caused by weakness of the abductor muscles of the lower limb, gluteus medius and gluteus minimus. People with a lesion on the superior gluteal nerve exhibit weakness of abducting the thigh at the hip.

Psoatic: Gait which causes flexion and laternal rotation of the leg or hip due to psoas muscle spasm.

Equinous: This gait is characterized by habitual tiptoe walking. This can be due to a leg length discrepancy.

Gluteus Maximus: Gait contracts at heel-strike, slowing forward motion of trunk by arresting flexion of the hip and initiating extension.

Other pathological gaits are hemiplegic, quadriceps gait, scissoring gait, and ataxic gait. A detailed presentation can be found in [257].

In this work we address patients with lower extremity prosthesis. There are several types of low-limbs amputations:

Knee Disarticulation: This pathology corresponds to an amputation performed between bone surfaces, rather than by cutting directly through bone. With this amputation, the residual limb generally can tolerate some weight bearing, providing a long mechanical lever controlled by strong muscles. By retaining a full length femur the thigh muscles tend to be stronger, because they are released at their distal attachments, rather than bisected in the muscle belly.

Transfemoral Amputations: This amputation is accomplished by cutting directly through the femur bone of the upper thigh. The primary prosthetic concern with this type of amputation is that the patient is required to control and manage two artificial joints: the knee and ankle. Also, weight bearing on the residual limb typically causes pain, so the socket design must accommodate for this.

Transtibial Amputations: This amputation is surgically accomplished by cutting through the tibia and fibula of the lower leg. A large portion of BK amputations are due to peripheral vascular disease and poor circulation of the lower extremity. Although retaining the knee joint improves overall function and ability, muscle contractures of the knee and hip can hinder recovery and physical therapy.

Partial Foot Amputation: This kind of amputation is often the result of advanced vascular disease, secondary to diabetes. A partial foot is characterized as an amputation occurring at the fore, mid or hind foot, not including the disarticulation of the ankle foot, which is specified as a Syme's. PFA's are historically known to have a significant failure rate and various complications, including ulcerations, skin breakdown, contracture and proximal amputations.

Gait deviations in lower limb amputees can be broadly broken into patient and prosthetic causes [255]. Patient Causes:

- Muscle weakness
- Contracture
- Pain
- Decreased confidence in the prosthesis or residual limb
- Habitual / learned behaviours

Prosthetic Causes:

- Prosthetic misalignment
- Poor fitting prosthetic socket

Most relevant biomechanics and physiological parameters during gait in lower-limb amputees were evaluated by Sagawa et al [86]. The main parameters used by other researchers in clinical studies belong to spatio-temporal ones (in total, 153 articles):

- self-selected gait speed (43 articles)
- cadence (19 articles)
- step length (17 articles)
- stride length (20 articles)

Kinematic parameters were less used that spatio-temporal ones on the average (in total, 78 articles):

- Ankle angles (22 articles)
- Knee angles (31 articles)

According to [86], additional spatio-temporal parameters include stance time, stance time ratio, and step time ratio. Several features are common for low-limbs amputees gait. During the gait cycle, the stance phase on the sound limb is slightly longer than on the prosthetic side. This contributes to a more asymmetrical gait.

Foot flat time is another spatio-temporal parameter that seems appropriate for low-limbs amputees gait analysis. Healthy subjects reached the foot flat phase at 12–17% of the gait cycle compared to subjects using a solid ankle cushion heel (SACH) foot, who were found to make heel contact only for the first 20% or 44.5%. Other parameters which usually are affected by the prosthesis are: reduced hip range of motion (ROM) in the sagittal plane, the dorsiflexion motion during the mid-to-late stance phase, Plantar flexion in the early stance phase, and the total ankle ROM in the sagittal plane.

However, the effects of amputation on the gait pattern stated in Physiopedia [255] are not limited to the parameters mentioned in [86]. According to [255], common deviations for prosthesis patients can be grouped in three categories listed in Tables A.1 A.2 A.3 in A. Proteor has mostly analyzed the Kinematic or Transtibial parameters as can be concluded from the patient study reports in A.2. The second group of parameters evaluated in patient assessments by Proteor is obtained by the GRF sensors.

2.1.4/ GAIT INDEXES

In clinical studies, some quantitative metrics, or so called gait indexes were proposed in order to assess pahtological gait. Such indexes commonly are based on the evaluation of a group pre-selected gait parameters. Then the final score, which quantifies the gait, is calculated from the single attributes. The most known of such indexes are Gait Deviation Index, Dynamic Gait Index and Gilette Gait Index.

The Gait Deviation Index (GDI) [57] is a multivariate measure of overall gait pathology. The calculation is based on kinematic gait data. Gait deviation index scales with respect to clinical involvement based on topographic Cerebral Palsy (CP) classification in Hemiplegia Types I-IV, Diplegia, Triplegia and Quadriplegia [71].

The Gillette Gait Index (GGI) is a summary measure incorporating 16 clinically important kinematic and temporal parameters. It cab be used for both computer-assisted and human-assisted gait impairment assessment [53, 47]. GGI is considered to be a useful outcome measure in patients undergoing gait analysis, although it is probably not a widely-practised analysis tool yet, except for the diagnosis of Cerebral Palsy [53].

Dynamic Gait Index (DGI) is used to measure for dynamic balance in people with chronic stroke [48] and evaluate the vestibular function in general [39].

There are more different Gait Indexes proposed by authors in the research literature. In her Ph.D thesis Caroline Hodt-Billington [105] evaluates many of them as the measures of gait symmetry in subjects with disease or injury related to one-sided affection. The estimation of the symmetry gait indexes using the 3D sensor is a very promising research direction.

Table 2.1: Available gait datasets

Dataset	year	Device	subjects	modalities	Comment
TUM Gaid [145]	2014	RGB-D	305	audio, image and depth	Person identification and for the assessment of the soft biometrics
Kinect Gait Biometry [137]	2014	Kinect v.2	164	skeletons	Gait Biometrics. User follows a semi-circular path and the sensor followed the movement using a spinning dish
SPHERE-Walking2015 [153]	2014	RGB-D (Kinect v1)	20	skeleton	Normal, PD and stroke simulating persons climbing stairs, 15 joints, noisy data
DAI [169]	2015	Kinect v.2	7	skeleton, orientations	Normal/abnormal (asymmetric) sequences performed by actors
UPCV Gait K2 [204]	2016	Kinect v.2	30	skeleton	Gait recognition database
Walking gait dataset [251]	2018	Kinect v.2	8	skeleton, point clouds	A treadmill is used, padding sole and attaching weight to simulate the pathological asymmetric gait

2.1.5/ GAIT DATASETS

According to EU law, the collection of patients data requires many permissions and a special laboratory setup. At the same moment, the research teams working on Computer Vision based motion analysis do not commonly have access to a clinical environment. Hence, there are very few benchmark datasets publicly available in the gait assessment field.

The absence of data is a big issue in clinical gait studies. Normal gait is more prevalent, since we can use data from the gait recognition domain, where less restrictions apply. The amount of publicly available gait data is small compared to the number of gait studies that have been performed over the years. The data that is available generally suffers from limitations such as little to a few subjects, few gait cycles, highly clinical, no raw data, lack of meta data, non-standard formats, and restrictive licensing [183]. Moreover, there are very few examples of the abnormal gait publicly available for a reuse. We collected the information about the gait dataset aquired with a RGB-D camera devices and shared with the other users. It should be noted, that there exist more databases collected with the other devices, such as 2D cameras, golden-standard systems, WS and others. This work [183] reviews many of these databases. Therefore, we concentrate here on the data acquired by an RGB-D sensors, and the Kinect camera in particularly. Information about the existing gait datasets is grouped in Table 2.1. It can be seen that commonly the datasets contain a small number of test subjects, and the gait-affecting diseases are simulated.

Since the goal of this dissertation is the automatic abnormal gait assessment using an RGB-D sensor, we will discuss a number of relevant datasets in detail:

TUM GAID [145]: This dataset was created for depth based gait recognition and assessment. The database simultaneously contains RGB video, depth and audio. TUM GAID contains 305 individual gait captures, acquired in different weather conditions and in a different context, i.e: the person walks normally or he/she is wearing a backpack or coating shoes. Some persons performed all the actions and some only a subset. The camera is located perpendicular to the subject.

Gait Biometry Dataset [137]: The dataset contains raw data from 164 subjects extracted using a Kinect sensor. Each subject walked at least once over a semi-circular path and the sensor followed the movement using a spinning dish. All walks were performed indoors with artificial lighting, but lighting conditions were not controlled. Folders contain all walks for each subject. The number of walks varies across subjects and the number of frames varies across walks.

SPHERE-Walking2015 [153]: The dataset contains normal and abnormal gait sequences collected with an RGB-D sensor. There are two parts: the bigger one corresponds to individuals climbing stairs, and the second one is acquired with a flat surface setup. A qualified physiotherapist, who was not included in the model training phase, simulated three standard scenarios of knee injury for actors to perform. The other abnormality is the freeze at some stage of the movement. Despite the fact that Sphere dataset joints are extremely noisy, the dataset was wholly or partly reused in many later studies.

DAI dataset [169]: The data contains 7 actors performing normal and abnormal gait. The abnormal gait includes two anomalies, a knee injury that implies that the knee cannot be bent but otherwise the gait is normal, and a second one where one foot is dragged towards the other, which usually happens when a mobility aid such as a crutch or a handrail is employed. The dataset design requires one to use only the normal examples for training and then detect the outliers, i.e pathological gait sequences, during the test. The first scenario uses all the actors while the second leaves 3 unseen actors for the test. The dataset is very small and some sequences contain only one gait cycle. However, the corresponding method was cited 10 times, and the dataset used by other researchers such as [199].

Walking gait dataset [251]: was recently proposed by Nguyen et al. It has been established to enable comparative studies on gait analysis, especially the problems of gait index estimation and abnormal gait detection. The dataset includes 9 normal gaits and 8 simulated abnormal (asymmetric) ones performed by 9 individuals on a treadmill. Abnormal gaits were simulated by attaching a weight to a foot and padding a sole. The dataset is very recent but it has potential since it contains a big number of data items. However, it is unbalanced towards abnormal gait samples.

The number of datasets dedicated to pathological gait available is very small, and the existing ones mostly contain very few trials and subjects. This makes it difficult to generalize the results and compare the existing algorithms. The gait pathologies are mostly simulated by actors [169, 153] or created by introducing artificial difficulties [251] such as sole padding. Most datasets store only the skeleton data, and no color, depth or even orientations data from the Kinect sensor. The biggest multi-modal database of all is the recent one by [251]. However, Nguyen et al. do not store all the data from the skeletons. Specifically, the joint states and orientations are missing. In addition, a treadmill was used to acquire this dataset, which slightly affects kinematic parameters [102]. Hence, this data cannot be used in the more common solid surface walking scenario.

To advance automatic gait assessment, more data is required. This data should satisfy the multi-modal criteria, contain a bigger number of test subjects, and be captured in normal conditions.

2.1.6/ Gait Parameters Conclusion

Gait is a complex process and can be characterized by different kinematic and spatio-temporal parameters. It is not possible to extract all the parameters using Computer Vision-based methods. However, a group of important and commonly used gait characteristics relevant to low-limbs amputees can be obtained. It was proven by a number of CV-based gait studies that spatiotemporal gait parameters can be calculated reliably from 2D and 3D sensor data. Kinematic parameters, however, were rarely assessed. In accordance with the information gathered for this thesis, we suggest that both transtibial and transfemoral parameters can be derived from kinematic parameters.

As stated earlier, currently, clinicians from Proteor perform subjective gait analysis or use GRF platforms and marker-based systems in more complicated cases. Special attention is dedicated to gait forces parameters and kinematic parameters, such as knee, hip and ankle flexion and moments. There is a lack of consensus between the parameters used by researchers and parameters assessed by clinicians, in particular practised by Proteor.

After studying the main gait parameters and existing gait analysis approaches, we decided to focus on those parameters that are tailored to detect gait symmetry. Specifically, kinematic parameters and all parameters related to the timing of specific gait phases. The symmetry is also the parameter indirectly verified by the clinicians in Proteor, when they compare the low-limbs flexion angles, since if a prosthetic limb shows different kinematic trends than the second limb, it usually signals a problem. The Proteor in there sight is interested in automatic gait classification and assessment.

The goal of this project is to propose an unsupervised gait classification tool. This task can be three-fold: we aim first to perform a binary gait classification, then a specific pathology detection, and then to complete a qualitative gait motion evaluation. Current work is mostly covering the problematic of binary and plural gait pathology classification.

2.2/ 3D Descriptors for Actions, Gestures and Gait

The first focus of our research is on 3D descriptors. This is a feature vector used to identify the shape of an item. This section aims to provide a comprehensive reference source on depth-based human motion descriptors. Motion description is a challenging problem, which became popular with recent advances in 3D Computer Vision. Our purpose is twofold. First, we introduce the main trends in human 3D motion descriptor design and evaluation. Second, we present a review of recent methods belonging to three different application categories: action recognition, gesture recognition and gait assessment. Selected categories have different specifics, which allow us to highlight aspects of a motion descriptor construction. A comparison of different methods by their main characteristics is provided. Finally, possible directions and recommendations for future research in 3D motion description are outlined.

2.2.1/ 3D Data

Visual or image descriptors are feature vectors of the salient regions of images and video sequences. Descriptors aim to describe the visual characteristics of present objects such as shape, color, texture or other and may replace an image as the input to a classifier. In general descriptors ideally are very short, descriptive and discriminative, yet highly robust. The latter requirement illustrates the divide between the general multi-media and cryptographic domains. Commonly, in the order of hundreds of individual descriptors vectors are taken from a single originating image source. Depending the application domain, researchers have designed a host of descriptor algorithms emphasizing one requirement over an other. Binary descriptors, for example, trade entropy and robustness against compactness and speed.

Specially designed and tailored descriptors can also represent motion, which is an essential part of algorithms in rather diverse applications, such as human activity recognition, gait recognition and analysis, motion tracking and 3D scene reconstruction to name a few examples.

With the advent of low-cost 3D sensing cameras and continued efforts in advanced point cloud processing, 3D perception has gained more importance in the vision domain. 3D sensing devices not only provide the user with a general projection of the 3D world to a

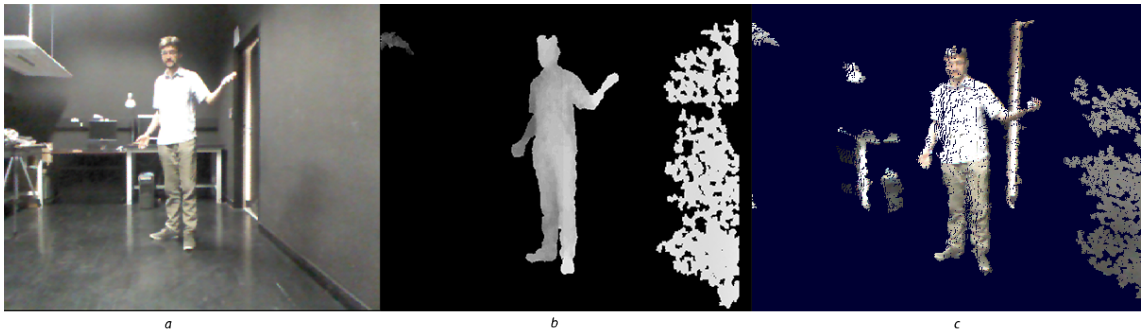
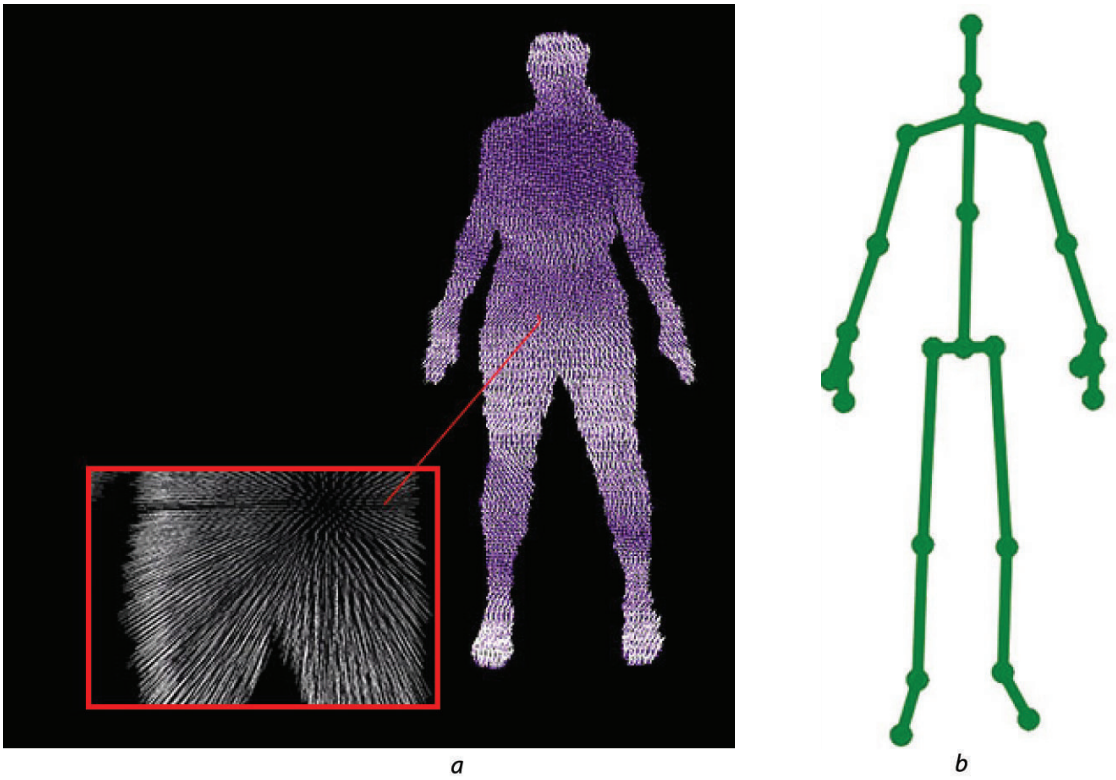


Figure 2.4: Images obtained with Microsoft Kinect v1 Sensor: a) RGB image; b) depth map c) 3D point cloud.

2D image plane as regular cameras do, but also acquire the 3D geometry, or depth. An example of information provided by the popular 3D Kinect sensor is shown on Fig. 2.4. Depth images deliver natural surfaces, which can be exploited to capture geometrical features of the observed scene with a rich descriptor. Compared to conventional color data, the additional depth information in RGB-D data helps to adjust for different lighting conditions, accustom for a different viewing point, remove background noise and simplify intra-class motion variations. Therefore, in general, RGB-D-based descriptors outperform the RGB-based ones [184].

The standard output of a 3D sensor is a depth map, which can be converted to a point cloud. One point in a point cloud contains XYZ coordinates and an optional color tuple. A descriptor specifically designed for a point cloud video sequence could serve as an important basis for motion characterization.

Information extracted from 3D point clouds is predominantly comprised of shape, color (or intensity) and the spatial relation between cloud points. Shape descriptors are the most popular 3D descriptors for point clouds, and amongst them normal-based descriptors are the most widely used. Normals are the perpendicular vectors to point cloud regions. They provide most of the shape and structure information of an object in 3D. There are many methods to estimate them, but the simplest is to find the normal of a plane tangent to the surface, which is a least-square plane fitting estimation problem. An example of a point cloud with normals is shown in Fig. 2.5 a.



a b

Figure 2.5: a) 3D point cloud with normals estimated by the analysis of the eigenvectors and eigenvalues of a covariance matrix created from the nearest neighbors of the query point [61]. b) 25 skeleton joints estimated by the Kinect Studio 2.0 software and Kinect v.2.

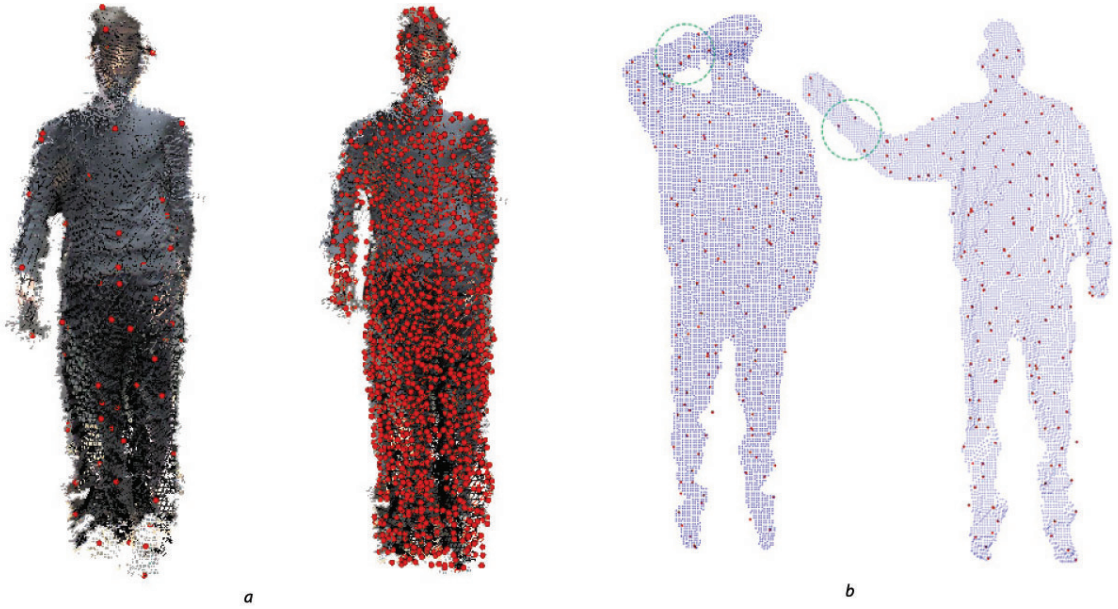


Figure 2.6: a) Key point detected with manually set parameters. Left: Harris 3D (64 points) and right: ISS 3D (2019 points) b) The resulting points detected by the Harris 3D algorithm on 2 depth frames of the same sequence from the MSR3D dataset. Due to the noise and non-rigid movement, a different set of key points is produced for the same region of an arm shown by a green circle.

Widely applied 3D shape descriptors include the Normal Aligned Radial Features (NARF) [75], the Spin image descriptor [26] and HOG3D [56], SURF and SIFT [52]. All of them were originally proposed for two dimensional planes and later modified for 3D data. These descriptors may be used to represent a motion trajectory when applied to individual consecutive frames [195] but mostly they are used for static point clouds. Before applying the descriptor, potentially interesting points should be identified, preferably, automatically. Seldom researchers try to select the most representative points by applying a descriptor to all the data points of the model and comparing the values to select 'rare' ones [42]. Commonly descriptors capture an object geometry around some key points or interest points, which can be selected by a detector. This is done to evade the high computational cost and storage space required to extract features from all points in a dense point cloud. Widely used key-point detectors are Intrinsic Shape Signatures (ISS) [65], NARF [131], 3D SIFT and Harris 3D, to name a few. Depending on the application requirements, interest point detectors using the color/intensity and/or geometric information (XYZ coordinates, normals) are available. It should be mentioned that most detectors are designed to use color or intensity information in addition to geometric information. Many 3D descriptors and detectors have been implemented in the Point Cloud Library [85]. An example of a popular key-points detector output is shown on Fig. 2.6. Harris 3D detector, which is based on surface normals and the ISS detector, which is based on the eigen values of the points position within a ROI with a specified radius, are used to find points of interest on a human silhouette point cloud.

The parameters of the detectors can be manually adjusted, which results in more/less points detected. The main problem of using key points detectors for human motion analysis is the fact that human body is not rigid. This complicates the use of standard key-point detectors such as Harris 3D, ISS and 3D-SIFT because their results are not consistent

between video frames. The following local descriptors will then be hard to match in order to find corresponding points. Fig. 2.6 b illustrates an example of the Harris3D detector applied on 2 video frames of an arm movement. It is possible that global descriptors (such as) 3D Zernike moments [37], shape histograms [25] and others would be better adapted for human movement scenario. We advise the work [142] for more information and a comparative evaluation of 3D key-point detectors.

Finally, 3D motion, sometimes referred to as 4D, combines 3D spatial scenes temporally. This data may be described via the calculation of the so-called Motion or Scene Flow, which is composed of 3D velocity vectors. Fig. 2.7 shows an example of a dense (i.e. calculated for each point) motion flow determined from 2 consecutive frames by the PD Flow algorithm [176]. Dense motion flow is information rich, but also computationally intensive to determine and carries a significant memory and storage burden. The use of motion flow data for classification tasks, is usually not straight-forward, and researchers experiment with different quantization schemes. Recently many alternative algorithms have been proposed that target at reducing this burden. This chapter aims to review and categorize this existing family of methods when applied to human motion analysis.

Motion analysis using depth data was extensively reviewed in [135]. However, in this work, researchers were focused primarily on activity recognition methods applied to RGB-D video sequences omitting (3D) motion descriptors. Therefore, we assumed a need for a specialized review dedicated to 3D descriptors applied to human motion, divided into the following subcategories: Action Recognition, Gesture Recognition, and Movement Analysis and specifically, Gait recognition. Action recognition classifies the full-body actions performed by a human subject, gesture recognition classifies local movements, and gait analysis evaluates the walking manner of a person. Three selected application areas allow us to cover the specifics of human motion descriptors for different analysis's level: from more general in action recognition for more fine and particular in gesture recognition and gait analysis. All three fields have been active research topics for many years and profited from the development of 3D cameras, which allows to remove the background easily and reduce the ambiguity of the 2D data. These three applications commonly require to be robust to intra-person variations and can be characterized by the motion cues. Hence, 3D data is the natural choice to proceed.

This chapter contains a review of characteristic methods for motion description. The papers reviewed were selected in an attempt to cover the most popular existing motion description techniques proposed between 2010 and 2017. In our research, we based our choice on their impact on the field for the earlier papers and the degree of novelties proposed in the later papers.

2.2.2/ Motion Descriptor Evaluation Criteria

This section introduces the main characteristics and evaluation criteria for 3D descriptors, focusing on 3D motion descriptors and existing benchmarks.

2.2.2.1/ Motion descriptor characteristics

Common requirements for 3D descriptors are invariance to transformations of the target object and variations in viewing geometries, user-independence, robustness to noise and

clutter next to storage efficiency (i.e., compactness). We propose to compare existing descriptors by a combination of the following characteristics, which highlights descriptor specifics and gives an idea in which scenarios their performance is optimal. The evaluation of the reviewed algorithms according to the chosen characteristics is given in Tables 2.3 and 3.2.

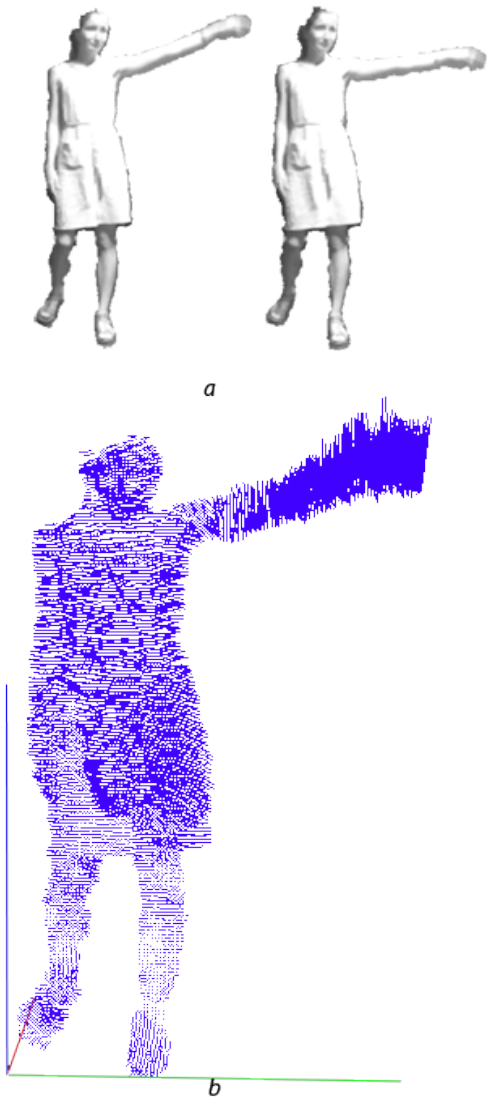


Figure 2.7: a) Corresponding intensity images. b) Motion flow calculated by the PD Flow algorithm [176] from 2 consecutive frames using intensity and depth data.

Application: Corresponds to the particular purpose for which the descriptor is designed. Application influences the requirements for a descriptor: action recognition in general deals which motions with large amplitudes, gesture recognition requires capturing very fine movements, whilst gait analysis combines the two. Datasets for different application categories are available.

Locality: A descriptor can be local (regional) or global based on the features it captures. If the descriptor is local, it is applied to selected points of interest, or on a local support region surrounding a basis point of an object. Commonly algorithms may be split in a detector and a descriptor part, where the detector locates regions that are perceived to be interesting or stable and the descriptor encodes a region of interest around the location of the feature. Using motion descriptors for local features generally provides one with invariance to geometric transformations and gives a reduction of the perturbations caused by variations in scale, rotation, and viewpoint. Using local descriptors also allows to reduce the computational complexity of the algorithms. However, local descriptors ignore the global spatial structure information of the scene.

Global descriptors aim to capture the overall structure of the object. It is applied to all points or to points selected by using a linear sampling scheme (i.e not special key feature points) or even a single point given point and is capable of capturing global spatial structure and affiliation.

Dimensionality: A characteristic which is directly connected to the size and distinguishability of a descriptor. The dimensionality of a motion descriptor is rarely a significant issue unless the descriptors for very long video sequences should be stored or the rich

motion flow based descriptor is used. In this case the general storage burden is predominantly determined by the number of descriptors and not their individual size. We consider the dimensionality as low when the final descriptor size is below 1000, middle from 1000 to 10000 and high otherwise.

View-invariance and scale-invariance: View-invariance of a descriptor entails its robustness against changes in a view-point. The appearance of 3D data allowed to progress a lot in object and action recognition with varying view-points in comparison with 2D. Many approaches for 3D shape retrieval and identification transform an object of interest into a canonical pose. Translating the center of gravity of the object into the origin and normalizing the area/volume or radius of the bounding circle/sphere/cube etc. This operation guarantees view-invariance to a reasonable extend.

Scale-invariance is the invariance of the descriptor for the objects size or the distance of the camera to the object. In Local Spatio-Temporal (LST) descriptors it can be, for example, the fixed support region of the feature point which is not adapted to linear perspective view variations.

Accuracy: This characteristic shows how well the proposed descriptor performes for a given task. In this work, we state the accuracy as reported by the researchers and also comment on the theoretical evaluation of the method proposed. When available, we report the results obtained on a benchmark dataset. However, we should mention that even when an identical dataset is used, the validation method used by each work might differ from the others [152], so a direct comparison is not always possible. In general, we consider the accuracy as high when it is above 90 percent, middle above 80 and low below 80 with some exceptions in case when a special evaluation scheme was selected by authors.

Computational complexity: The computational resources required by an algorithm. Sometimes computational complexity is reported by the authors or could be approximated by the specifics of the algorithm. This criteria is important for real-time applications. We evaluate the complexity by the number of operations needed in order to calculate the descriptor only, not taking into account possible prior calculations (such as skeleton extraction). We assume the computational complexity as low when the descriptor is applied only on the sparse data and simple mathematical operations are used, and high in the case of complex pixel-wise operations on multi-modal data.

Dataset: Motion descriptors reviewed in this work were proposed for three particular applications and evaluated using task corresponding test data. The choice of a dataset obviously affects the reported accuracy of an algorithm. Datasets are evaluated by their difficulty and their data scope.

Classifier: A classifier is a hypothesis or discrete-valued function that is used to assign categorical class labels to particular data. More specific, it is an algorithm that implements the recognition task in an application. The use of a different classifier can influence the final recognition score significantly, however, with a set of discerning features (possibly, with a corresponding encoding scheme [80]) even a very simple classification method such as linear regression can show good results. Classifier selection can be specified by

the algorithm design and the amount of the training data available. As this review shows, SVM is widely used with simple yet powerful BoW model built on descriptors which have a histogram form [112, 70, 144]. CNNs and Random Decision Forests (RDF) are used when a significant amount of testing data is available which is still difficult in the case of 3D data. HMMs and LSTMs are employed when temporal order is of a great significance.

2.2.2.2/ Datasets for Motion Descriptors

When proposing a new algorithm, its performance usually needs to be evaluated in comparison to existing ones for a given application. Publicly available data and established evaluation procedure is a standard research practice. A common database, if it is appropriate for a given task, can be used for designing, training and evaluating of a descriptor. Construction and annotation of a new database is often a long and arduous process, therefore it is preferable to only do this when an existing benchmark can not be used.

Table 4.2 illustrates a list of popular Benchmark datasets most widely used for 3D human motion descriptor evaluation, next to their specifics and designation for a particular task.

MSR Action3D: This is one of the most used RGB-D human action-detection and recognition datasets [112, 128, 111, 110, 113]. The main reason for this is the fact that it is one of the first RGB-D datasets capturing motions (dated 2010) and it contains the biggest amount of different actions. The MSR Action3D Dataset [70] consists of 20 action types performed by 10 subjects 2 or 3 times. The actions are: high arm wave, horizontal arm wave, hammer, hand catch, forward punch, high throw, draw an x, draw tick, draw circle, hand clap, two hand wave, side-boxing, bend, forward kick, side kick, jogging, tennis swing, tennis serve, golf swing, pick up & throw. The resolution of the video is not very high, namely 320x240. The frame rate is just 15 fps. The data is recorded with a depth sensor similar to the Kinect device and contains color and depth video sequences. The sequences are pre-segmented into background and foreground. Skeleton joints data is also provided acquired with a higher frame rate than the depth data. However, there are many erroneously estimated joints in the dataset. An example of wrongly estimated skeleton joints is shown in Figure 2.9. It is a challenging dataset to test an algorithm on and to compare the results with earlier proposed methods but future users should take note of the different evaluation schemes used by other authors.

MSR Daily Activity 3D: Also captured using a Kinect device, it includes 16 activities performed by 10 subjects 2 times, in a standing and a sitting position [111]. The actions are: drink, eat, read a book, call a cellphone, write on a paper, use a laptop, use a vacuum cleaner, cheer up, sit still, toss a paper, play a game, lie down on a sofa, walk, play a guitar, stand up, sit down. There is a sofa in the scene. RGB and depth channels are recorded, and also the skeleton joint positions are extracted. However, the RGB channel and depth channel are recorded independently, so they are not strictly synchronized. The dataset is more challenging than MSR Action3D, because it represents natural everyday activities, which are harder to distinguish. Most activity sequences involve human-object interactions. MSR Daily Activity 3D is a good choice to evaluate a real-life scenario on an application and compare the results with other algorithms [222, 111, 128]. Alternative is the similar more recent dataset Watch-n-Patch [193] captured by Kinect 2. However, it is less used so by researchers so far.

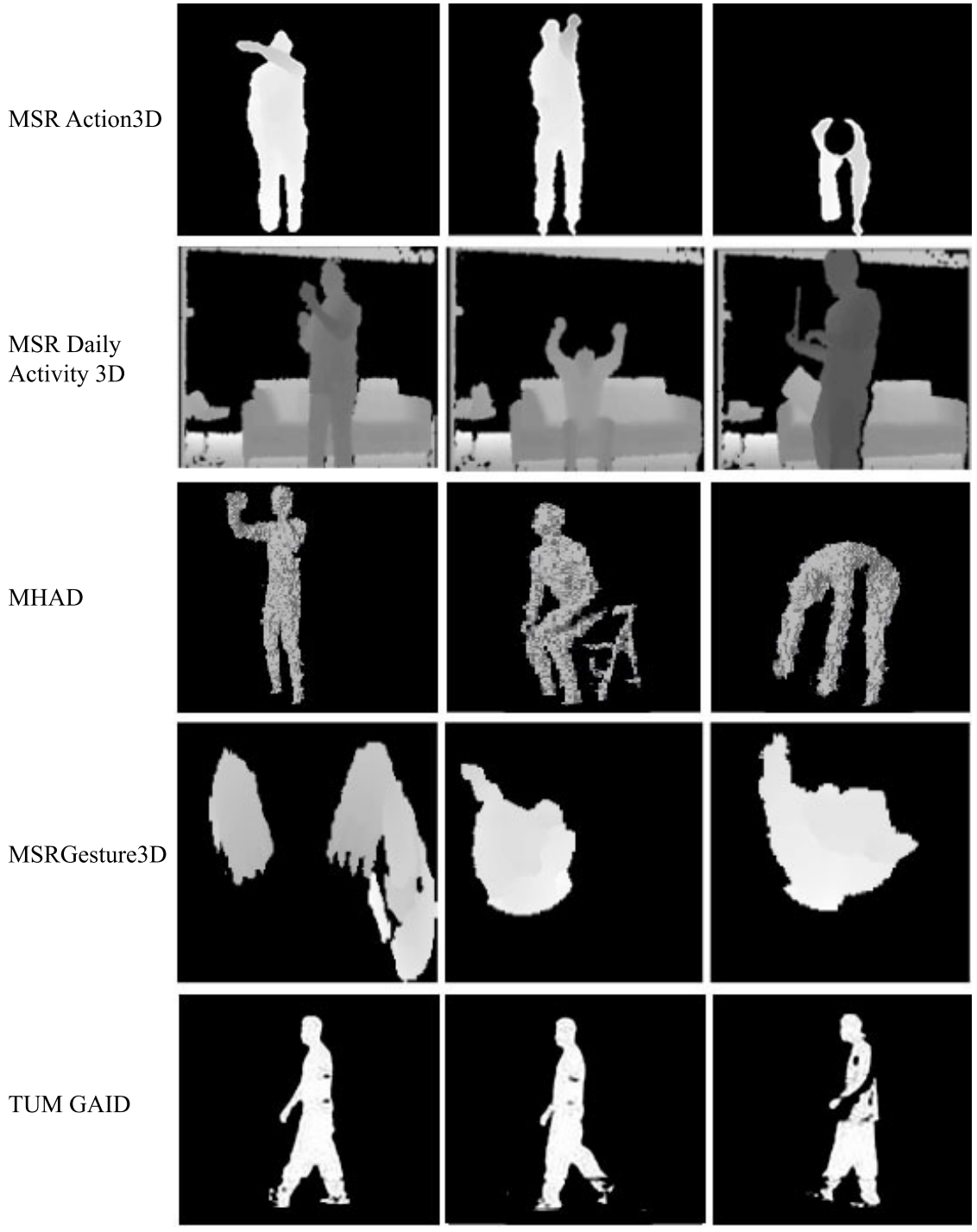


Figure 2.8: Examples of depth maps from 3D motion datasets.

Berkeley MHAD: The Berkeley Multimodal Human Action Database (MHAD) [127] is a complete and general purpose dataset which consists of temporally synchronized and geometrically calibrated data from an optical Motion Capture system, multi-baseline stereo cameras from multiple views, depth sensors, accelerometers and microphones. The dataset contains 11 actions performed by 7 male and 5 female subjects (they are in the age range of 23-30 years except for one elderly subject) 5 times. The total recording time is 82 minutes, which makes this dataset one of the biggest by the amount of video

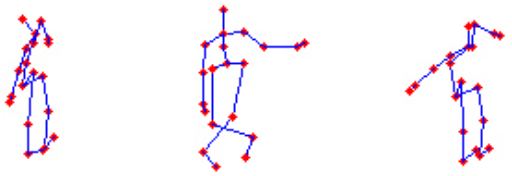


Figure 2.9: Examples of skeleton estimation for the MSR 3D dataset, actions 'High Arm Wave', 'Horizontal Arm Wave', 'Hammer'. In reality, person is always standing straight facing the camera, legs not crossed.

sequences it contains. The specified set of actions is comprised of the following: (1) actions with movement in both upper and lower extremities, e.g., jumping in place, jumping jacks, throwing, etc., (2) actions with high dynamics in upper extremities, e.g., waving hands, clapping hands, etc. and (3) actions with high dynamics in lower extremities, e.g., sit down, stand up. Berkeley MHAD is popular [222, 195], probably due to the fact that it allows to perform a multi-modal analysis of human motion and is very easy to use with all the data access scripts provided on the official dataset website.

MSRGesture3D: This dataset was captured by a Kinect device. There are 12 dynamic American Sign Language gestures performed 2-3 times by 10 people [108]. The hand sign language represents a concentrated entity in motion, combining both the overall movement of hands and the inner variability of fingers. The hand portion (above the wrist) is segmented. An alternative to this dataset with less gestures is the Sheffield Kinect Gesture (SKIG) dataset [122], which captures forearm gestures under different hand poses, background and illumination variations from 6 subjects. This makes SKIG more applicable to a real-life scenario than the MSR Gesture 3D dataset. However, the latter remains the most widely used for gesture recognition research [108, 141, 197].

TUM GAID: For depth based gait recognition and assessment, this challenging multi-modal recognition database [145] was proposed in 2014. This database simultaneously contains RGB video, depth and audio. The database contains 305 individual gait captures, acquired in different weather conditions and in a different context, i.e: the person walks normally or he/she is wearing a backpack or coating shoes. Some persons performed all the actions and some only a subset. The camera is located perpendicular to the subject. Other databases for depth gait analysis are available. However, TUM GAID is the most cited and most currently used. It is the only database which allows for multi-modal gait recognition using video, depth and audio features along with different acquisition conditions.

Table 4.2 summarizes the specifics of each dataset in numbers and Fig. 2.8 shows examples of the data.

Alternative datasets: In the previous section we provided the most used datasets. However, there are many more available. The most popular ones are mentioned below. For researchers interested in 3D Skeleton data only, there are many other databases not mentioned in this review.

Table 2.2: Popular 3D Video datasets and their characteristics

Dataset	Actions	Persons	Calibration synchroni-zation	Annotation	Sequences	Citations
MSR Action 3D	20	10	no	skeleton joints, segmentation	567	708
MSR Gesture 3D	12	10	no	segmentation	336	187
MSR Daily Activity 3D	16	10	no	skeleton joints	320	717
Berkeley MHAD	11	12	yes	multi-modal	660	132
TUM GAID	3	305	no	metadata	3370	50

- Action recognition: UTKinect Action [112], MSR ActionPairs [128], ACT4 Dataset [94], Cornell Activity Datasets: CAD-60 & CAD-120 [121];
- Gesture recognition: G3D [108], WorkoutSu-10 Gesture [126];
- Gait analysis: K3Da [180] (a limited number of gait sequences is available).

With the technological progress and constant improvement of depth cameras and skeleton estimation methods, there is definitely a need for new challenging multi-model datasets in all 3 domains, but especially in gait analysis.

2.2.3/ Motion Descriptors for Action Recognition

Motion description is an essential part of human activity recognition. From a computational perspective, actions are best defined as four-dimensional patterns in space and in time. Methods able to discriminate the class of an action being performed are based on the analysis of a video sequence combining a motion descriptor with a classifier. Most of the work on human action recognition published up to today relies on information extracted from 2D images and videos. However, with the availability of affordable depth sensors, this research area enlarged considerably with new studies dedicated to 3D. For a detailed review on action recognition, the works [91] and [74] could be referred. These reviews, however, are not particularly oriented on RGB-D-based methods as ours.

Motion descriptors can be categorized with respect to various criteria. A general classification feature-based scheme is shown in Fig. 2.10. The two main groups of methods are Skeleton joints based and Depth map and color based. The skeleton can be extracted from the depth data using the famous method proposed by Shotton et al. [130] or with a better accuracy by different motion capture systems with markers. Commonly, the software for popular depth sensors such as Kinect v1-2, provides the algorithms for automatic joint extraction. A skeleton obtained with the Kinect v.2 and OpenNI [256] is shown on Fig. 2.5 b. The software estimates 25 joints for each body. Each joint has a position and orientation. In other literature authors also name similar classifications as model-based and model-free. In this case human model-based approaches represent an action with body segments, joint positions, or pose parameters. The Depth map and

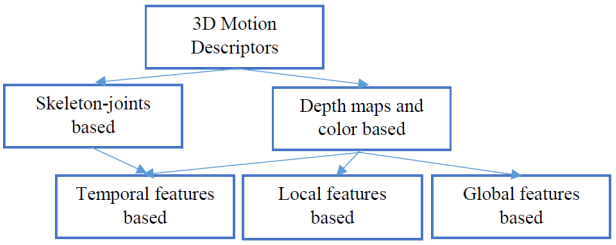


Figure 2.10: Motion descriptors classification.

color based approaches can be build on local or global features. We also introduce the temporal group which uses HMM to model data temporally. Note that HMM can be used with both skeleton data and depth features. In the following sections 2.2.3.1, 2.2.3.2 and 2.2.3.3 we introduce 10 different motion descriptors for action recognition representing popular strategies and choices made by researchers in this field.

2.2.3.1/ Skeleton-joints based Methods

A very common approach for human action recognition is to track human skeleton joints derived from depth maps [136, 111, 116]. The joint information is then used to model human activity directly using the joint positions or relative geometry between joints. Ding et al. [173] use Spatio-Temporal Feature Chain to represent the human actions by trajectories of joint positions. Vemulapalli et al. [160] proposed a skeletal representation based on the 3D relative geometry of the points described using the rotation and translation required to take one body part to the position and orientation of the other. A simple yet effective example of a joint-based activity recognition approach is the algorithm by Xia et al. [112]. A view-invariant posture representation was devised using histograms of 3D joint locations (HOJ3D) within a modified spherical coordinate system. The positions of the joints in time form a 3D spatial histogram. HOJ3Ds were re-projected using Linear Discriminant Analysis and clustered into k posture visual words. The temporal evolutions of these visual words were modeled by a discrete HMM. A sophisticated action recognition method based on skeleton joints and a probabilistic graphical model is presented in the work by [90]. The temporal evolution of human movements is represented by a hierarchical model with 2 layers: the first layer classifies sub-activities and the second layer forms activities on a more general layer. Joint-based methods are popular, but will fail if the initial joints were estimated wrongly, which is still an issue, especially in the case of severe self-occlusions. Moreover, if a very fine action is to be recognized, for example, a gesture or a slight movement, the joint-based methods lack precise information on shape and movement. For this reason, low-level attributes in depth images often outperform more high-level representations [128].

2.2.3.2/ Depth Maps based Methods

While high-level representation based methods are very popular, their main disadvantage is the difficulty to represent subtle motion, which leads to the research in low-level features based motion descriptors. In 2010 Li et al. [70] proposed a new depth-map based method build on local features. They use an expandable graphical model to explicitly model the temporal dynamics of actions and propose to use a bag of 3D points extracted from

a depth map to model postures. To select the points, they project a depth map onto orthogonal Cartesian planes and further sample a specified number of points at equal distance along the contours of the projections. These selected points are then clustered in order to obtain salient postures. A Gaussian Mixture Model (GMM) [40] is used to model postures globally by a distribution of points, and an action graph is constructed from the training samples to encode all actions that need to be recognized. This method gives better results than the 2D silhouette based action recognition [78], and the final descriptor is very compact. However, reported cross-subject activity recognition results are low due to the fact that the proposed sampling scheme is view dependent. Secondly, the descriptor loses spatial context information between interest points, which could be a problem when using the method in a real life scenario.

Simplified (not dense) motion-flow based descriptors could also be used [81, 144]. In this case either a rough motion estimation is used or a dense motion flow is calculated and then encoded into a more compact representation. Munaro at et al. [125] proposed a global descriptor which takes the direction and magnitude of motion of every body part into account. To do that, they first identify a human figure on a point cloud and then center a 3D grid around it. This grid divides the space around a person into a number of cubes. The flow information (direction and force) is extracted from each cube in the form of a mean and a summary of motion vectors. Motion flow is calculated by using the KD-search algorithm and a color distance in ?? space. A single descriptor is concatenated for every frame of a video sequence. The published results are good, however, the descriptor is not view-independent and the task of aligning the video frames in time is not addressed. Moreover, we can imagine the color information to not be of great use when the person tracked has solid color clothes, making it hard to establish point correspondences based on color similarity.

Hadfield et al. [144] proposed a novel local motion descriptor for v.2 RGB-D video sequences. The descriptor encodes the 3D orientation of flow vectors around established interesting points extracted from evenly spaced regions. The nature of the local motion field is described using a spherical histogram in a velocity domain: the contribution of each flow vector to the histogram is weighted based on the magnitude of the flow vector. To remove 3D rotation ambiguity, an invariance to camera roll is encoded. The direction of the flow vectors within sub-regions is made rotation-invariant. An interesting approach is also to perform PCA on the local region of the motion field, which finally leads to a descriptor which is invariant to all 3 types of camera viewpoint change, next to being robust to outlier motions. To obtain a global descriptor, a video sequence is divided into space-time blocks, each encoded independently to provide a final description of a sequence. This work proves that normalization and adaptation of the features so that they are scale and view point invariant improves the overall recognition of the system. However, the algorithm is applied on local interest points, discarding relational information about the movement. Sequence time block-division schemes can also lead to wrong results in the recognition step.

When dealing with depth maps, normals are an important source of information about the shape of an object. They were successfully exploited in many motion descriptors [162, 128]. The advanced HON4D (Histogram of Oriented 4D surface Normals) descriptor [128] is analogous to the histogram of gradients in color sequences and extends the histogram

of normals in static depth maps. The extended surface normal is then formulated by:

$$n = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial t}, -1 \right)^T \tag{2.1}$$

where depth is expressed by $z = f(x, y, t)$. In order to construct HON4D, a 4D space (XYZ, t) is initially quantized in order to get the grid representation using a regular 4D extension of a 2D polygon, namely a 600-cell Polychoron. This cell partitioning selection was not justified, but selected experimentally. Then in the 4D space a normal to the surface is computed and normalized. To form a final motion descriptor, researchers compute the corresponding distribution of 4D surface normal orientation with a 120 bin histogram. Videos are divided in parts and the final descriptor is a concatenation of individual descriptors. The results show that the motion descriptor outperforms several earlier descriptors [113, 111] However, as it is based on the change of depth, it does not take the 3D movement into account in the y and x directions. It was later outperformed by a similar normal-based method proposed by Yang and Tian [162], where the authors came up with a similar descriptor called Super Normal Vector (SNV). They used an adaptive spatio-temporal grid instead of regular grid partitioning followed by sparse coding to encode the features. Finally, aggregation was done using Fisher kernel [73] representation of the final feature vector. Both HON4D and SNV show very good results on available 3D datasets. The success can be explained by the use of an informational rich normal based feature. SNV also employs a temporal alignment scheme which improves the results a lot on datasets where actions have very different lengths.

A notion of hierarchy can be successfully employed in 3D motion descriptors. Kong et al. [179] improve the algorithm [128] using kernel descriptors alongside with the surface normals. They present a 3D gradient kernel descriptor which is a low-level depth sequence descriptor with the ability to capture detailed information by computing a pixel-level 3D gradient. The descriptor captures change in shape of the 3D surface in time in the following way: the 3D normals are computed and projected to a learned set of compact KPCA (modified PCA) basis vectors [24]. The kernel measures the similarity between orientations of the gradient for corresponding pixels from different patches of a video. A BoW method is used to build a hierarchical structure upon the low-level patch features to produce mid-level feature vectors. The efficient Match Kernel [59] is employed over the output of the 3D kernel descriptor to represent patch-based features which can be fed into the classifier. The method achieves state-of-the-art performance for action recognition using depth data, but it is computationally expensive. It also lacks an explanation of how exactly the correspondences between pixels in different frames of the video are estimated.

Spatio-temporal features based descriptors gained lots of attention recently [222, 195]. This group of methods usually localizes spatial features and describes their temporal evaluation. Yang Xiao et al. [161] proposed a 3D trajectory shape descriptor for unconstrained RGB-D videos. To extract the 3D dense trajectory feature, candidate feature points are first densely sampled in each RGB frame and tracked using optical flow. Then, by mapping the 2D positions of the RGB trajectory points to the depth map, motion information along the depth direction can be intuitively captured to form 3D trajectories. This method is a good illustration of how to enrich the 2D optical flow. To obtain the global representation of RGB-D videos, researchers combine several earlier approaches: Motion Boundary Histogram along the depth direction and trajectory shape descriptors [134] encoded using the Fisher Vector [73]. However it is not justified whether the estimations

of the 3D flow using the 2D and depth information give good results in general.

2.2.3.3/ Multiple Features based Methods

During the first years after the appearance of non-expensive depth sensors, researchers were more interested in depth information to characterize motion and rarely used information coming from the color sensor, apart from for sub-tasks such as motion flow computation or tracking. Recently, methods [222, 230] which fuse color and depth features gained popularity. Zhang et al. [222] propose discriminative and robust LST features named 4-D color-depth (CoDe4D) that incorporate both intensity and depth information acquired from RGB-D cameras. The feature detector constructs a saliency map by applying independent filters in the XYZt dimension to represent texture, shape and pose variations, and selects its local maxima as interest points. A Multichannel Orientation Histogram (MCOH) adaptive descriptor is applied on a 4-D support region of each interest point. Each support region is adaptive to linear perspective view changes. Then, the image gradients of color-depth patches within the support region are computed and quantized using a spherical coordinate-based method to form a final feature vector. The method is interesting, however, the use of the color and texture information in the descriptor can be a disadvantage in a general real-case intra-person scenario application. In this case it is proposed to tune the weighting parameters of the descriptor in order to weight the depth information more than the color one. Different weighting schemes are proposed for the selected datasets.

With the recent advances in human activity recognition, researchers also addressed the challenging task of a group action recognition. Zhang et al. [195] propose to use LST features which they call Adaptive Human-Centered (AdHuC) features. As in the previous method, their features are adapted to depth. To incorporate spatio-temporal and color-depth information in the XYZt space researchers use a cascade of three filters: a pass-through filter to encode cues along the depth dimension, a Gaussian filter to encode cues in XY space, and a Gabor filter to encode time information. Then the color and depth cues are fused to form a saliency map. Local maximums on this map form the LST features. The spatial-color-time based descriptor is then calculated for each point. The HOG3D [56] descriptor is modified in order to incorporate multi-channel information. The final descriptor has a histogram form and is concatenated of the per-channel descriptors.

The combination of skeleton data and LST features is a strategy recently employed by [245]. Each depth map is divided in cells by a regular grid. A Histogram of Temporal Gradient-Histogram of Oriented Gradient (HTG-HOG) descriptor captures the information of the current and neighboring frames in the following way: First, HTG is used to collect spatio-temporal information from each map; second, a modified HOG is applied on HTG over time. The HTG is simply the gradient magnitude and orientation of the temporal depth data calculated for each pixel. Then all the HTGs are aggregated to form a 2D array on which HOG is applied to capture the temporal changes and make an even more compact descriptor. The spatial relationships between skeleton joints and the descriptor obtained on depth map data are then used to find the key events in every action. Sparse coding and max pooling are applied to reduce the data dimensional and descriptiveness by deleting redundant cells. A Random Decision Forest is used to classify the data. The resulting image feature is compact, descriptive and noise-free. The recognition results on MSRAction3D dataset are high but lower than the results obtained with the HOG2

descriptor. Due to the powerful encoding scheme the descriptor is small but this also can affect its generality for different types of data. There are also many parameters to tune.

Table 2.3 summarizes the specifics of action recognition methods according to the proposed criteria.

2.2.4/ Motion Descriptors for Gesture Recognition

Nowadays computer applications require new ways of interaction, especially within the growing Virtual Reality domain. For that reason, human-computer interaction and particularly, gesture recognition, became a very popular field of research in the last few years.

A gesture can be defined as a physical movement of hands, arms, face and body with the intent to convey information or meaning [50]. Research in hand gesture recognition aims to design algorithms that can identify explicit human gestures. Gesture recognition dedicated motion descriptors are similar to the activity recognition ones with the difference that the descriptors should be capable of capturing very fine movements. Hand gestures are more difficult to recognize than body gestures due to the fact that the motions are more subtle. There are more degrees of freedom and serious occlusions occur between fingers. A detailed recent review on the advances in this area can be found in [188]. In this review, all the specifics of the gesture recognition task are discussed along with the proposed algorithms. However, this paper tends to capture all the steps of gesture recognition from detection up to classification, but is not particularly dedicated to dynamic gesture recognition using 3D motion descriptors. The major part of the methods reviewed do not exploit the 3D point cloud based gesture recognition, which became particularly popular after 2012, and dynamic 3D gesture recognition which emerged even later. For this reason, we also include here an overview of several motion descriptors applied to gesture recognition on 3D data.

Similarly to full body motion descriptors, many different types of visual features have been proposed for hand gestures. Early work used 2D information and build descriptors for a 2D silhouette of a hand. Due to the ambiguity of 2D data, the accuracy of such methods was not high. The latest dynamic gesture recognition methods use 3D depth information and its 2D counterparts.

Model-based methods are very popular. In contrast with full body action recognition, where skeleton estimation methods are well developed, none significantly superior to other automatic hand-joints position estimation algorithm are available. Scientists experiment with different hand models and approaches for static [157] and dynamic scenarios [218]. Obtained hand skeleton joints can then be used as features for the following classification step. An alternative approach is to characterize motion patterns from direct depth cues. Various geometric features can be extracted from a depth sequence [141]. For example, a shape silhouette can be used as a descriptor [84] or the cell occupancy information [76] similar to [110] can be used as a feature. This results in approaches which are less dependent on separate segmentation and tracking algorithms.

Recently, 3D dynamic gesture recognition methods similar to action recognition ones started to avoid the use of human body models and focus more on depth-based model-free features. When talking about gesture recognition, a Dynamic Time Warping (DTW) and its variations should be also mentioned. They are used to align two sequences based on preselected features. In general, DTW is applied at the classification stage and does not affect the descriptor design [198].

Cirujeda and Binefa [141] propose to use a Covariance matrix composed of the selected features from a 3D depth video sequence frame. Their descriptor doesn't use the absolute features themselves, but exploit representations of complex interactions between variations of 3D features in the spatial and temporal domain. It helps to make the descriptor robust to inter-subject and intra-class variations. The feature vector is the result of experimenting with several low-level cues and includes information about depth itself combined with other coarse observations such as the first and second image derivatives, gradient magnitude, curvature and temporal information. The idea of their descriptor is to measure how several variables change together, capturing the intrinsic correlation between distributions of the involved cues. The final sequence descriptor is a concatenation of the three scene-wise covariances in its vectorized form. The descriptor captures the global motion patterns. The final descriptor is scale-invariant, but not view-invariant. The method is easily generalizable for action recognition and outperforms [108, 128]. It is also independent of the sequences length and from the cluttered background. However, it doesn't explicitly use the information about the movement (i.e. force and direction), which still can be useful in action recognition.

Recently Ohn-Bar and Trivedi [184] proposed a combination of global low-level spatio-temporal features for gesture recognition in naturalistic driving settings. Their feature set combines other earlier features: Motion History Image (MHI) [30] and HOG features. Several descriptors are tested along with different fusion schemes to establish the better-performing ones. Moreover, only compact descriptors are used, so the resulting one is small in dimensionality. For this work, depth and color data descriptors were extracted separately and their performance was compared. The best combination is Extended HOG2 [151] + Extended MHI paired with Dense Trajectories and the Motion Boundary descriptor. The work shows that the approach to use a descriptor on the color and depth data separately and fuse them in the final step of the algorithm works very well. The resulting descriptor is also fast to compute.

Widespread use of deep learning techniques in the Computer Vision domain also touched 3D gesture recognition. In [221] researchers propose a complex NN architecture which helps to estimate the skeleton joints. Another NN is then used on the obtained skeleton data, depth frames and RGB frames separately. For a final descriptor, different fusion schemes for multi-modal data are proposed. A HMM is used to classify the data. Deep learning of the features on training examples of 3D hand gestures data gives very promising results, but a big amount of various samples is needed to generalize them to all datasets.

Table 3.2 presents the evaluation of reviewed descriptors in a more compact form.

Table 2.3: Motion descriptors for action recognition. **Reported accuracy is for MSR Action 3D dataset when available as reported by authors.* ***Accuracy reported for other dataset.*

Descriptor	Year	Locality	Dimensionality	View-invariance	Accuracy, %	Complexity	Classifier	Dataset
HOJ3D [112]	2012	local	mid(≈ 1008)	yes	low (78.97*)	low	HMM	MSR Action3D, custom (10 actions).
DS feature [245]	2017	local	low	yes	high (96*)	high	RDF	MSR Action 3D, MSR Daily Activity 3D, MSR Action 3D Pairs.
HON4D [128]	2013	global	high (≈ 22680)	yes	mid (88.89*)	high	SVM	MSR Action 3D, MSR Gesture 3D, MSR Actionv 3D Pairs, MSR Daily Activity 3D.
3D kernel descriptor [179]	2015	hierachical	high (≈ 13824)	yes	high (92.73*)	high	SVM	MSR Action 3D, MSR Action 3D Pairs, and MSR Gesture 3D.
Bag of 3D points [70]	2010	local	low	no	low (74.7**)	low	NN	Custom, 20 actions, complex combinations.
3D grid-based descriptor [125]	2013	global	mid (≈5760)	no	mid (87.4**)	low	NN	IAS-Lab Action Dataset (15 actions, 12 person).
3D Flow descriptor [144]	2014	local	low(≈ 144)	yes	mid(36.9**)	high	SVM	Hollywood 3D (14 actions, multi-cam setup).
3D Trajectories [161]	2014	global	low (≈ 96)	no	low (29.76**)	low	SVM	Hollywood 3D.
CoDe4D [222]	2016	local	high (≈ 21600)	yes	mid (86**)	high	SVM	Berkeley MHAD, $ACT4^2$, MSR Daily Action 3D, UTK Action3-D.
AdHuC [195]	2015	local	low	yes	mid (85.7**)	high	SVM	Berkeley MHAD and $ACT4^2$.

Table 2.4: Motion descriptors for gesture recognition and gait analysis. **Reported accuracy is for evaluated dataset when available as reported by authors. **Ground-truth correspondence.*

Descriptor	Year	Locality	Dimensionality	View-invariance	Accuracy, %	Complexity	Classifier	Dataset
4DCov [141]	2014	global	mid(3105)	no	high (92.89*)	low	Sparse Collab. Classifier	Gesture3D, Action3D, SKIG, WorkoutSU10.
Combined features [184]	2015	global/ local	low(≈128)	no	high (97.90*)	low	SVM	Cambridge hand gesture dataset.
STMD [148]	2014	local	low (≈120)	no	high (93.89*)	high	SVM, MLP	Custom, 22 individuals.
2.5D Gait Voxel Model [158]	2014	global	high	no	high($\approx 98^{*}$)	high	Simple distance metric	Custom, 100 individuals.
3D kernel descriptor [181]	2015	local	low	no	mid (86.5**)	low	n/a	n/a (MOCAP data).
HMM KF Leg tracking [186]	2015	local	low	no	n/a	low	HMM	custom, 6 person.

2.2.5/ Motion Descriptors for Gait Analysis

Lately, gait recognition and analysis from 3D data became popular. Observation of gait can provide early diagnostic clues for a number of movement disorders such as Parkinson's disease, cerebral palsy, stroke, arthritis, chronic obstructive pulmonary disease and many others. Depending on the field of research, different gait parameters are evaluated. Parameters of a gait could be divided into kinematic ones (such as knee flexing angle) and spatio-temporal ones (such as speed). A common approach is to experiment with different combinations of gait parameters and select the most representable for a given task as in the work of Agosti et al [224], where authors performed spatio-temporal and kinematic gait analysis for patients with Frontotemporal dementia and Alzheimer's. A general review on the subject of gait analysis is [212], whilst [150] overviews recent advances of skeleton-based gait recognition. In this literature review section, we provide information on the use of motion descriptors in 3D gait assessment and recognition tasks.

Descriptors for gait recognition commonly include the biometrics parameters, because intra-person variability is no longer an issue as in the case of action recognition. Motion information is a part of information used to describe a gait pattern. For this reason usually the motion descriptors used for 3D gait recognition are more simple and compact than the ones used for action/gesture recognition. Despite the fact that RGB-D cameras are popular tools in gait assessment tasks, it is quite common to use 2D projections in order to obtain a gait descriptor. Moreover, a vast majority of modern gait recognition and analysis methods perform a 3D-2D transformation of the depth sequence to form a final gait descriptor [148, 158, 106] or use 2D sensors directly [165]. The most well-known 2D gait descriptor is a Gait Energy Image [45], which is basically the average silhouette over one gait cycle. It was lately upgraded to a Gait Energy Volume (GEV) [88] by using information obtained from 3 Kinect sensors. GEV is derived by averaging all the voxel volumes over a gait cycle.

2.2.5.1/ 3D based Gait Descriptors

Similar to action and gesture recognition, 3D gait recognition methods can be categorized as methods based on skeleton joints [139, 123, 247] (model-based) and methods based on depth images [181] (model-free). Skeleton-based methods [247, 117, 68] are similar to analogue methods for action recognition: descriptors are based on the spatial and temporal position of the human skeleton joints or a human body model is used. Depth based gait features use detailed information about shape and depth variation of a walking individual and do not require model fitting. For gait recognition and analysis the use of HMMs (and more recently, LSTMs) trained on different features is a popular trend.

Kwolek et al. [148] proposed a view independent motion-based algorithm for gait recognition using a multi-camera setup. They use particle swarm optimization for full-body motion tracking. A 3D human-body model is also proposed in order to improve the results. The model is based on segmentation of body parts. The final descriptor is a gait signature composed of the distances between joints projected to a 2D plan and evaluated through time of a single gait cycle. This approach is interesting because it uses the multi-camera setup in order to fit a 3D human model. However, the accuracy of the 3D model is limited due to the use of 2D images without depth information.

Tang et al. [158] introduced a 2.5D voxel gait model that only includes a one-side surface

portion of the human body. A 2.5 gait model corresponding to a gait cycle is obtained from several Kinect depth frames. View-invariance is obtained by simply rotating the 2.5 gait model and synthesizing obtained views. The final descriptor is a color 2D image based on a combination of Gaussian and mean curvature [63] of the point cloud data. The method shows good results and avoids the high computational cost of 3D gait modeling. However, a 2.5D gait model cannot address the problems of the lack of robustness to covariates such as differences in appearance due to various clothes etc.

Lim et al. [181] propose a real-time model-based gait tracking and analysis method using a depth image sensor installed on a robotic walker. The particle filter is adapted to the depth camera video sequences to obtain the spatio-temporal gait parameters. Segmented leg regions of the point cloud are tracked using particle filtering, improved by implementing a simple harmonic motion model which corresponds to the human walking manner. To simplify the problem, each particle represents the predicted leg model part. Spatio-temporal parameter data (namely, physical parameters such as stride length) and gait phase can be deduced from the tracked leg pose parameters. The method is validated by a comparison with parameters obtained by a Motion Capture system. This paper proposes a computationally effective gait analysis algorithm suited for clinical gait assessment. However, a specific setup is considered and the proposed segmentation scheme might not work, when a person has a movement disorder.

Symmetry of the gait is an important criteria of a pathological gait. The following diseases could lead a non-symmetrical gait: cerebral palsy, stroke, hip arthritis and leg length discrepancy [168]. Therefore gait asymmetry can consequently be used to identify pathology and track recovery. In [168] authors propose a longitudinal index distinguishing asymmetrical gait and prove that index is correlated with the motion capture measurements. A Kinect camera is used to track a patient who is performing a walking test on a treadmill. The key concept of the proposed asymmetry index is to compare the spatial position of the left and the right legs at comparable times within their respective step cycle. Researchers use orthographic projections of the 3D surface of the subject to locate the lower limb zone using anthropometrical estimations of the human body parameters. Then the gait is temporally segmented into steps and then each step is divided into ten 10 intervals and a representative mean depth image MDI is computed for each interval. The mean image is actually an average depth image compensated for the lateral movements of the walking person and centered laterally. The symmetry of the movements is then defined by comparing the MDI from left and right foot (displaced laterally for a precise comparison). The experimental results show that this gait asymmetry index measured with a Kinect is low cost, easy to use and is a promising development for clinical gait analysis.

The design of a gait descriptor is strongly dependent on the specific application. For some applications the skeleton data based analysis is sufficient and some need more accurate shape tracking and description [168]. The 3D gait analysis is the least explored area among three reviewed which would benefit from more 3D motion cue description.

2.2.5.2/ HMMs AND LSTMs IN GAIT RESEARCH

Hidden Markov Models (HMMs) and Long-Short Term Memory nets (LSTMs) [23] create a model taking into account the temporal parameters of the data. HMMs and LSTMs are very popular machine learning tools representing a sequence of events, such as

action [110, 112, 209] or gait recognition [34]. HMM and LSTMs can be used for data classification but also for data encoding, so we propose a more detailed evaluation of their use in gait analysis since we believe the methods have a great potential in this field.

A HMM is defined as a double embedded stochastic process with an underlying process that is hidden (not observable). The hidden process can only be observed through another set of stochastic processes that produce the sequence of observations. An excellent tutorial for HMMs, their parameters, and their usage is [16]. This review provides only concrete examples of the usage of HMM in gait research. HMMs have long been used for gait recognition and analysis for the 2D data [34, 27, 107, 118] because of their statistical properties and their ability to reflect the temporal state-transition nature of gait. HMM models capture temporal shape and dynamics of a walking person and are used for the description or identification of gait, based on different characteristics extracted from the human silhouette such as the distances from the center to outer points [107], the width of an outer contour [34], LBP flow [118] and others. In 3D gait research HMMs are less exploited yet but there are some successful examples [186, 68, 153] and probably many more to follow since at the moment many successful 2D methods are re-designed in order to use 3D data. Papageorgiou et al. [186] use coordinates and velocities for the right and left leg obtained with a laser sensor that follows the subject motion to collect observation data for an HMM. A GMM is used to model data distribution. Lastly, another HMM is used for gait modeling and classification. Viterbi decoding algorithm is then used to find gait events which are hidden states in the HMM. Ngoyen et al model normal and pathological gait using hidden Markov model [211] based on the skeleton data from a Kinect sensor.

Lately LSTMs became popular and replaced HMMs in many applications. Both HMMs and LSTMs can capture transitional features along with the structural features extracted beforehand. HMMs are more applicable when the number of features is not excessive. That is why they are often applied on skeleton-based methods [247]. LSTM [23] recurrent neural networks can be used to learn the features for gait recognition or classification. Usually they are used in the same applications as HMMs, but lately researchers reported higher performance in traditional HMM domains, such as handwriting recognition [60]. Similar to HMM, there are more examples of their usage with 2D data, or data coming from wearable sensors [202]. Recently Feng et al. [201] trained an LSTM model on human joints data to characterize gait. An ?? is used to model temporal parameters of the gait sequence. The hidden activation values of the Neural Network represent the final gait feature. Researchers use 2D data, although the method can be easily extended to 3D.

HMM and LSTM based methods are very promising ways to describe the 3D motion data as they reflect the temporal states in actions, gestures and gait. There are many successful examples of their usage in all three fields covered by this review and probably there are many more to follow with the general tendency of the adaptation of 2D methods to 3D with the advance of 3D cameras. We choose to talk about HMMs and LSTMs in context of gait analysis due to the fact that this is the main area of our research.

2.2.6/ Conclusion 3D Descriptors

This literature survey reveals the progress made in the last decade in the field of 3D motion descriptors for point cloud data. The appearance of low-cost depth sensors in-

fluenced significantly research in action, gesture and gait recognition and analysis. 3D cameras are in a phase of constant improvement and development. Meanwhile, more and more applications are choosing to use them instead of conventional 2D cameras. The main issue for human motion based applications is still the constrained depth range imposed by the technology used. This problem is less relevant for hand gesture recognition applications but limits the usage of 3D sensors in gait analysis and action recognition research. The limited depth range of 3D sensors can be overcome by stereo cameras and there are several interesting solutions proposed (for example, ZED Camera) in this area recently. Therefore, in the coming years the number of applications using 3D data will only grow.

It is clear that 3D motion descriptors are progressing towards more general and efficient descriptors. However, fully automatic motion analysis and classification remains an open problem. General applicable motion description algorithms are sought after. The latest approaches are compact, transformation and view invariant to the target object and robust to noise.

According to published results, approaches that model spatial and temporal statistics holistically for point cloud data show less promising results than LST feature points and projection based methods. Model-based methods always have great potential but do not necessarily outperform the low/mid level features based methods as stated earlier in the application for natural images [119].

The main issue for many methods using low-level features remains the time needed to extract features from point cloud sequences. The features with the best recognition performance are often costly to compute. A detailed comparison of the time complexity of several popular descriptors is addressed in [184].

Perspectives

In action recognition, some latest methods tend to use multi-modal information to achieve better accuracy. It helps to improve the recognition results but introduces other difficulties, such a need of a synchronization and additional computations. Gesture recognition and gait analysis may also benefit from such additional information. Overall, gesture recognition while being the most prominent area for the 3D data usage in human motion analysis due to the limited range of modern 3D sensors is the least explored amongst the applications reviewed. There are many works focused on gesture recognition in 3D but they are mostly dedicated to static gestures and do not explicitly use motion cues. Since motion can add additional information helping to classify a gesture, we think that this direction needs to be explored further in the future.

Methods based on joint estimation usually provide compact and meaningful descriptors, and with the advances in skeleton joints recognition have great potential. Model-fitting methods remain popular for gait analysis and gesture recognition but for action recognition the main focus has shifted towards model-free methods. With the discontinuation of popular Microsoft Kinect sensors, there is a great need for a reliable open source skeleton joints estimation method. A skeleton is a compact, effective and reliable 3D representation of a human figure. The widely-known and used today skeletonization algorithms operate on a single image and do not consider the temporal information. Moreover, commonly used skeletons provided by the Kinect, are reliable only in a very small distance range between 1.5 and 2.5 meters and become error prone otherwise. All application areas can benefit greatly from a more reliable 3D skeletonization algorithm. The most common strategy for skeleton extraction at the moment is deep learning. While showing

good results and real time performance, once trained, such nets cannot take into account all the possible cases or abnormalities in the data. Therefore this data is not reliable for clinical applications. A promising strategy would be to return to classical skeletonization methods such as Median Axes [243]. For the deep-learning based methods, a promising direction is the open source package called OpenPose (based on [220, 244, 225]), which estimates the posture from 2D images. It demonstrates very good results and can be paired with a 3D camera in the future to enrich the initial 2D estimations. Without an accurate model-fitting method available, the model-free methods seem to be more prominent for all three areas.

Several methods [144, 124] reviewed use Scene Flow to encode the motion present in the video. Dense scene flow can give a complete information about motion present, but it is prone to errors, computationally expensive and not fully explored in the case of 3D data. Currently, the field looks on open-source real time algorithm estimating 3D scene flow from a point clouds sequence.

Overall, there is still a room for improvement in 3D motion description in all three reviewed areas. The most developed field in 3D human motion description is action recognition, with the biggest amount of research works published and a fair amount of datasets available. There is a need for an alternative 3D gesture and, especially gait datasets. The ideal dataset should contain data of different modalities and a significant amount of test subjects along with an established performance evaluation procedure.

3D data based applications have not been affected much by the deep learning based classification methods such as Convolutional Networks and Generative Adversarial Networks in the 2D domain. Some interesting works appeared recently [237, 240], and this research direction has potential, since 3D data analysis tends to follow proved trends from the 2D field.

Issues that must be addressed in future work in all three areas our opinion are: integration of all the cues for the better performance; computationally less expensive solutions; temporal alignment for a descriptor computation or classification stage; applications of LSTMs and HMMs for 3D features.

2.3/ SKELETON BASED GAIT DESCRIPTORS

The previous section is dedicated to 3D motion descriptors and only talks briefly about skeleton-based descriptors applied for motion analysis. As explained earlier, skeleton is a model of a human figure composed of rigid parts, connected by joints. A well-known algorithm to extract skeletons from depth images from 3D sensors exist [130]. Researchers often use skeleton data for different purposes, from action and gestures recognition to different medical applications. In gait analysis and recognition numerous skeleton-based research studies were published. For the RGB-D sensor based recent gait assessment studies, the skeleton oriented papers by far outnumber the ones using depth or color data directly.

This section focuses on the use of skeletons specifically for three gait-related applications: gait recognition, gait classification and gait assessment. To start with an example of skeleton data, Figure 2.11 provides with an example of a sequence of skeletons obtained by a Kinect v.2 sensor of a subject walking normally.

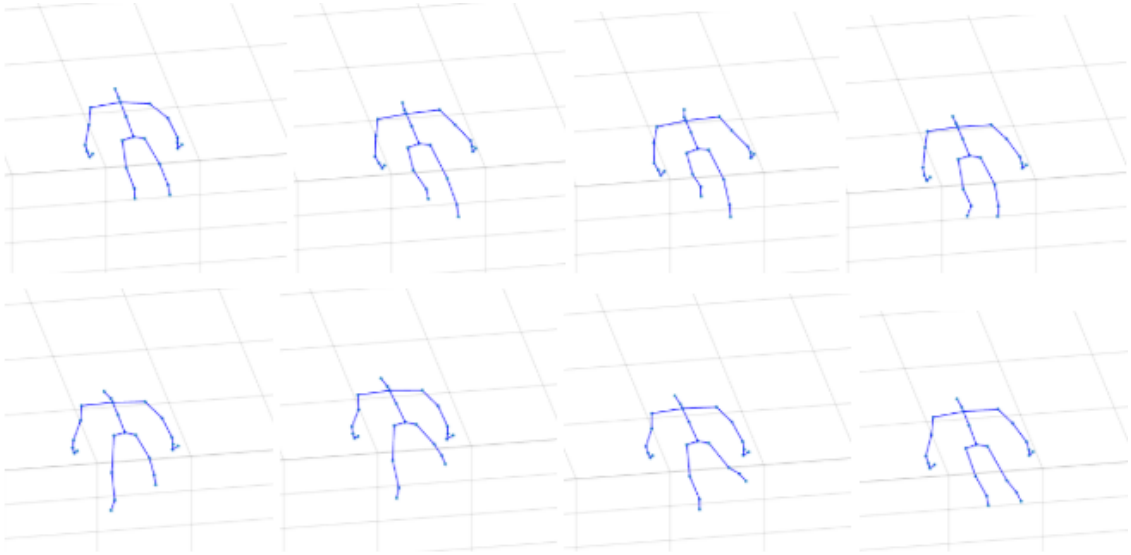


Figure 2.11: Several skeletons provided by a Kinect v.2 sensor of a person walking on a solid substrate from our MMGS dataset. Kinect v.2 provides with the joint coordinates, and the connected segments are added for the better visual representation. Frames are selected manually from a complete gait cycle.

2.3.1/ Abnormal Gait Detection

First step of gait analysis can be viewed as a general assessment of normal and abnormal gait. Abnormal or pathological gait deviates from normal gait patterns. There could be many different reasons for such a deviation, the most common are different neurological diseases. We introduced the pathological gait earlier in section 2.1.3.

Binary gait classification is a popular subject for researchers, who mainly use skeletons to build their frameworks.

Paiement et al. [153] target abnormal gait detection by analyzing the skeleton data. A non-linear manifold learning technique is employed to reduce the dimensionality of the noisy skeleton data. Then a statistical model is learned from the normal gait samples and the detection is performed based on matching the new observations to the model following Markov assumptions. This method is then applied to detect simulated gait anomalies on subjects, which are climbing a flight of stairs. Both event-based and frame-based classifications are performed. The event-based results are considerably better with 2 false alarms and 2 missed anomalies in a total of 58 sequences.

Chaaraoui et al. [169] perform abnormal gait detection with RGB-D devices using so called Joint Motion History (JMH) features. Skeletons are used to track motion over a segment of frames based on the three-dimensional location of their joints. Each skeleton (obtained by a Kinect v.2 sensor) is normalized for location, size and rotation. The skeletons are then accumulated using a sliding window approach. For each window, all skeleton joints are shown as 3D coordinates of a volume, and by the occupancy of this volume voxels the JMH is described. To reduce the size of JMH, researchers use a dimensionality reduction method based on axis projection. The BagOfKeyPoses is used to detect abnormal behaviour based on JMH features. It is a variation on the bag-of-words model and represents the most common feature instances present in the learning

data. The temporal relation is obtained via learning templates of sequences of key poses, which indirectly captures the gait events. The method is original, but it is complicated and doesn't identify the source of the gait pathology. The datasets used are not large enough to generalize the results.

Devanne et al. [199] analyze the motion trajectory for each step separately. Full body posture described by joints positions is used for the analysis. The trajectory made by the skeleton joints during each step is obtained by concatenating all the feature vectors with joints of the corresponding time intervals. A custom distance metric is then used to compare different gait sequences and remove the outliers. A statistical model of the normal distribution in a Riemannian shape space is build by analyzing shape variations among all samples belonging to left and right foot's steps. New gait sequences are then evaluated using the two learned step models. Experimental results demonstrate the effectiveness of the proposed approach in the context of asymmetric gait detection. However, the resulting trajectory feature relies on prior segmentation and can be affected by the arms motions and erroneous estimations of the feet joints.

In order to analyze the entire sequence, the work proposed by Meng et al. [210] employs a sliding window technique to explore the evolution of interjoint distances as spatio temporal intrinsic feature. The feature vector employed is composed by the all pairwise distances between the joints. Abnormal gait detection is performed by the Random Forest classifier. The results for the tests using Dai dataset [169] are good, but it is the only dataset used to validate the algorithm.

Nguyen et al. [211] proposed a normal gait model build on skeleton joints to detect abnormal human gait. Authors decompose each gait sequence in cycles. Each human instant posture is represented by a feature vector which describes relationships between pairs of lower body bone joints. Selected features are: hip angles, knee angles, ankle angles, and two feet angles. The angles corresponding to flexion angles (one for each joint) are calculated via two planes computed with the three joints of each leg. The feature vectors are then converted into codewords using a clustering technique. The normal human gait model is created based on multiple sequences of codewords corresponding to different gait cycles. A Left-to-Right Markov chain is used to model the skeleton data. In the detection stage, a gait cycle with a normality likelihood below a selected threshold, which was determined automatically in the training step, is assumed to be an anomaly.

The method shows state-of-the-art results, however, we are not sure that with the current difficulties for the Kinect sensor to detect low-limbs joints and, particularly, ankles and feet joints, it is sufficiently reliable to base a feature vector on them. Many researchers report the inaccuracy of feet position estimations and corresponding angles [228] coming from Kinect v.2. Available gait datasets [251, 204] also suffer from this issue.

Li et al. [250] proposed a method to classify normal and abnormal gait based on skeleton data. They are focused on the group of motion anomalies characteristic for Parkinson and Hemiplegia (tremor, partial paralysis, gestural rigidity and postural instability). At the heart of their descriptor lie 2 covariance matrices. One for 24 mutual joint positions and another one for motion rate of the same 24 joints. A K-NN based classifier and a custom covariance distance metric were used. Sequences were temporally segmented by a sliding window. A distinctive classifier was learned for the time windows to yield an importance weight assignment. The method gives good results with an average accuracy 80%. However, the custom dataset used is not publicly available. Although the full list of body joints is commonly used to compose a covariance matrix, we believe that in many

cases upper joint motion can be a source of ambiguity for the gait assessment task. In our work we propose to use compact covariance features based on low-limbs flexion angle data only. Later, in chapter 5.4 we choose to use angles over joints directly because they are more robust and vary less between individuals.

2.3.2/ Quantitative Gait Assessment

Gait assessment is strongly related to the binary gait classification. However, the difference is in the fact that gait assessment proposes qualitative evaluation, focusing on extracting different gait parameters and specifying the degree of deviation for new sequences from a gait model. The second stage is optional, since commonly only the extraction of subjective gait parameters is considered. When just parameters of the gait are examined, we can talk about quantitative gait assessment.

The most detailed work dedicated to the gait assessment subject by our knowledge is the one by Salbach et al. [189]. Researchers provide with a detailed overview of reference values and regression equations for time-and distance-limited walk tests and describe the methodology used to obtain them. The authors concentrate on ”normal” walk evaluation, i.e. not taking into account the ”abnormal” walk patterns due to patients disease etc. Multiple and simple linear regression are used to develop reference equations for walking patterns based on reviews of earlier studies on the subject. The study is mostly concentrated on the ”longer” walk distances studies (i.e. 30m-45m). The study shows the importance of different parameters in the task of gait assessment. A very general review of the gait assessment methods, devices and applications could also be found here [150]. With the increasing popularity of the depth sensors, many researchers use them for gait assessment tasks. A short review about the state-of-art depth-based methods including the reliability of the parameters obtained could be found in [217].

RGB-D sensors and the skeleton data they provide, are often used in gait analysis. Earlier studies have shown that the Kinect sensor can be exploited to examine gait. Such an analysis system based on skeleton joints was proposed by Gabel et al. [99]. Using regression models, authors successfully estimated the stride duration and arm angular velocity from the joints data. Rocha et al. [156] evaluated the possibility to use skeleton joints data from the Kinect v.1 sensor for gait assessment. First, the acquired sequences were temporally segmented in cycles. Then for each cycle, velocities, accelerations and intra-distances for the selected joints along with the flexion angles were obtained. In the next step, the mean, median and variance of these values were calculated, resulting in a total of 136 parameters. Then the most characteristic features for Parkinson patients were identified. The proposed research approach is interesting, however, the used test data was quite small with just 3 training and 3 testing subjects.

Eltoukhy et al. [229] evaluated gait parameters using a treadmill. Both kinematic and temporal parameters of the gait were extracted. Spatiotemporal measures included step length and width, step and stride times, vertical and mediolateral pelvis motion, and foot swing velocity. Kinematic outcomes included hip, knee, and ankle joint angles in the sagittal plane. Researchers used the skeleton data to extract all the parameters, and then compared them with the same parameters extracted by a marker-based system.

Overall, mostly researchers agree that RGB-D sensors and Microsoft Kinect v.2 in particular suit to the goal of gait assessment. In chapter 3 we provide with the detailed research on the Kinect validity for such an analysis. However, very few works propose the concrete

evaluation criteria, leaving this aspect for a human observer.

2.3.3/ GAIT RECOGNTION

Gait recognition is a popular direction in biometric studies. Very early researchers came to an idea of Computer Vision assisted human identification. Human identification is itself a big field of research with its roots going back into the early days of Computer Vision. Human recognition means establishing an identity of a person using different biometric features. Biometrics is the science of establishing the identity of an individual based on the inherent physical, psychological or behavioral characteristics associated with this person. A biometric-based system operates by capturing the person via appropriately designed acquisition module, extracting the interesting features and comparing them with biometric samples in the database in order to determine the identity of a person [177].

Gait recognition has received an increasing attention as a remote biometric identification technology, because it can achieve identification at a long distance, where other identification technologies can't work, and it needs little subject cooperation. There is a great number of works dedicated to 2D gait recognition, but the 3D cameras are more promising tools for this task thanks to the scale-invariance and 3D information. With the appearance of the Kinect sensor, many researchers based their gait identification works on it [166, 253].

The work [138] is interesting by the fact that it is designed to compare the marker-based and markerless solutions in terms of accuracy in the application to gait recognition. Researchers mount markers the ends of different parts of the foot. Knowledge of anatomical landmarks spatial location enables automatic calculation of anthropometric measurements necessary for joint kinetics while tracking the person movements [49]. Researchers conclude, that the marker-based approach is accurate and significantly outperforms the markerless method it is compared with.

For gait recognition, the static features and dynamic features can be used. Static features correspond to some physical parameters of a person such as height, bones length ratio etc, and dynamic correspond to the given movement pattern. We provide with some examples of existing gait recognition algortihms, while concentrating on the dynamic features used, since they are most interesting for the goal of our work.

It is quite popular to use the skeleton data as features directly [146, 178, 123]. Jiang et al. [146] propose to use static and dynamic features of gait cycle. Neural Networks classifier performs the final classification, however for the dynamic features the DTW algorithm is used to calculate the sum of cumulative minimum distance between a new sample and sequences in the database. Prior to feeding the features into the neural network, the dynamic and static features are fused together to guarantee the best classification score. Researchers report the Correct Classification Rate equal to 82 percent. Milovanovic et al. [123] used the coordinates of all the joints captured by Kinect to generate a RGB image, combined such RGB images into a video to represent the walking sequence, and identified the gait in the spirit of Content Based Image Retrieval.

Other researchers went a bit further and extracted meaningful gait parameters from the skeleton data. The most popular parameters for gait recogntion were low-limb kinematics [253, 166], but researchers also used relative joint distances [164]. For example, Sun et al. [253] use the length of some specific skeleton parts as the static features, and the

angles of swing limbs as the dynamic features to propose a view-invariant gait recognition framework.

2.3.4/ Conclusion Skeleton based Methods

Using machine-learning algorithms, the high-dimensional depth data can be reduced to 25 lower-dimensional three-dimensional body points. Skeleton is an extremely powerful representation of the human figure, allowing for all kinds of analysis, including the gait assessment. Our literature review demonstrates that such skeletons calculated from depth data from a Kinect camera can be successfully used for gait analysis and gait recognition. In the same moment, skeleton data from the Kinect can suffer from noise, which can limit the accuracy of the proposed methods. There is always a need of a simple gait assessment method robust to the presence of noise. In particularly, a full system, containing the reglamented acquisition setup, gait parameters extraction and assessment modules can be targeted. Despite a big number of research works using skeletons for gait analysis and recognition, it is hard to compare their results because it is a common practise to use a custom proprietary dataset.

2.4/ Conclusion

This chapter reviewed the main fields related to the subject of our research. First, we provided a review about normal and pathological gait and its' parameters. We addressed the task of gait assessment and listed different gait deviations. Gait analysis can be considered as an area of more general human motion analysis. Therefore, we considered 2 research areas for the conducted literature review. First, we studied 3D motion estimation and description using point cloud data. Second, we overview the use of skeleton data coming from the Kinect sensor in gait research.

We can say that skeleton-based descriptors are more popular in motion assessment applications than the ones directly using 3D information based on the strength of the literature review performed. As for the 3D data, the latest motion descriptors use it along with the color information for a better result. Gait-assessment algorithms are usually based on skeleton data and rarely employ the 3D data.

Based on the review of the state-of-the-art, we can highlight two main directions for the future research: gait analysis using skeleton data or the adaptation of more general 3D motion descriptors. Both categories were already exploited by researchers in different applications with promising results. In gait analysis, marker-based systems are still the most used due to their accuracy. However, we discovered many methods adopting an RGB-D camera to acquire the data.

We studied the specifics of the gait parameters and different diseases affecting the gait, with special attention to prosthesis gait. The most important parameters used by researchers and clinicians are kinematic based: low-limbs flexion and extension angles. The second important gait characteristic is speed. These parameters match the gait evaluation performed by Proteor. They use knee and ankle angle dynamics, velocity, force, and speed to evaluate the quality of a patient's gait. Therefore, we will target the estimation of these parameters for objective gait evaluation in our work as well. However,

it should be noted, that the final set of features should be general and not depend on the the biometric parameters of the gait.

An alternative approach can be to accurately capture human motion using a 3D sensor, followed by machine learning techniques to estimate the normal/abnormal gait factors empirically. This research direction is very promising, but requires a significant amount of labelled data.

In order to select an optimal research strategy, we first need to select the acquisition setup. The next chapter reviews the available 3D sensors, the specifics of their functioning, and the advantages and disadvantages of a single or multi-sensor setup.

3

Acquisition Platform

The previous chapter introduced gait parameters, 3D, and skeleton based motion descriptors. 3D data and simplified skeleton representations are most commonly provided by RGB-D sensors. In this chapter, we will select an optimal sensor package for our gait assessment tool, based on a literature review and the technical specifications of the 3D sensors available.

The acquisition platform selection and development is a core part of this work, as we aim to provide a complete gait analysis platform. We evaluated and tested different depth cameras available on the market in section 3.1. We studied the possibility to use a single camera or a multi-camera setup, and specified the pros and cons of each.

Once an acquisition device was selected, we study its validity in application for the gait assessment in section 3.2. We provide instructions on how to extract kinematic gait parameters from skeleton data. We study the possibility to use anterior-posterior direction (forward-backward displacement), the vertical direction (up-down displacement) and medio-lateral (side to side or from median to lateral) direction angles extracted from orientation data as a gait symmetry feature. we perform the validation of kinematic parameters between the camera and golden standard Vicon MOCAP system.

Section 3.3 describes the experiments performed with single and multi-Kinect acquisition setups.

Contents

3.1/ HARDWARE

The difference between 3D cameras and conventional 2D cameras is the fact that the latter provide scene structure information, or depth. The depth is usually provided as an image in which each pixel corresponds to the real world measurement of the distance from the viewed object to the camera. Such a depth image can be converted to a 3D point cloud when the calibration parameters of the camera are known. A 3D point cloud is a data structure containing points in a three-dimensional coordinate system. Points in point clouds are always located on the external surfaces of visible objects, because this are the spots, where ray of light from the scanner reflected from the objects. 3D cameras have great potential for human movement analysis in 3D, and there are several non-expensive solutions available on the market today.

For simplicity, in the following we would use the term 3D sensor/RGB-D sensor to talk about cameras using active sensing technology. Cameras delivering depth measurement can be classified as stereo information based and infrared based. The infrared group, which is also called active, refers to techniques that use a controlled source of structured energy emission and a sensor sensible to the emitted signal.

3.1.1/ RGB-D CAMERAS

We reviewed existing consumer RGB-D cameras available on the market in 2016 to select the most suitable for our research purposes. The list of the cameras reviewed is summarized in the Table 3.1. Devices are shown in Figure 3.1. We set the following requirements for the acquisition device: sufficient frame-rate to capture the motion, sufficient distance range to allow to capture a gait sequence, low price, technical support and reasonable resolution.

For gait assessment, only 4 out of 7 tested cameras satisfy the requirements set: the R200 from Intel, Orbbec Astra and Kinect v.1 and v.2. The information about selected devices is summarized in Table 3.1. Each camera has its advantages and disadvantages. The Orbbec Astra promises an excellent depth range, but is poorly supported and provides data containing many errors in depth estimation. The SR300 and F200 from Intel are small and portable, but have a slightly reduced depth range and are more suitable for 3D scanning. We also detected many holes in the single depth images provided by the two cameras. The Structure Sensor gives high quality 3D scans, but only when placed very close to an object. The Kinect v.1 and v.2 are massively supported and the most widely used, but have a slightly smaller depth resolution than newer Intel sensors. Overall, after testing all the devices in our laboratory, we selected the Microsoft Kinect family for usage in our project. The choice was based on the characteristics of the devices and the fact that the newer alternative cameras are not significantly better according to their specifications. The next section gives more details on the selected acquisition devices and their specifics.

3.1.2/ KINECT V.1 AND KINECT V.2

The Kinect is a markerless affordable technology introduced to public in 2010 by Microsoft. It is based on a software kit developed by Rare [10], a Microsoft Game Studios

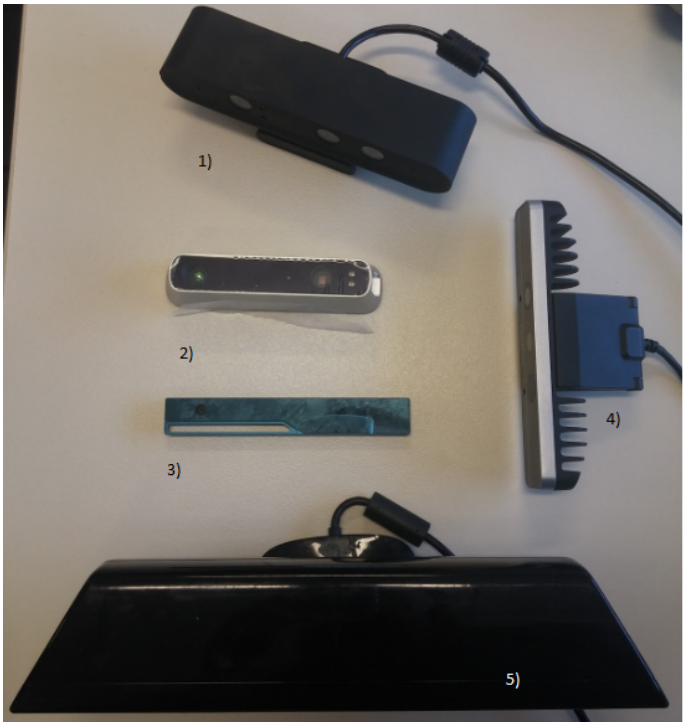


Figure 3.1: 3D sensors: 1) Orbbec Astra 2) Structure Sensor 3) Microsoft SR300 4) Microsoft F200 5) Microsoft Kinect v.1.

affiliated company, next to the active sensor technology of PrimeSense cameras. In the beginning Kinect primarily targeted the entertainment market. However, it was quickly adopted by the research community.

The Kinect is composed of two main sensors. The Kinect color camera uses an 8-bit Video Graphics Array (VGA) resolution, which corresponds to 640x480 pixels. It is possible to change the resolution up to 1280x1024, but with a lower frame rate. The monochrome depth sensor has a VGA resolution of 11 bits that allows 2048 sensibility levels. The depth sensor is composed of an IR projector and IR camera. The relative geometry between the IR projector and the IR camera and the projected pseudo-random IR dot pattern are known. This allows the reconstruction of the dots in 3D, using triangulation similar to stereo vision approaches. The Kinect provides per-pixel depth readings that are generally sparse. It has an approximate depth limitation from 0.7 to 6 meters. The horizontal angular field of view is 57^0 and vertical is 43^0 . The horizontal field of view has a minimum distance of around 0.8 meters and 0.63 meters in vertical, giving the Kinect an approximate resolution of 1.3 millimeters per pixel [98]. The bandwidth required for 30 fps is 21.8 MB/s of raw data per Kinect [97].

A Kinect device gives 3 kinds of images: an infrared, depth and color image. An example of those can be found in Fig 2.4. To use one kinect device for the estimation of 3D point clouds, the intrinsic parameters of its sensors and extrinsic parameters for the camera system should be known. For a coarse estimation the build-in camera calibration parameters are sufficient.

Kinect can be thought of as a 3D markerless motion capture system, because with the supplied software it is able to do simplified skeleton tracking in real time. No other special equipment is required. However, due to the nature of the depth sensor, the environment

Table 3.1: RGBD devices tested. **Maximum frame rate possible. Can be lower with increased color resolution.*

camera	frame rate (FPS) col/depth	depth resolution	angle (col/depth)	range, meters	color resolution	OS	software
R200	60*	480x360	77°x43°x70° 70°x46°x59° (Cone)	0,51 - 4	640x480	minimum windows 8	SDK Real Sense
F200	60*	640 x 480	77°/90°	0,2-1,2	1920 x 1280	minimum windows 8	SDK Real Sense
SR300	60*	640 x 480	N/a, should be similar to f200	0,2-2	1080p	windows 10	SDK Real Sense
Astra	30	640x480	60° x 49.5° x 73°	.4-8	280x960	Windows, Linux, Android	Orbbec Astra SDK + OpenNI
Structure sensor	60*	320 x 240	58x45	0.4-3.5	640 x 480	MAC OS, Windows	Windows SDK, +apps, OpenNI
Kinect v.1	30	320x240	57x43	.5-5.5	640x480	Windows 7 and up, Linux (non-official)	Microsoft SDK, OpenNI
Kinect v.2	30	512x424	70x60	.5-5.5	1920x1080	Windows 8 and up	Microsoft SDK, OpenNI

should still be controlled, i.e. the tracking should be performed indoor and within distance range of the Kinect. Further more, there are ample open-source libraries that work with Kinect data.

Although a Kinect camera provides many advantages, there are also some points and limitations, which we outline here. The first obvious disadvantage is the fact that the Kinect is not suitable for use outdoors due to the nature of its depth sensor. Secondly, reflecting surfaces and complex geometries may results in unpredictable random measurements and anomalies. This includes partially reflective clothing, and shiny shoes.

The distance range and the frame rate of the devices currently available on the market are not good for all the applications. In addition, the Kinect's v.1 depth estimates are very noisy and the field of view is not wide. One of the examples of consecutive depth frames acquired with the Kinect v.1 is shown in Fig 3.2. Since the depth data is not accurate, the resulting post processing algorithms also fail to deliver accurate joint estimations.

Figure 3.2: Consecutive frames taken with the Kinect v.1 and an applied body mask. The person is moving towards the camera, the distance is about 3.5 meters. There are many missing depth values, especially in the data originating from the right limb.

Despite its limitations, the Kinect sensor has become very popular in human movement analysis. It successfully eliminated the need for markers and calibration procedures, thereby enabling fast and patient-friendly 3D full-body motion registration.

Gait acquisition experience with the Kinect v.1

A small database with ordinary people walking, was build using the Kinect v.1 sensor as the main capturing device. It served as a testing bed for both the somewhat older sensor and our motion flow estimation algorithms.

Eight persons performed five walking trials in front of the Kinect camera, placed perpendicular to the walking direction. The resulting gait data was segmented in cycles, and the corresponding motion flow was extracted from RGB and depth images with the PD Flow algorithm [176].

An example of such data from our experiment is shown in Fig 3.3. Due to the PD flow algorithm design, it delivers erroneous results the moment when depth data is missing from consecutive frames. Such errors were omnipresent in the acquired sequences. The skeleton data obtained from depth images with an OpenNI based application was also visually estimated as not sufficiently reliable for gait analysis. Therefore, the older Kinect sensor was retired in favour of the more modern Kinect v.2 platform.

In 2013 the second generation of the Kinect camera was released as part of the Xbox One gaming console, which was later also introduced as a stand alone version for Windows PC. The Kinect v.2 has several advantages in comparison to the Kinect v.1. Next to an improved resolution, it uses a novel depth acquisition technique based on time-of-flight (TOF). It measures the distance between points by computing the phase-shift distance of modulated infrared light. The intensity of the captured image is thus proportional to the distance of the points in 3D space. The TOF technology, as opposed to structured light, inherently provides a dense depth map.

However the results can suffer from various artifacts caused by the reflections of the light

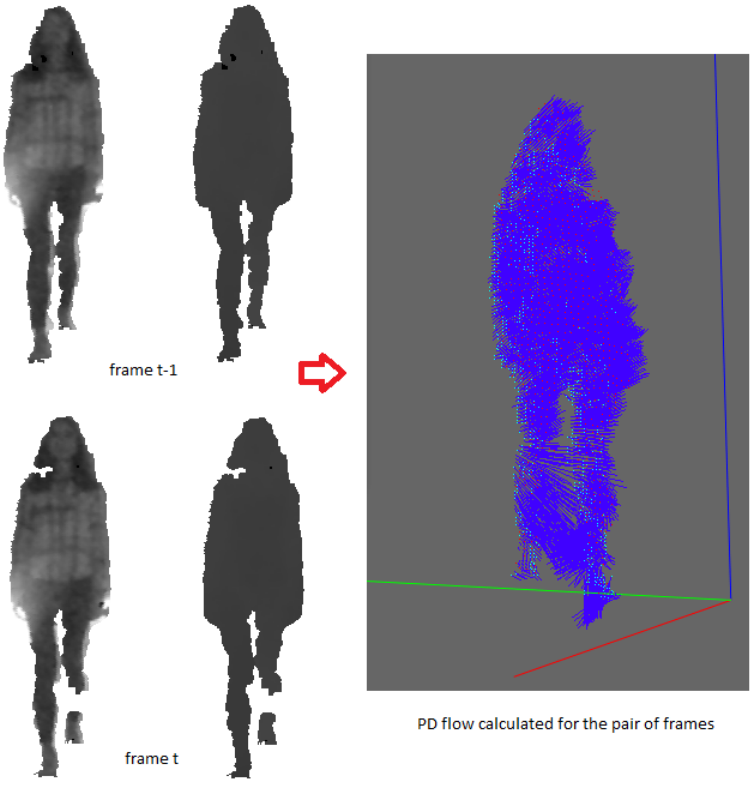


Figure 3.3: Data from the Kinect v.1 and resulting erroneous Motion Flow. The algorithm matches points from the left foot to the right foot. Left: color and depth from consecutive frames of a person walking used as an input to PD Flow. Right: Resulting PD Flow with wrongly estimated motion vectors.

signal from the scene geometry and the reflectance properties of observed materials. A comparison of the main parameters of the two cameras is summarized in Table 3.2.

Table 3.2: Characteristics of Kinect v.1 and v.2.

Characteristic	Kinect v.1	Kinect v.2
Color Resolution	640x480	1920x1080
Depth Resolution	320x240	512x424
FOV	57x43	70x60
Joints	20	25
Max Depth Distance	4.5 m	4.5 m
Min Depth Distance	0.5 m	0.5 m
Frame rate	30fps	30fps
Method to calculate depth	Structured Light	Time of Flight
Multi-kinect	yes	no

Despite the significant improvements, there are some known issues for the Kinect v.2, especially related to the depth data. Images appear jittery and fuzzy around edges. This is because it's not representing a visual feed, and instead false colouring calculations of time. These have a wide margin of error when done real time. This fuzziness of the image increases at large distances. Although the depth acquisition is done based on a different principle, the newer Kinect also remains highly sensitive to even minor reflective surfaces.

3.1.3/ Human Skeleton Joints

The Microsoft official SDK provides body skeleton joints, located from a depth image using machine learning based methods, such as [87]. Skeletonization was briefly mentioned in Sections 2.2.1 and 2.3 of this dissertation. Here we come back to this subject and provide more details focusing on the particular sensor. We address the skeleton data provided by Kinect v.2 sensor and corresponding Microsoft SDK.

Using skeleton tracking, the Kinect sensor can track the human body with various joint points. There are 25 joints mapped for each body. Each one gives a position and orientation. Figure 3.4 shows the estimated joints provided by the Kinect SDK. Skeleton data contains the skeleton joints 3D coordinates (X, Y, Z), their state (i.e., *tracked, inferred, not tracked*), and joint orientations (q_w, q_x, q_y, q_z). The Kinect v.2 can track up to 6 persons simultaneously.

The Kinect v.2 has more robust and more accurate tracking of human pose as compared to Kinect v.1 [191], moreover the sensor tracks more joints. Thanks to the improvement of the camera characteristics, the Kinect software SDK now also receives features such as enhanced hand-tracking capabilities, more detailed facial landmarks and better skeleton tracking. The main interest for our work is the improved skeleton tracking. The skeletal tracking method implemented in Kinect SDK v.2.0 has not been fully disclosed by Microsoft. However, it appears to follow similar methodology as for Kinect v.1 while taking advantage of GPU computation to reduce the latency and to improve the performance. In their review, Wang et al.[191] conclude that overall Kinect v.2 gives more accurate skeleton data in comparison to the first generation of the Kinect camera. Researchers [191] perform a side-by-side comparison of the two cameras using a marker-based optical motion capture system as the ground truth. Some researchers report that the new improved hardware of the second version can enhance the results of automated tracking of anatomical landmarks needed for movement analysis [182].

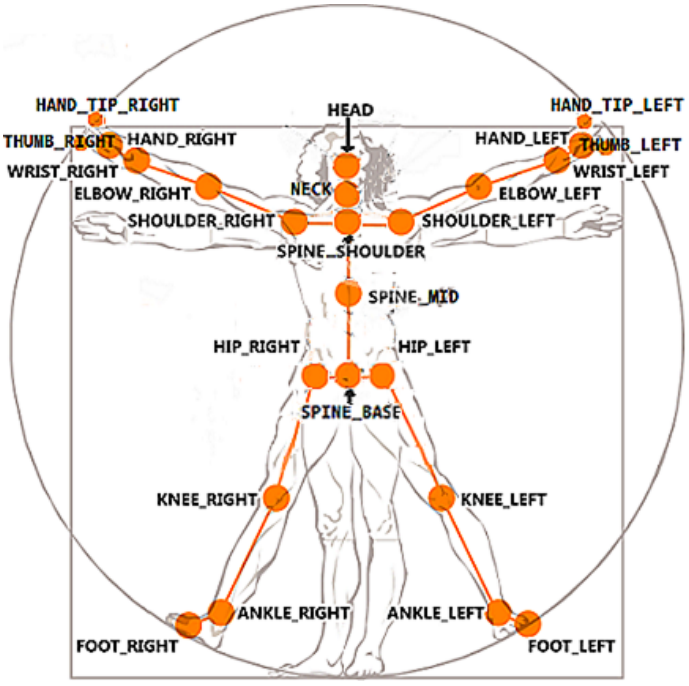


Figure 3.4: Skeleton joints estimated by the Kinect v.2 and Microsoft Kinect SDK. Image is taken from the Microsoft official website.

3.1.4/ KINECT VALIDITY

The Microsoft Kinect might provide a low cost gait analysis method, and researchers do explore its validity for the estimation of common gait characteristics. Before exploiting the Kinect camera in our work, a literature review was conducted to prior ascertain its capabilities for movement and gait analysis. We include here both information concerning the Kinect v.1 and v.2 sensors but concentrate on the later version. In most works related to the subject, researchers only use the skeleton information extracted from Kinect depth images. We discovered only one study where researchers use the RGB-D data from the Kinect and not the skeletons and report its accuracy. The authors in [154] propose a foot-tracking method to be used for a virtual reality application, and compare the data with their motion tracking marker-based system, concluding that the data from the Kinect depth images are reliable enough for this task.

Recently, [217] evaluated the literature describing the concurrent validity of using the Microsoft Kinect as a gait analysis instrument. Researchers came to the conclusion that the Kinect is valid only for estimating a subset of spatiotemporal gait parameters. Kinematic parameters measurements show large errors and are not consistent enough for clinical assessment. However, in the same year, in a very detailed report by [213], Otto et al. conclude positively on the possibility to use the Kinect v.2 for motion assessment for a healthy population. Nonetheless, they mention that signals of feet and ankles had the lowest accuracy according to the mean 3D distance and correlation analysis in XYZ dimensions in comparison to the golden standard measurement. According to the research, these landmarks are very unstable and thus not reliable.

Several studies have demonstrated that spatiotemporal gait parameters can be reliably obtained using a single Kinect v.1 sensor [143, 115] and lately using a single [182, 217] and multi [175, 191] Kinect v.2 sensor. Researchers attempted to both measure the accuracy of Kinect landmarks and gait parameters extracted from them. Since kinematic and spatial parameters estimated from Kinect data seem to have different accuracy levels, we treat the findings for the two groups separately.

Landmarks and spatio-temporal gait parameters accuracy

Clark et al. [115] estimated that gait speed, step length, stride length and landmark location linearity from the Kinect and a marker-based system exhibited excellent similarity. The foot swing velocity, step time and stride time possessed excellent and modest relative agreements, but modest overall (we assume the authors mean absolute) agreement.

Contrarily, later research by Mentiplay [182] has shown an excellent step time similarities for the Kinect v.2 sensor and Vicon 3D motion analysis. [191] measure mean offsets for all the joints for Vicon, Kinect v.1 and Kinect v.2 skeletons and report a big offset for the low-limbs joints for both camera versions, especially on the side of the skeleton that is turned further away from the camera as the occlusion of joints increases the uncertainty of the pose detection. The orientation of the feet is stated as unreliable. Geers et al. [175] performed a kinematic validation of a multi-Kinect v.2 instrumented 10-meter walkway for quantitative gait assessments and concluded that the Kinect values correspond well to the Optotrac [7] active-marker 3D optical tracking values.

The between-systems agreement was examined for raw data, i.e., body point's time series, and spatiotemporal gait parameters, e.g., step length, cadence and step width. In addition, the between-systems agreement for the performance measure of the 10MWT, i.e., time to walk 10 meters, is compared. The test showed good to excellent alignment in body point's time series between the two motion-registration systems according to intraclass correlation indexes [22]. Later Geerse at al. extended their previous work to validate foot placement locations from ankle data of a Kinect v.2 sensor in comparison with a golden standard system [231]. Researchers found significant between-systems differences in foot placement locations and step lengths in a distance of 2 meters. This detected effect was likely caused by differences in body orientation relative to the Kinect sensor.

Kinematic parameters

Kinematic parameters are popular amongst clinicians but rarely assessed by researchers. Re-

cently, Guess et al. [232] performed a comparison of 3D joint angles measured with the Kinect 2.0 skeletal tracker versus a marker-based motion capture system and concluded on a good agreement between both systems for knee and hip flexion. All test subjects were free from musculoskeletal pathology. The movement parameters were examined on the data squired in the experiment where subjects perform a drop vertical jump. The test subject was always placed on an optimal distance from Kinect, which the researchers established to give the most accurate results.

Analyzing the findings made earlier, it can be said that the Kinect can be used for gait analysis if all its limitations are taken into account. A limitation of the Kinect v.2 is the relatively low sampling frequency of 30 Hz and the quality of the depth image in the distance. Despite its limitations, the Kinect provides many advantages over other sensor-based devices for examining gait. It is low-cost, widely available and does not require any markers or sensors to be attached to the body, which significantly improves clinical feasibility and reduces testing time. By simply mounting a Kinect to a set position in a clinic, spatiotemporal gait characteristics can be assessed quickly and easily using automated analysis algorithms. Although the conclusions on the Kinect v.1 accuracy vary, researchers [175, 231, 182, 217] agree that the Kinect v.2 can be used for clinical gait parameters assessment. We summarized the weak points of the skeleton based estimations detected:

- Kinect's estimate of the ankle position is not reliable and seems to gradually change during the gait cycle in relation to the distance from the sensor [231].
- Foot placement locations are not very accurate [231, 213].
- Gait kinematic parameters measurements estimated from skeleton data are not reliable. [217, 182]
- Ground contact time is not reliable [182]

Overall, the Kinect v.2 can be used to assess the main spatio-temporal parameters, but is not sufficiently accurate for kinematic parameters. The main erroneous spatio-temporal estimations originate from ankle and foot joints. We provide an additional assessment of the kinematic parameters further in this work in section 3.2.3.

A single sensor is capable in capturing the gait parameters, but an ideal setup is required to get the correct skeleton locations with the ML method used. In particular, the Kinect should face a patient. In all the other cases, the distant body parts may be occluded and inferred. When the used joints are inferred, extracting kinematic features from lower limb angles as gait features is inaccurate. A possible solution to this problem is described in [191], where authors use the angle-based features from both sides using two Kinect devices simultaneously. The 3D data does not suffer from this issue.

From a clinical perspective, factors that mitigate against the use of a gait analysis tool in clinical settings are the lack of availability, reimbursement, and training [92]. The use of an accessible depth sensor can overcome these issues. A gait assessment using a Kinect v.2 camera is then viable, while taking into account its limitations and specifications.

3.1.5/ Multi-Kinect vs Single Kinect

The second question we faced was the possibility to use a multi-sensors setup instead of a single Kinect camera. A single Kinect provides information from only one side of the objects facing the sensor. In case when 3D data is used, a multi-kinect setup can be exploited in order to capture complete volume information.

The Multi-Kinect setup can also be very interesting in the case when a treadmill is used. Then with two or three cameras the 3D figure can be restored, similar as was done in earlier work in our lab [174]. A single Kinect, however, often has to be placed on the side or at an angle due to the

fact that modern treadmills commonly have a control panel in the front. These panels block and distort the measurements beyond recovery, rendering them useless.

On the other hand, the delivered skeleton data is less reliable when the camera was places to a side. The skeleton data is estimated using the depth image, and, depending on the view point, the pre-trained machine learning model performs differently. A solution is to use depth data from several depth sensors capturing the scene simultaneously from different sights. In this scenario the depth data from the cameras should be used. Alternatively, several sensors can be used to increase the acquisition volume. In this scenario, the Kinects can be switched when the person is leaving the field of view of one camera and enters the next one.

According to the Kinect for Windows SDK 2.0, individual machines are required for each individual Kinect v.2 device. Although other libraries, such as libfreenect2 [6], allow to connect two devices, it decreases the frame rate and it also does not provide the body skeleton data, which could be necessary for some applications. We tested several available solutions and found the frame rate unsatisfying for gait assessment.

The second solution is to use two separate machines to acquire the data and then to fuse two resulting point clouds in a post-processing step. We study such a setup later in this work in Section 3.3.2.

3.2/ KINEMATIC GAIT PARAMETERS FROM KINECT

The goal of our research work is a clinical gait analysis system which can be an alternative to expensive golden standard systems. Many researchers have evaluated spatial parameters of the gait calculated from Kinect's data. A general conclusion was the moderate to excellent reliability of the Kinect for the estimation of the most popular gait characteristics. However, to our knowledge, no one assessed the full set of kinematic gait parameters.

In this section, we assess the validity of the Kinect sensor for kinematic parameter estimation, comparing it to a golden standard Vicon camera. We also give a recommendation on how to obtain the flexion angles from Kinect orientation data.

3.2.1/ EXPERIMENTAL DESIGN

This test was performed on the platform of the CHU Dijon, where a Vicon camera is employed to perform movement analysis. A single experienced researcher placed retro-reflective markers on every participant. All makers were placed directly on the skin for higher accuracy. The acquisition setup is shown in Figure 3.5. The person was walking towards the Kinect from a distance of 6 meters and stopped at 1.5 meter in front of the sensor. The Kinect detects a person approximately from a distance of 5 meters, when he is walking normally. This gives us an average of about 2-3 meters of walking distance. All subjects walked bare-foot. They were instructed to perform 11 trials.

The acquisition process was started simultaneously with the Kinect and Vicon. Marker trajectories from Vicon were saved with a sampling frequency of 120 Hz. The Kinect data frequency levelled at 30 Hz. We later downsampled the data from the Vicon camera to have the same frequency.

The Vicon and Kinect v.2 software give a different set of markers. In addition, all Vicon markers are located on the body surface, and the skeleton joints returned by the Kinect are 'shifted' inside the imaginary body of a person closer to where the real bones are supposed to be. Figure 3.6 shows the Kinect and Vicon skeleton joints. We selected 20 joints from both systems for further analysis. Table 3.3 describes the mapping of the Vicon markers to the nearest Kinect landmarks. The marker name abbreviations are given as specified by Vicon and Kinect official software.

The body points, represented by Vicon's virtual markers were selected to closely match the

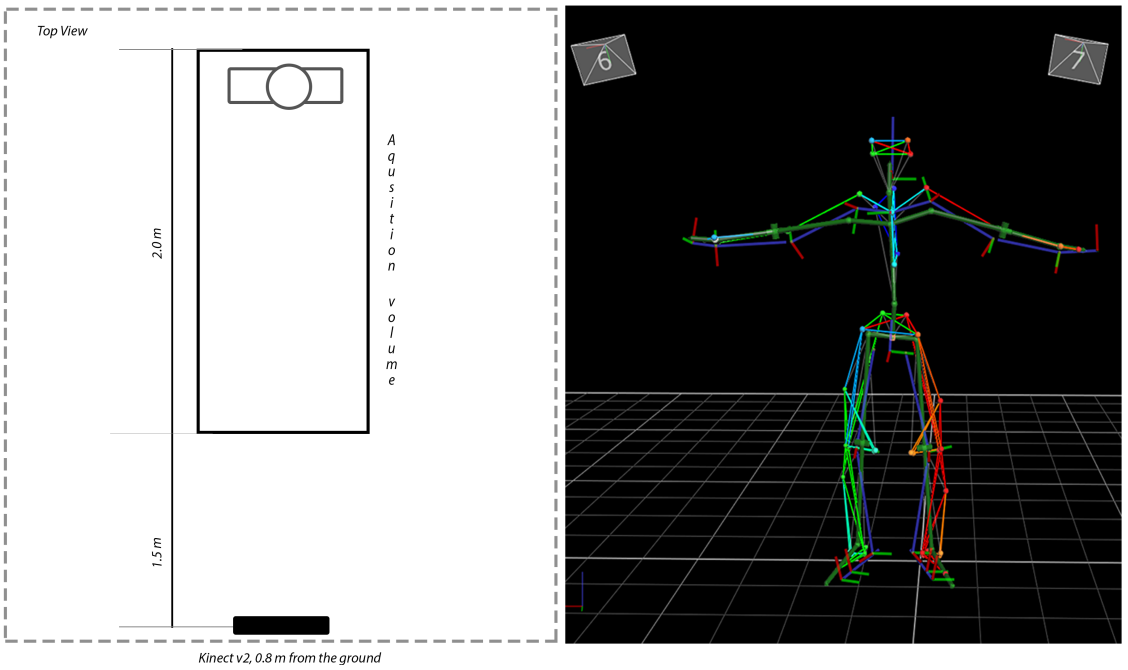


Figure 3.5: Acquisition setup and a skeleton figure provided by Vicon. The skeleton is formed by the lines connecting markers, attached to different body part. The colouring of each element is provided by Vicon software for a better visibility.

Table 3.3: Joints used for the comparison test between the Kinect v.2 and the Vicon.

nbr	Joint	Kinect	Vicon
1	Head	HEAD	RFHD + LFHD
2	Spine-Shoulder center	SPINE SHOULDER	CLAV
3	Spine-Middle	SPINE MID	(STRN+T10)/2
4	Spine-Base	SPINE BASE	(LPSI+RPSI+LASI+RASI)/4
5	Left Hip	HIP LEFT	LASI
6	Right Hip	HIP RIGHT	RASI
7	Left Knee	KNEE LEFT	LKNE
8	Right Knee	KNEE RIGHT	RKNE
9	Left ankle	ANKLE LEFT	LANK
10	Right Ankle	ANKLE RIGHT	RANK
11	Left Foot	FOOT LEFT	LTOE
12	Right Foot	FOOT RIGHT	RTOE
13	Shoulder Left	SHOULDER LEFT	LSHO
14	Shoulder Right	SHOULDER RIGHT	RSHO
15	Elbow Left	ELBOW LEFT	LELB
16	Elbow Right	ELBOW RIGHT	RELB
17	Left Wrist	WRIST LEFT	LWRA
18	Right Wrist	WRIST RIGHT	RWRA

Kinect's body points, although sometimes arbitrary positional differences between the body point's time series of the two motion-registration systems are possible. Both Kinect and Vicon markers values were converted into meters. We also did a standard normalization by subtracting the center of spine joint from the all frames of a skeleton sequence.

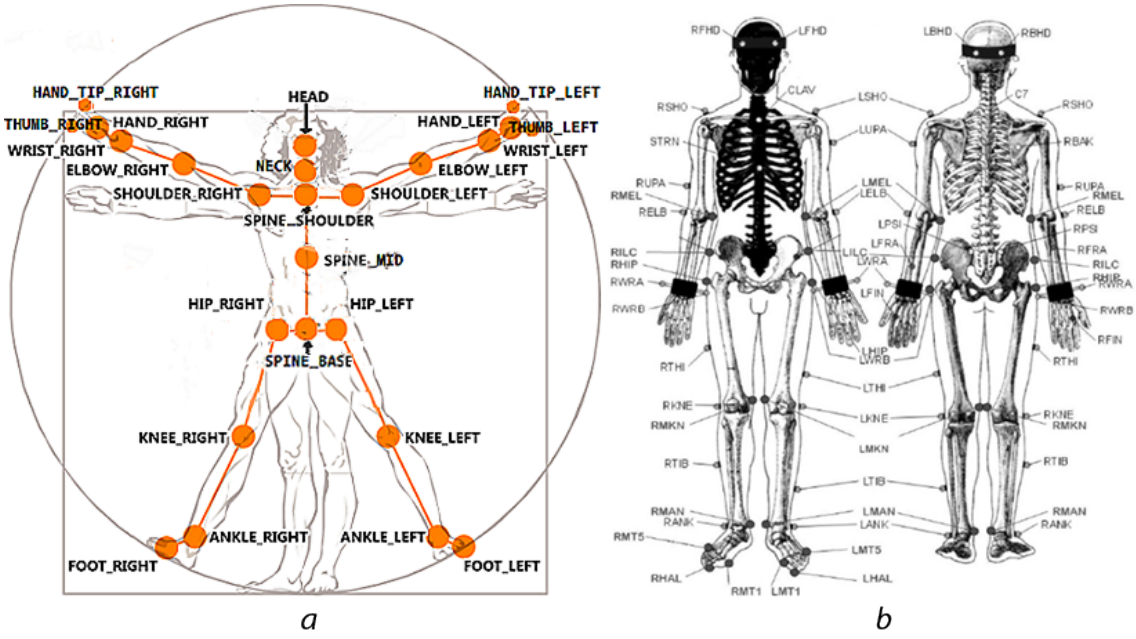


Figure 3.6: Skeleton joints provided by the Kinect v.2 (a) and the Vicon (b).

3.2.2/ KINEMATIC GAIT PARAMETERS FROM VICON

First the data from the Vicon and both Kinects was manually segmented into cycles using the method described earlier in section 2.1.1. Nguyen et al. [211] segment the cycles by identifying corresponding intra-feet joints distance peaks for the Kinect and Vicon. We simply used the right foot marker Z coordinate from the Kinect and Vicon, and obtained peaks needed to divide the gait sequence in cycles. An example of the resulting data is shown in Figure 3.7.

Low-limbs flexion angles are common features in gait analysis. Clinicians are assessing flexion, rotation and abduction angles to evaluate a patient's gait. For the reference, the report on the gait assessment for one of the Proteor patients is given in Annex A .

Earlier studies have shown that such kinematic parameters obtained with a Kinect camera are not reliable [217]. Nonetheless, Guess [232] compared the low-limbs flexion, extension and abduction data obtained via Kinects' v.2 quaternions and Vicon. The comparison has shown that reasonable measurements of hip and knee flexion angles, as well as isolated hip abduction angles, can be obtained from the Kinect skeletal tracker for a person performing a jumping exercise. Inspired by [232], we adopted the proposed method to calculate low-limbs joint angles from Kinect v.2 orientation information. Commonly, researches [211, 250, 153, 156] only use Kinect joints data, whilst the quaternions are mainly used for avatar animation. Yet quaternions can also be used to obtain limbs flexion angles according to ISB recommendations [203]. Here we use quaternions to obtain the angles that match the Plug-in Gait kinematics data coming from the Vicon software.

Kinect provides the joint orientation values in the form of quaternions. A quaternion is a four-element vector composed of one real element and three complex elements. Quaternions are commonly used to encode any rotation in a 3D coordinate system.

A 6D pose can be described as a displacement in 3D plus a rotation defined using a specific case of Euler angles, which can be represented via quaternions. Quaternions find uses in applied mathematics, in particular for calculations involving three-dimensional rotations such as in three-dimensional computer graphics, and computer vision. Euler Angles are limited by a phenomenon called "gimbal lock," which prevents them from measuring orientation when the pitch angle approaches +/- 90 degrees. Quaternions provide an alternative measurement technique that does not suffer from gimbal lock [12].

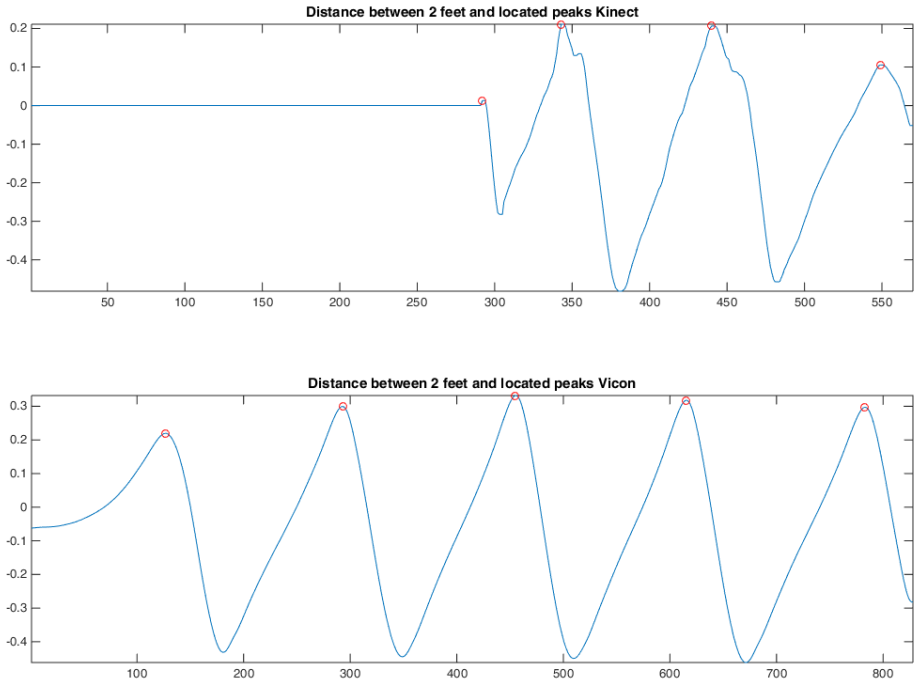


Figure 3.7: Evaluation of the Z coordinate of the right foot marker for the Kinect and Vicon. The Kinect detects the patient from a distance of 5 meters, hence the zero Z values in the first frames.

A quaternion is a set of 4 values: x, y, z, and w. Using the orientation quaternion, we can calculate the rotation of the joint around the Z (ϕ), X(ψ), and Y(χ) axis. This yaw, roll, and pitch correspond to flexion, abduction, and rotation. A useful interpretation of quaternions is that of a rotation of θ radians around the axis defined by the vector v= $(vx, vy, vz)(qx, qy, qz)$. Figure 3.8 demonstrates the angles for the knee and hip used later.

The Microsoft Kinect SDK version 2.0 provides us with the orientation of a bone in a 3D camera global reference system of coordinates. The orientation of the bone is relative to the child joint while the hip center joint still contains the orientation of the person.

Therefore, we followed the recommendations of [232] to calculate the flexion, extension and abduction angles for the low-limb joints. We calculate the knee and hip angles of a walking person via quaternions in the follwing way. First we re-orient the hip quaternion to match the Vicon hip orientation as advised in [232]. The initial orientation given by Kinect quaternions and the re-oriented hip quaternions are shown in Figure 3.9 a.

To calculate the knee joint angles, we need to use the knee joint and its child: the ankle joint orientations. These orientations are q_k = $(w_k + x_k i + y_k j + z_k)$ for the knee quaternion and q_a =$(w_a + x_a i + y_a j + z_a k)$ for the ankle quaternion. The quaternions provided by Kinect 2.0 SDK satisfy the definition of a unit quaternion:

$$N(q) = 1; N(q) = N(w_0 + x_0 i + y_0 j + z_0 k) = w^2 + x_0^2 + y_0^2 + z_0^2 = 1 \tag{3.1}$$

Each rotation, q_k and then q_a is done in the absolute frame of reference so we use $q_a^{-1} \times qk$ to

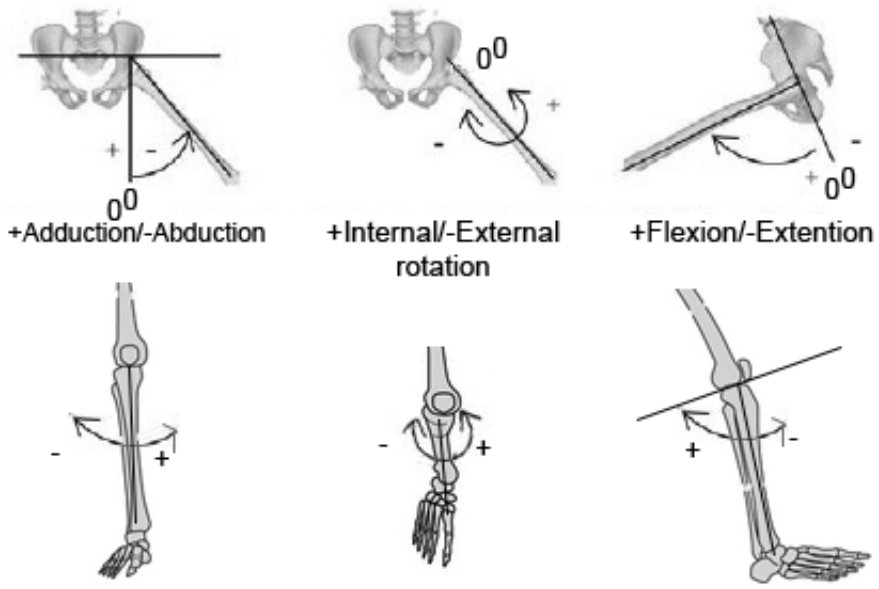


Figure 3.8: Upper row: hip abduction, rotation and flexion angles. Bottom row: knee abduction, rotation and flexion angles. Adduction refers to a motion that pulls a structure or part towards the midline of a limb. Abduction refers to a motion that pulls a structure or part away from the midline of the limb. Internal rotation refers to rotation towards the axis of the body, external rotation refers to rotation away from the center of the body. Flexion describes a bending movement that decreases the angle between a segment and its proximal segment.

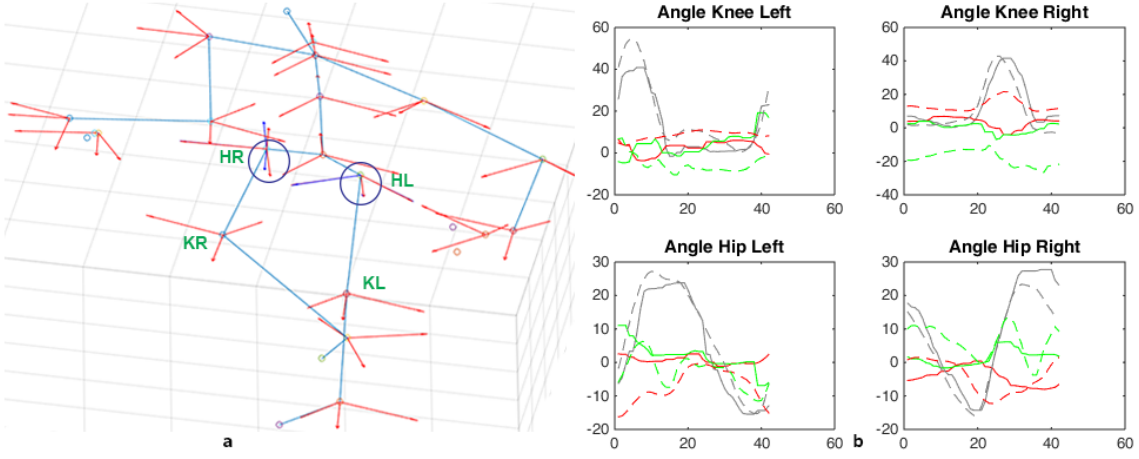


Figure 3.9: a) Re-orientation of the hip. Red - initial orientations, navy - modified to match Vicon orientations, blue - skeleton data. b) Angles calculated for 1 gait cycle by the Kinect v.2 and Vicon (dash). Gray - flexion, green - rotation, red - abduction.

obtain a quaternion q_{ka} (w_{ka}, x_{ka}, y_{ka},z_{ka}) corresponding to the knee flexion angle:

$$\begin{aligned} w_{ka} &= (w_k \times w_a - x_k \times x_a - y_k \times y_a - z_k \times z_a), \\ x_{ka} &= (w_k \times x_a + x_k \times w_a + y_k \times z_a - z_k \times y_a), \\ y_{ka} &= (w_k \times y_a - x_k \times z_a + y_k \times w_a + z_k \times x_a), \\ z_{ka} &= (w_k \times z_a + x_k \times y_a - y_k \times x_a + z_k \times w_a). \end{aligned} \tag{3.2}$$

Note that q_a^{-1} our case is equal to the conjugate of a quaternion. Then q_{ka} is converted to Euler angles in a predefined order, specified as 'ZXY'. The conversion between quaternions and Euler angles is:

$$\begin{bmatrix} \phi \\ \chi \\ \psi \end{bmatrix} = \begin{bmatrix} tan^{-1}\left(2\frac{w_{ka}z_{ka} + x_{ka}y_{ka}}{1 - 2(y_{ka}^2 + z_{ka}^2)}\right) \\ sin^{-1}(2(w_{ka}y_{ka} - x_{ka}z_{ka})) \\ tan^{-1}\left(2\frac{w_{ka}x_{ka} + y_{ka}z_{ka}}{1 - 2(x_{ka}^2 + y_{ka}^2)}\right) \end{bmatrix} \tag{3.3}$$

A particular case arises when $|q_w q_y - q_x q_z|$ is equal to $1/2$ [66]. In order to calculate the angle for the hip joint, we need to use the re-oriented hip joint and its child, the knee joint orientations. Let the q_h be the quaternion of a hip joint. Then q_{hk} is equal to:

$$q_{hk} = q_k^{-1} \times q_h \tag{3.4}$$

Equation 3.3 is used to obtain flexion angles in radians, which are then converted to degrees. The knee-ankle and hip-knee angle obtained with Kinect and Vicon for a gait cycle are visualized in Figure 3.9 b. Visual evaluation shows that flexion angles by Kinect and Vicon are similar in amplitude and dynamics, but the rotation and abduction vary.

3.2.3/ Comparison of Gait Kinematic Parameters for Vicon and Kinect

In the previous section, the procedure to calculate Kinematic parameters for Hip and Knee joints using Kinect's orientation data is given. This section quantitatively assesses the correlation between angles obtained with Kinect and Vicon systems based on our gait data.

Reliability is defined as the extent to which measurements can be replicated. In other words, it reflects not only the degree of correlation but also the agreement between measurements. Historically, the Pearson correlation coefficient, paired t test, and Bland-Altman plot [13] have been used to evaluate reliability. However, the paired t test and Bland-Altman plot are methods for analyzing agreement, and Pearson correlation coefficient is only a measure of correlation, and hence, they are non-ideal measures of reliability according to [206]. The intra-class correlation coefficient (ICC) is an index to measure a reliability metric which reflects both the degree of correlation and agreement between measurements.

McGraw [22] defined 10 forms of ICC based on the Model (1-way random effects, 2-way random effects, or 2-way fixed effects), the Type (single rater/measurement or the mean of k raters/measurements), and the Definition of relationship considered to be important (consistency or absolute agreement).

Following the recommendations of [206] and [22], for our intra-systems comparisons we selected the correlation indexes to be used. These are coefficient for consistency for single rater/measurement (ICC(C,1), also known as norm-referenced reliability) and correlation coefficient for absolute measurement (ICC(A,1), also known as crite -rion-referenced reliability) defined by the Table 4 in [22]. Even knowing the Vicon and Kinect frame-rate, we did not obtain a perfect cycles alignment with our segmentation algorithm. Therefore, we decided to use the ICC(C,1) correlation index between the measurements. We decided to provide the absolute agreement index below as well. However, the values might be slightly lower due to our imperfect resampling scheme. The index was calculated for 32 cycles taken with Kinect and Vicon simultaneously using the setup described in 3.2.1.

Tables 3.4 and 3.5 summarize the ICC values obtained for consistency and absolute agreement measurements. They signify worse overall agreement than the ones reported by [232], but follow the same trend: flexion/extension and internal/external rotation angles are correlated with the Vicon measurements while the abduction/abbuction angles are not. The smaller values of the ICC in our experiment can be due to the fact that the Kinect provides less accurate measurements

Table 3.4: ICC(C,1) and its 95% confidence interval correlation index for Vicon and Kinect angles.

Joint	Flexion/Extension ICC(C,1)	Abduction/Adduction ICC(C,1)	Rotation Internal/External ICC(C,1)
Left Knee	0.69(0.67-0.71)	-0.12(-0.15-0.09)	0.35(0.31-0.38)
Right Knee	0.73(0.71-0.75)	-0.37(-0.04-0.33)	0.05(0.01-0.08)
Left Hip	0.85(0.84-0.86)	-0.18(-0.22-0.15)	0.45(0.42-0.48)
Right Hip	0.76(0.74-0.78)	0.14(0.10-0.17)	0.32(0.28-0.36)

Table 3.5: ICC(A,1) correlation (index) for Vicon and Kinect angles.

Joint	Flexion/Extension ICC(A,1)	Abduction/Adduction ICC(A,1)	Rotation Internal/External ICC(A,1)
Left Knee	0.68(0.63-0.72)	-0.04 (-0.10-0.04)	0.14(-0.08-0.36)
Right Knee	0.73 (0.71-0.75)	-0.12 (-0.24-0.12)	-0.03(-0.06-0.00)
Left Hip	0.75 (0.26-0.89)	-0.026 (-0.07-0.05)	0.42(0.31-0.50)
Right Hip	0.72 (0.58-0.80)	0.11(0.03-0.17)	0.18(0.06-0.37)

within specific distance ranges [231]. Guess et al. [232] used an optimal capture space for their experiment. However, we need more space to perform a gait acquisition. Nonetheless, we obtain moderate reliability for knee flexion and good reliability for hip flexion. Rotation angles signify poor reliability, but can still be interesting as an additional characteristic for gait analysis. Abduction angles were judged as not correlated with the exception of the right hip data.

3.3/ Acquisition System Design

We experimented with a single Kinect v.2 sensor setup and built an experimental multi-Kinect platform. To increase the accuracy of our acquisition system, it should be calibrated. This section is dedicated to single sensor and stereo setup calibration procedures. We experiment with different approaches and report on the findings. We also talk about the difficulties arisen in each scenario.

In all the following experiments, we used the Kinect for Windows Software Development Kit [4] to obtain the 3D time series of 25 body points, color and depth images, and also the user masks by means of the in-built and externally validated human-pose estimation algorithms. The Kinect data was saved using custom-written software utilizing the SDK 2.0 and OpenCV library.

3.3.1/ Single Kinect

In most of the studies dedicated to human movement assessment using an RGB-D cameras, a single device is used. This is the simplest scenario, and commonly authors just use the out-of-the-box camera with build-in calibration parameters. In this case researchers may face slight mismatching between color and depth data, which is not critical for most applications. To increase the quality of the matching of different modality data, sensor calibration should be performed. This

subsection is dedicated to a calibration procedure required to obtain the most accurate data from a single Kinect camera. Then we provide the recommendations for camera installation and setup for optimal gait sequence acquisition.

3.3.1.1/ Camera Calibration

Geometric camera calibration estimates the parameters of a lens and image sensor. Lens parameters are called distortion parameters, and image sensor parameters are the camera intrinsics. Normally, the calibration parameters of RGBD cameras are known beforehand, as they are measured during fabrication. However, they can be slightly inaccurate. We noticed, that typically there is a small shift between the color and depth image, linearly dependant on the distance from a sensor. Moreover, this inaccuracy increases in case of a multi-camera setup, possibly introducing significant errors. Therefore, we investigated the sensor calibration methods available. Before discussing calibration methods, we will introduce relevant camera and lens parameters.

Distortion

The kinect camera has a lens which allows for a large field of view, but introduces distortion to the image. The two major distortions are radial distortion and tangential distortion, but for Kinect device the radial distortion is the most important one. A scheme of the radial distortion is shown in Figure 3.10.

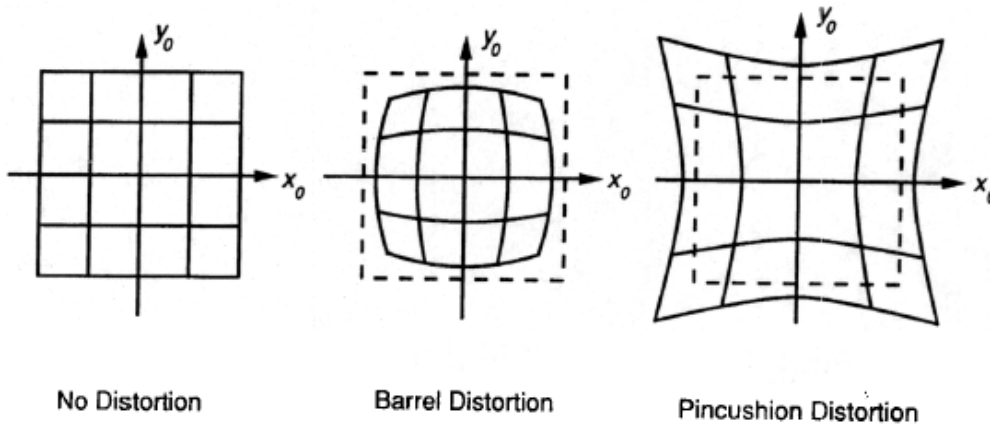


Figure 3.10: Radial Distortion examples. Taken from [9]

An equation reproducing the radial distortion for each pixel (x,y) is represented as follows [1]:

$$\begin{aligned} x_{distorted} &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \\ y_{distorted} &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \end{aligned} \tag{3.5}$$

where r is the radius from principal point to the predicted point and k are the distortion coefficients.

Another distortion is the tangential distortion which occurs because image taking camera lens is not aligned perfectly parallel to the imaging plane. It is represented as:

$$\begin{aligned} x_{distorted} &= x + [2p_1 xy + p_2(r^2 + 2x^2)] \\ y_{distorted} &= y + [p_1(r^2 + 2y^2) + 2p_2 xy] \end{aligned} \tag{3.6}$$

Tangential distortion is rarely considered in the calibration processes.

We need to find five parameters, known as the distortion coefficients=(k_1,k_2,p_1,p_2,k_3). Although the optical distortion induced by the Kinect is pretty minimal, it is a non-trivial task to find the optimal

parameters via camera calibration, because on practice different distortion models (for example, higher-order coefficients) should be tested.

Intrinsic

The next group of parameters that should be known about a camera are its intrinsic parameters. Intrinsic parameters are specific to a camera. It includes such information as focal length (f_x,f_y) and optical centers (c_x,c_y) et cetera [1]. It known as the camera projection matrix. Intrinsics depend on the camera only, so once calculated, it can be stored for future purposes. It is expressed as a 3x3 matrix:

$$K = \begin{bmatrix} F_x & s & c_x \\ 0 & F_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.7}$$

In our case s is the skew factor and it can be neglected, so s=0. For the Kinect camera, the above stated parameters should be estimated for each sensor (RGB, IR) separately.

Extrinsic The next group of parameters is called the extrinsic parameters. Extrinsic parameters denote the coordinate system transformations from 3D world coordinates to 3D camera coordinates. In the scope of this work, we are interested in the stereo setup. In this case we seek to define the position of the camera center and the camera's heading in world coordinates towards the second camera. When intrinsic parameters for all individual cameras are known, RT parameters between two sensors should be estimated. Extrinsic parameters correspond to the rotation and translation vectors which translates the coordinates of a 3D point to another coordinate system.

With these parameters we can map the depth image to the color image and back, in order to find matches. Using the intrinsic matrices, a 3D point cloud can be obtained.

3.3.1.2/ METHODS TO PERFORM THE CAMERA CALIBRATION

Camera calibration consists in the estimation of a model for an uncalibrated camera. The objective is to find the external parameters: the position and orientation relatively to a world co-ordinate system, and the internal parameters of the camera: principal point or image centre, focal length and distortion coefficients.

For the Kinect SDK users, Microsoft provides the calibration parameters for each device, which contain all the distortion parameters plus the intrinsics of the sensors. These are enough to get a 3D point cloud and find coarse correspondences with the matching color image. However, the matching between depth and the RGB image is not not perfect, so in the case when color information is used along with depth, camera calibration is an essential step. Camera calibration is an ubiquitous task in computer vision implemented by many different algorithms. The most used camera calibration techniques are the ones proposed by Tsai [15] and Zhang [29]. Its algorithm requires corresponding 3D world points and 2-D image points. The correspondence can be found if we know the geometric parameters of the 2-D image points. A so called calibration pattern with a objects whose dimensionality is known, is commonly used for calibration. It can, for example, be a chessboard pattern or a pattern with circular dots.

In this work we use a chessboard pattern, in which the corners provide 2-D features that can be computed as the intersection of the lines as shown on Fig 3.11.

It uses a two-stage technique. First, the position and orientation are ascertained, followed by the internal parameters of the camera. In this work, we are using two ready solutions, namely the Camera Calibration Toolbox from Matlab [46] and the OpenCV calibration utilities [77].

Apart from classical calibration methods, there are works addressing the calibration of a Kinect sensor in detail [103].

a b

Figure 3.11: a) An example of target acquired by a Kinect camera and used for the calibration. The image used is the color image mapped to the depth one, hence the missing values. b) The A2 10×7 checkerboard pattern used for our test and the 70 corners detected by OpenCV software [1].

3.3.1.3/ Results for the single Camera Calibration

In practise, Kinect camera calibration is a challenging task even when using the available software and a set of different calibration targets (A4, A3, A2). With 70 images containing the target at various poses and distances relative to the camera, the average reprojection error was about 1 to 3 pixels with the OpenCV library, and with up to 1 pixel, when using the Matlab Calibration toolbox. When creating the point clouds using the resulting calibration matrices, we didn't obtain a better visual quality than by using the factory-set calibration parameters of our cameras.

Two Kinects v.2 are possessed by out laboratory. For the following work, we preferred to use the following calibration parameters for the two kinects. Intrinsics:

$$K_1 = \begin{bmatrix} 366.59 & 0 & 256.67 \\ 0 & 366.59 & 205.25 \\ 0 & 0 & 1 \end{bmatrix}. \quad K_2 = \begin{bmatrix} 366.69 & 0 & 257.51 \\ 0 & 366.69 & 204.79 \\ 0 & 0 & 1 \end{bmatrix}.$$

Distortion coefficients:

D_1 = [-0.27, 0.09, 0, 0, 0.09]

D_2 = [-0.27, 0.09, 0, 0, 0.09]

Kinect does not provide users with the tangential distortion parameters, so we consider the tangential distortion as non-existing.

The parameters mentioned above are later used for the multi-camera calibration.

3.3.2/ Multiple Kinects

Multiple Kinect devices can be used simultaneously to capture complete point cloud. Such complete point clouds contain more information about the scene arrangement and can be used in further 3D-based applications.

3.3.2.1/ CAMERA CALIBRATION

Two cameras observing the same scene from different positions are called a stereo camera system. Although a Kinect is a multi-sensors device and thus already a stereo camera by itself, here the stereo system is a pair of individual Kinect v.2 devices. Two devices allow to capture a bigger scene volume from different view points, which assures more complete 3D information.

A possible stereo-system setup is shown in Fig 3.12. An essential step of any multi-camera setup construction is an accurate stereo calibration, which is a non-trivial and sometime unattainable step. A common practice is to improve the calibration results further by a so called Iterative Closest Point algorithm [20].

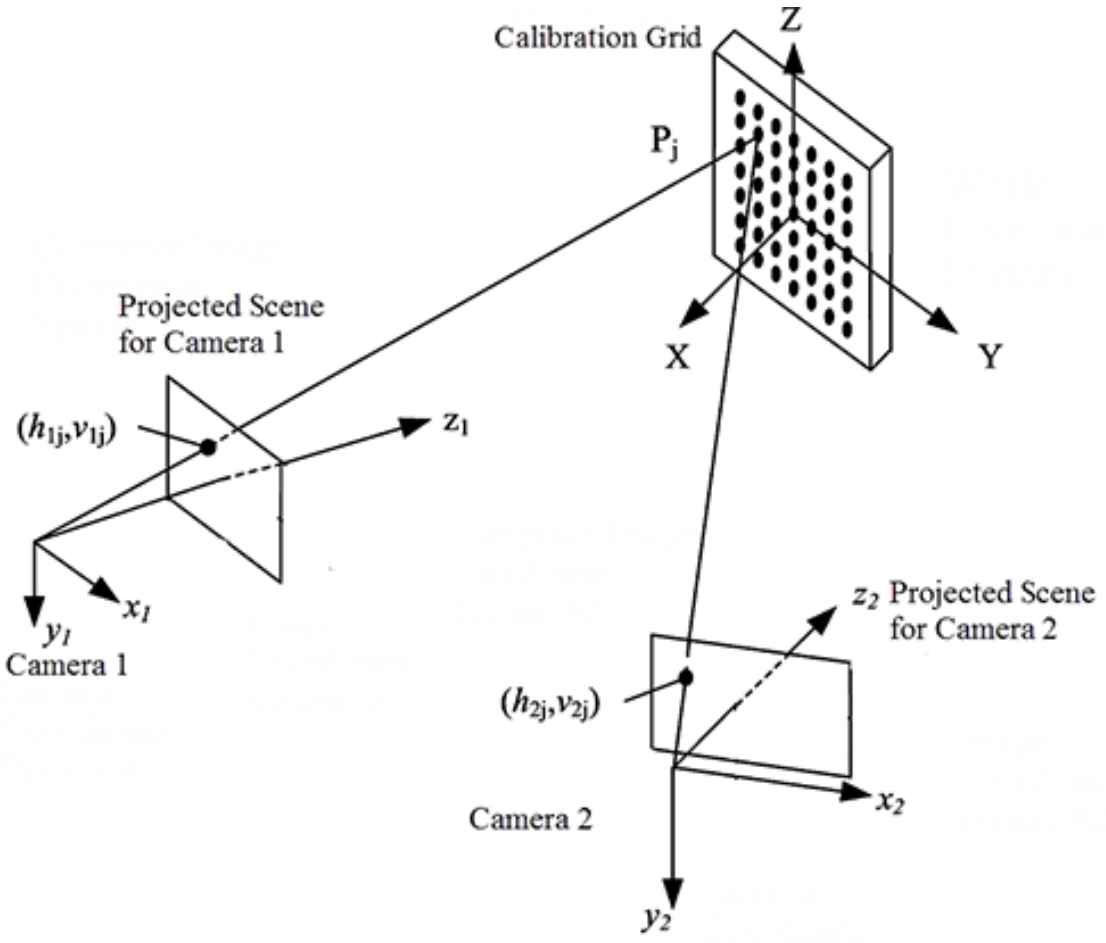


Figure 3.12: Stereo Setup. Two pinhole cameras are facing the calibration pattern with dots. The 3D point $P_j(X, Y, Z)$ is projected to 2D camera planes points (h, v). Source of the picture [11]

Before coming to this step, each Kinect camera was calibrated separately as described in the previous section.

In order to restore the scene information using a stereo setup, the point correspondences in two images should be established. For pinhole cameras without distortion it can be described by a matrix called the fundamental matrix [35]. When the camera lens has significant distortion, we need to take the distortion into account or rectify the images beforehand using the parameters estimated in the previous step. All the information about the classical stereo system, its parameters and calibration can be found in [35].

However, our goal is to work with the point clouds obtained by two Kinect v.2 cameras, therefore we are only interested in finding the rotation and translation between two cameras. Solving for R, t in equation:

$$B = R \times A + t, \tag{3.8}$$

where R, t are the transforms applied to dataset A to align it with dataset B. R is the rotation matrix which rotates an object, and T is the translation to bring the object into the world reference frame. R and T have the following dimensions:

$$R = \begin{bmatrix} R_{11} & R_{12} & R13 \\ R21 & R22 & R23 \\ R31 & R32 & R33 \end{bmatrix},$$

$T = [(T_x\ T_y\ T_z)]$.

To calibrate the stereo camera system means to find R and T matrices, which allow to align the data from one camera to the coordinate system of the second camera.

We are interested in the calibration between the two infrared sensors of the Kinect stereo system. There are two possibilities to proceed. Either use the data from the infrared camera directly to calibrate the system, or use the color images and apply the known Color-to-Depth transformation to obtain the depth image coordinates later. The first method is more desirable because we can increase point cloud registration accuracy by eliminating the error caused by the RGB to depth space mapping.

The Kinect v.2 depth images are computed from the captured infrared images and therefore, both images have the same optical specifications. However, there are many difficulties in using the infrared images itself, which need a gamma correction applied on images originating from distances greater than 1 meter. After initial tests, we conclude that on a distance greater than 2 meters, the IR image of the kinect sensor becomes very low contrast, so even gamma correction doesn't allow for a calibration target's corners detection. Therefore, the solution is to use the color images from Kinect.

For the construction of a scene structure, i.e. a point cloud transformation, we use the depth image. Unfortunately, to calibrate the camera based on the depth image, we need a special calibration target with different depth of squares. Other methods are also proposed in literature such as [54]. But usually they are not straight forward to implement. We choose the other way and calibrate the system based on the color image, which is mapped to the depth image beforehand.

That is to say, we obtain the color image from Kinect and map all the pixels to the corresponding depth image using the known calibration parameters, i.e. a homography to do the color to depth image mapping. The resulting original and rectified image is shown on Fig 3.13. By collecting numerous image pairs with the stereo system we can obtain the calibration parameters of two cameras and thus the R and T matrices.

A custom OpenCV based application was written. It estimates the R and T between two sets of 3D points, which are obtained from the 2D image target corners. The program works as follows: We detect the corners of the calibration target on two images separately using an OpenCV function [55], project them to 3D using cameras intrinsics and find the correspondence between the 3D points to obtain the R and T.

The final mapping RT can then be found as a linear transformation between two sets of points. We use the stereo Calibration function from the OpenCV library. The function estimates transformation between two cameras making a stereo pair. The function calculates the position and orientation of the second camera relative to the first camera. It computes R and T such that:

$$R_2 = R \times R_1 T_2 = R \times T_1 + T. \tag{3.9}$$

Here indexes corresponds to the first and second camera. The function minimizes the total re-projection error for all the points in all the available views from both cameras. Using the method

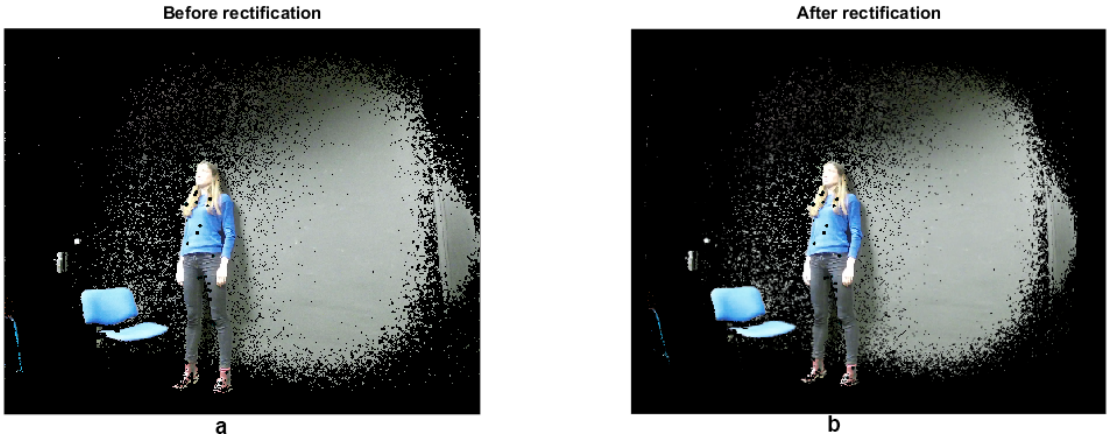


Figure 3.13: Mapped depth to color image. a) Initial image. b) The color image is slightly changed after the automatic rectification from the Matlab toolbox. The visual difference is very small and can be mainly seen in on the wall region.

described, we proceed the stereo Kinect calibration. The calibration was performed only between a stereo pair to calculate the R and T. The intrinsic parameters of the cameras were fixed. Then both point clouds are transformed to a common reference space and fused together. We considered one Kinect's coordinate system as a reference, and then the other Kinect's relative pose was estimated using translation T and rotation R transformations. Resulting aligned point clouds are shown in Fig 3.14.

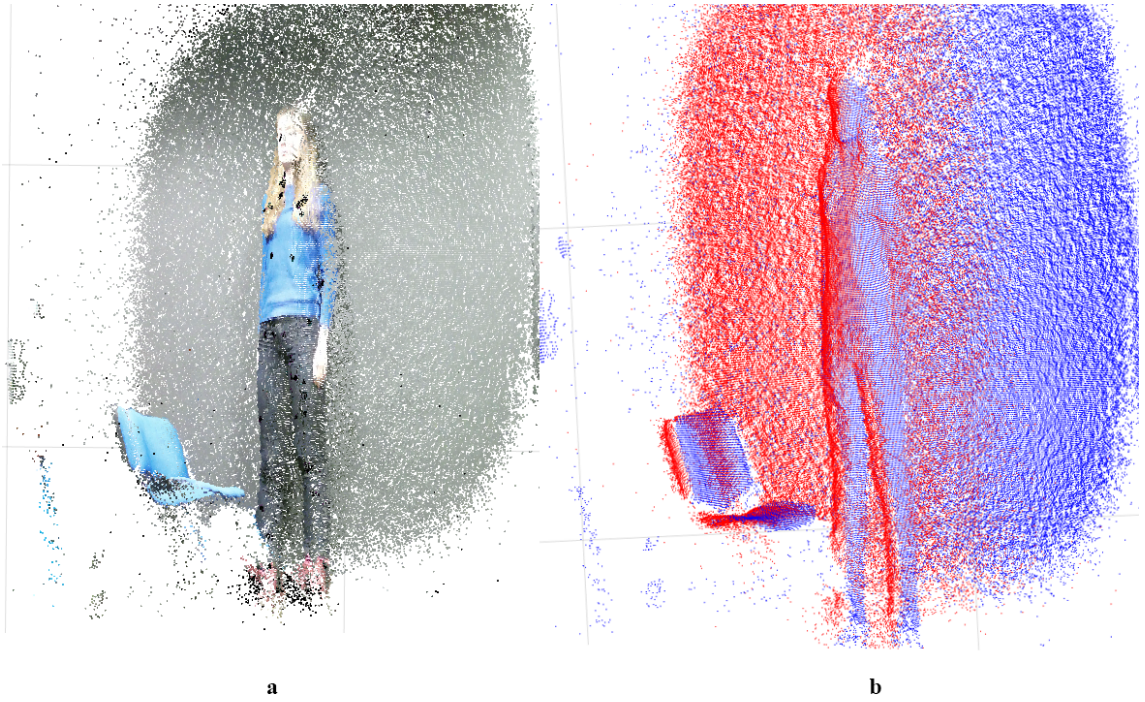


Figure 3.14: Resulting point cloud from two Kinect cameras. a) Point cloud with spatial and color information XYZ + rbg b) Point cloud with spatial information only XYZ, colored for the visualization purposes: blue - camera 1, red - camera 2. The cameras are installed on the left and right side of the person. The visual calibration result is satisfactory.

Another example when the person is placed between two cameras is shown in Fig 3.15.

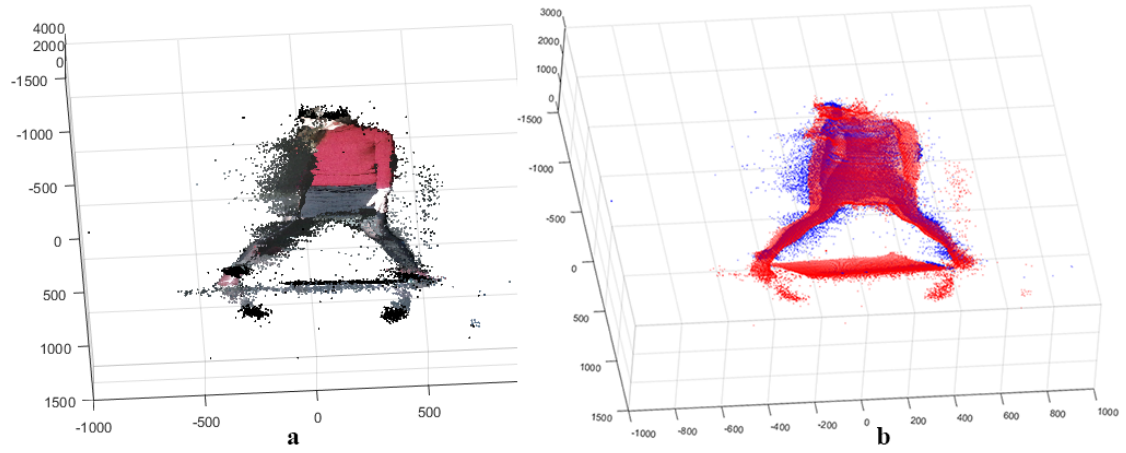


Figure 3.15: Resulting point cloud from two Kinect cameras, the background is removed. a) Point cloud with spatial and color information XYZ + rbg b) Point cloud with spatial information XYZ, colored for the visualization purposes: blue - camera 1, red - camera 2. The person is placed in between two Kinect sensors facing each other.

We experimented with different camera placements and did not handle the interference problem, since the data quality was satisfying even in case when two kinects were facing each other. The alignment results were evaluated visually and by an automatically calculated error during the calibration process, which was about 1 pixel. Visually the resulting point cloud borders are smooth, without visible misalignment.

3.3.2.2/ SKELETON ALIGNMENT

The calibration method described in the previous section is very efficient, but requires a big calibration target and additional procedures and computations. In this section we test an alternative approach, a calibration performed by a direct comparison of skeleton joints. This idea was inspired by other geometric calibration methods, where rigid transformations that help to align the 3D point clouds from each RGBD camera are computed. Computed transformation can be used to transfer the two Kinects' skeletons and, initially, point clouds to a joint coordinate system in the registration and reconstruction stage.

Since both Kinects in our stereo system provide us with series of skeleton joint coordinates, we suppose we can use their correspondences to calibrate the system. For example, knowing the most stable joints positions, we can find the RT matrix data using Singular Value Decomposition (SVD) using the method proposed by [62].

Finding the optimal rotation and translation between two sets of corresponding 3D point data, so that they are aligned, is a common problem in computer vision. Since we know that some skeleton joints are more reliable than others, we can get estimate the Rotation and Translation matrix from stable joints correspondence.

Given S1 = $j_1, j_2, \ldots, j_n$ and S2 = $q_1, q_2, \ldots, q_n$, where j and q correspond to skeleton joints 3D coordinates, we want to find the best, i.e the one which gives the least square error, rotation and translation that will align the points in skeleton S1 to skeleton S2:

$$(R,t) = armgin \sum_{i=1}^{n} w_i ||(Rj_i + T) - q_i||^2, \tag{3.10}$$

where $w_i > 0$ are weights for each point pair. This transformation is called the rigid transform,

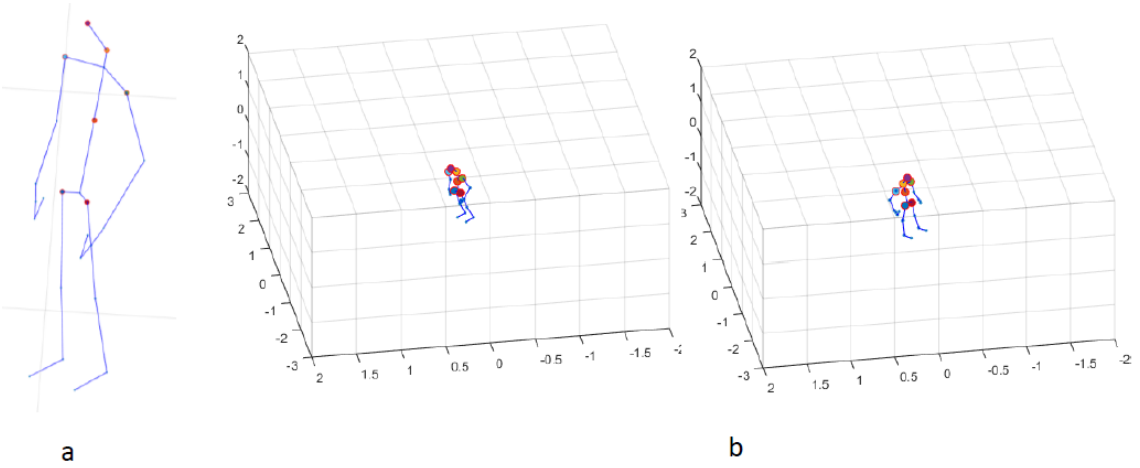


Figure 3.16: a) Selected joints. b) Initial Skeletons S1 and S2 (blue) from 2 Kinects with selected joints coloured.

because it preserves the shape and size of our 3D skeletons.

First both datasets are re-centred so that both central spine joints are at the origin of our coordinate system. We do it by computing the relative difference of each joint' triplets $[jX_i,\ jY_i,\ jZ_i]$ with the position of the root joint $[jX_{root},\ jY_{root},\ jZ_{root}]$.

Then we selected the following seven joints, which are claimed to be the most reliable for the Kinect v.2 camera skeleton estimation algorithm: Head, Neck, Spine Mid, Spine base, Left Shoulder, Right Shoulder, Left Hip, Right Hip. The corresponding joints are shown in Fig 3.16. The joints were selected based on previous studies. In particular, we took the most stable joints according to Otte et al. [213], who performed a kinematic validation of movement signals from Kinect landmarks against Vicon marker locations by means of the 3D Euclidian distance diff3D, Pearson correlation coefficients per dimension and signal-to-noise ratio SNR per dimension. Similarly, Geers et al. [175] obtained good results for these joints with a multi-Kinect walking assessment system. We also tested a reduced joint set with hip joints excluded.

Normalization of the joints towards the central hip joint removes the translation component, leaving only the rotation. The next step involves accumulating a matrix, called H, and using SVD to find the rotation as follows:

$$H = \sum_{i=1}^{n} S_1 (S_2)^T. \tag{3.11}$$

SVD will factorize a matrix with point correspondences (H), into 3 other matrices, such that:

$$[U, S, V] = SVD(H), \tag{3.12}$$

$$H = USV^T, \tag{3.13}$$

$$R = VU^T. \tag{3.14}$$

Our experiments show that estimation of the R matrix using the skeletons joints only, does not give very accurate results in terms of alignment. This is probably due to the fact the body is not a rigid object. In addition, skeleton joints positions estimated by two the Kinects are rather different, so to use this method this problem should be addressed first. In the case of the stereo camera setup, it is probably more interesting to experiment with new ways of 3D skeleton fitting using the point cloud data. An example of the resulting alignment is shown in Fig 3.17. It can be seen that the resulting alignment is poor, although the method can be used for a coarse alignment.

Figure 3.17: The examples of resulting aligned Skeletons. a) Front view. b) Back view. The skeletons ratios and estimated postures delivered by two Kinect cameras are slightly different, so perfect alignment is not possible.

The alignment of two point clouds and corresponding skeletons is shown in Fig 3.18. The obtained 3D scene is visually less accurate that the one obtained with the stereo calibration methods described in the previous section, with huge shifts between point cloud data coming from two sensors.

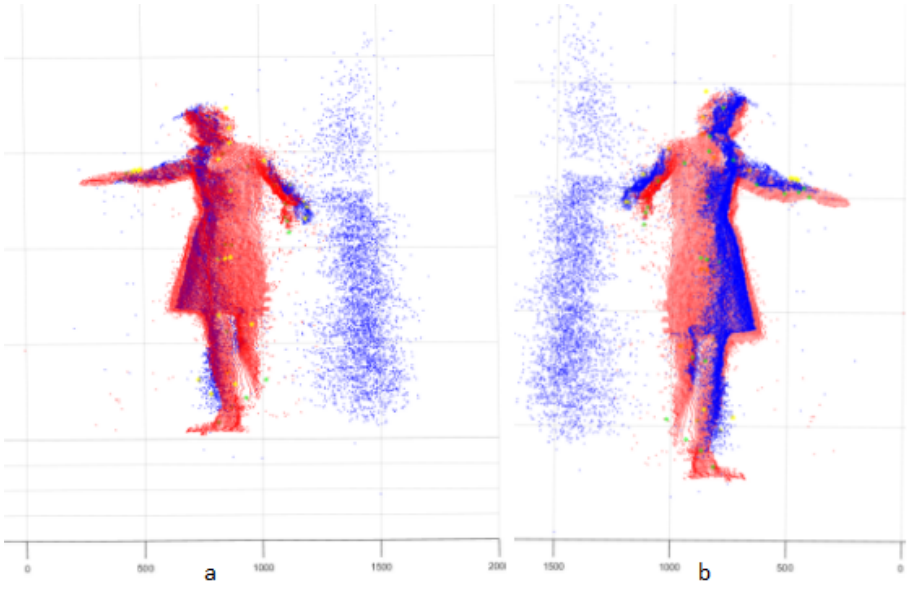


Figure 3.18: Aligned by R and T from the calibration skeletons joints, in green and yellow, and corresponding point clouds in red and blue. The person is in between two Kinect sensors. a) Front view. b) Back view.

3.3.2.3/ TIME ALIGNMENT

One of the common limitations in motion capture systems is the necessity to have synchronized cameras. Existing RGB-D sensors, such as the Microsoft Kinect v.2, do not have the possibility to trigger the image acquisition. Moreover, the frame-rate of computers can be very different and

depends on the speed of the computer hard drive and the processor.

For the two PC's available for this project, the frame rate differs with a range up to a factor two, when using the same custom software created for the acquisition. The two PC are saving the color and depth images with a resolution of 512×424, next to a binary mask of the detected human body and plain text file with the temporal data, skeleton joints coordinates and orientations. The average frame rate for the computers is 35.4505 fps and 28.440. Moreover, the peak frame rate can vary during a run.

In order to use data from the two Kinects simultaneously, the two systems have to be synchronized. We assume that the acquisition starts at the same time on both machines.

There exists many solutions to synchronize two acquisition machines, amongst which the LAN protocol is the most commonly used. However, this solution needs further hardware and software knowledge and preparation time.

Another easy software solution would be to use a timer, which will be run to write the data from two Kinects with a pre-defined diminished frame rate. Such a timer was successfully implemented to acquire the images of a calibration target automatically. However, due to the clock difference, the synchronization up to a ms rate is not possible, unless we slow down the frame rate to 15 fps, which is not desirable for human motion acquisition. Plus, for some reason, certain packages of data arriving from the camera may be lost, or the writing process can last longer, all which introduces miss alignments in the data which are hard and costly to recover from in post processing.

When finally writing data to a disk, we use the original frame rate for both machines and then align the data using the slowest machine as a reference.

What we want is to find the frame from camera B which corresponds the most for each frame of camera A. We can use the time data for that, which are stored in the acquisition time-stamp.

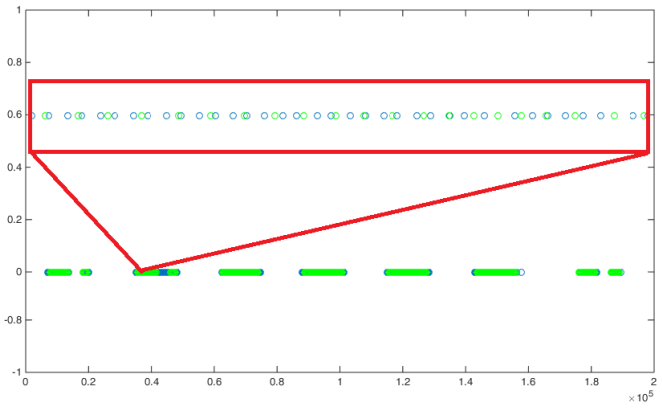


Figure 3.19: Time after alignment. Green dots are the frames from PC with higher frame rate, blue dots are frames the PC with lower frame rate. The red triangle zooms the resulting aligned frames. The person walks to and from the sensor, sometimes leaving the view range.

We store the time in ms for each frame acquisition for each Kinect, starting the counter from the moment acquisition has started. Then for each frame of the camera with lower frequency, we find the closest frame from another device, using the Euclidean time-wise distance. The distance between matching frames can then be used to evaluate the alignment error. The time series coming from both Kinect devices are shown in Figure 3.19. We do not obtain a perfect alignment with this method, but the final average error is 15 ms, so we consider it negligible for the following motion analysis.

3.3.3/ Conclusion on Single and Multiple Kinect Setups

In this chapter we examined different approaches to calibrate a single and stereo camera system of our two Kinects v.2. Single Kinect calibration did not outperform the build-in calibration data in the case of 3D scene restoration visual analysis.

The attempt to use the skeletons for the stereo setup calibration was unsuccessful. The two skeletons delivered by two Kinects facing the same person from different sights have shown to have different configuration. We can draw the conclusion that it is not possible to use skeleton data from two Kinect sensors to estimate the RT parameters for a stereo setup.

Stereo calibration was then performed using custom build software based on the OpenCV calibration feature and a standard calibration target. The resulting calibration is accurate enough to perform a 3D scene restoration and is sufficient for the following analysis of the data. Visually the resulting 3D scene point cloud is very accurate but contains some noise around objects' edges. The best calibration results were obtained using the factory intrinsic parameters of the Kinect sensors and executing stereo calibration using a standard calibration target. We also propose a simple post-processing time synchronization approach in order to roughly match the frames between two cameras.

3.4/ Conclusion

Amongst the commercial 3D cameras presented on the market in late 2016, the Kinect v.2 from Microsoft is the most suitable for human movement analysis. The main drawback of this sensor is the fact that it has a limited distance range. Using Microsoft software, a skeleton representation of a human can be extracted from the viewed scene. These skeletons are error-prone, but can be used for motion analysis when the person is directly facing the sensor. We demonstrated how kinematic parameters commonly used in gait research can be calculated from Kinect v.2 skeleton data. The following comparison with Vicon sensor has shown that Kinect v.2 has potential for a gait assessment system.

We experimented with different setups of motion acquisition using the Kinect v.2 camera. Several discoveries were made during this step of our work. When the Kinect is not placed directly in front of the patient, the reliability of the estimated joints and orientations positions decreases. The body part orientations calculated from Kinect images can be used to calculate kinematic parameters, comparable to those of golden standard MOCAP systems.

Skeletons estimated by two different Kinect devices do not match and can be used only for rough alignment. A stereo Kinect v.2 system can be calibrated using a standard calibration pattern and color camera data. However, the frame rate of Kinects using the official SDK is not constant and depends on hardware parameters and platform particularities. Therefore, for ideal time alignment, the Network Time Protocol (NTP) protocol for PC clock synchronization should be used to capture dynamic 3D scenes.

Overall, the usage of two or more Kinect sensors is only feasible in the case when the 3D scene information will used in the following analysis. The skeletons from different devices do not correspond to one another, and estimated body postures can be very different.

Unfortunately, there is no method for accurately estimating a 3D skeleton from point clouds, which could have justified the use of two cameras otherwise.

Due to the limited Kinect v.2 range, it is feasible to use a treadmill to capture the gait data. However, modern treadmill design usually has a control panel in front of the person. This requires the Kinect placement under an angle, which results in less accurate skeleton estimations. In the same moment, other researchers use such a setup, and it allows for longer acquisitions and more data. Probably, if the initial training data for posture learning algorithm used in Kinects SDK will be enriched with side views, the posture estimation results will improve.

The two cameras scenario is perfect in combination with the treadmill, when both cameras are installed statically from both sides of the patient. Through our experiments, we did not detect significant degradation of the depth data due to the interference of two sensors.

In the single camera scenario, we advise to remove the control panel of the treadmill if possible, and place the Kinect in front of the patient. Alternatively, the patients can just walk towards the camera on the solid static substrate, but the number of steps is then limited.

In the next two chapters of this thesis we propose a novel 3D posture descriptor for a complete point cloud analysis and a skeleton-based normal gait model for a single Kinect setup scenario.

4

3D Posture descriptor

The previous chapters introduce gait parameters, RGB-D sensors and specifics of their usage. 3D sensors provide with different types of data, among which a point cloud is the most interesting, since it reconstructs the original scene in 3D. Point clouds are used to represent volumetric data in many applications related but not limited to medical imaging, architecture, engineering, construction and others. As shown in the previous chapter, a single RGB-D sensor or a multi-sensor setup can be used to capture the 3D scene containing a human subject. The Kinect sensor is also capable to capture human motion with a frame rate about 30 fps. Human motion, including the gait, can be seen a sequence of related static postures. This chapter is dedicated to the static human posture description based on point cloud data. We are searching for the way to characterize the person's motion using 3D information directly, to avoid exploiting skeletonization algorithms. The latter are proven to be error prone, plus the skeleton estimation algorithms usually do not use complete 3D information (they treat the depth maps and not the point cloud data). Human gait can be considered as a sequence of repeating characteristic static postures. Detecting the postures, or gait key-events from the 3D data can be valuable for the following analysis.

This chapter presents a simple yet powerful algorithm for global human posture description based on 3D Point Cloud data in section 4.3. The algorithm is addressed mostly for the cases when the skeleton representation is not reliable (i.e, in case of the Kinect v.1 sensor, or when the patient is not facing the camera), or for the multi-camera setup when a complete point cloud (for example, the one obtained by a stereo Kinects system from Chapter 3) can be obtained.

The proposed algorithm preserves spatial contextual information about a 3D object in a video sequence and can be used as an intermediate step in human-motion related Computer Vision applications such as action recognition, gait analysis, human-computer interaction. The proposed descriptor captures a point cloud structure by means of a modified 3D regular grid and a corresponding cells space occupancy information. The performance of our method is evaluated on the task of posture recognition and automatic action segmentation in section 4.5.

Contents

4.1/ 3D Human Pose Estimation

3D pose estimation is a common task in Computer Vision applications. In the case of a rigid object, pose estimation seeks to capture the appearance of an object under certain viewing conditions. This task is challenging for natural images due to the ambiguity of an object representation in 2D, poor texture and varying view-points. With the introduction of consumer 3D sensors, this problem has been revisited by researchers developing a broad range of new descriptors. They may be both handcrafted [104] or automatic [192], and capture information from both global and local scales.

Non-rigid object pose estimation is inherently more complicated. A human body is an articulated object, and its motion can be build up from rigid and non-rigid motion parts. Articulated pose estimation seeks to estimate the configuration of a human body in a given image or video sequence. Recognition of body postures is an important step towards the fully automatic classification of human motion. Many researchers tried to use information from individual frames which are subsequently joined under some temporal structure modeling [112, 110]. Most 3D based gait analysis algorithms are build on the skeletons derived from static depth images and seek to analyze the posture evolution in time [211, 250, 169].

A canonical work on human posture estimation using RGB-D camera data is the one by Shotton et al. [130], which proposes a real-time algorithm which segments a human body from a corresponding depth map and locates skeleton joints in 3D. The algorithm belongs to ML category and uses hundreds of thousands of training images to learn the body configuration. All the training images are manually labeled into 31 body parts (could be changed based on a task). To achieve maximum accuracy, deep randomized decision forests are combined with pre-selected features which are robust to depth/scale and translation variations, invariance from the data to camera pose, body pose, and body size and shape. This algorithm shows good results and its variations are widely used today. The success of the method could be explained by the amount of training data, covering different human body configurations. However, it has certain limitations: in presence of severe occlusions and noise, the positions of the joints cannot be estimated correctly; it gives approximate joint positions and therefore coarse pose estimation and is not able to capture very subtle variations between postures. For this reason, joint-based posture estimation methods, although simple and powerful, will fail if the initial joints were estimated wrongly, which gives the way to low-level attributes based methods. The most difficult situation for modern algorithms estimating skeleton articulations is the side view. Figure 4.1 shows an example of ankle joint coordinates estimated by the Kinect v.2 of a person walking on a treadmill, the camera is placed under an angle (the picture from this setup is shown in Figure earlier 1.4, more joint positions are visualized in A.4). These data are hard to use for analysis, although it should be mentioned that feet and ankles are the most noisy joints among 25 provided by Microsoft Kinect SDK.

This chapter proposes a simple yet effective descriptor for pose recognition based directly on point cloud data. The algorithm takes a holistic pose estimation approach, capturing the slightest posture changes using accumulated point cloud features. Our descriptor is based on the space occupancy for cells of a modified 3D regular grid, super-imposed on a point cloud. It is translation, scale, and rotation invariant.

Originally, we aim at a descriptor which can be used for gait analysis. As mentioned earlier in the 2.2.1, we are interested in automatic gait classification and analysis. The proposed design should be able to detect reliably different postures in human gait, where the precision of skeleton data is not sufficient (the Kinect reliability is evaluated by [171] for the side and front [182] views). The second problem addressed is the symmetry of the gait which should be evaluated based on the point cloud data. In a normal gait, the left and right foot should move symmetrically. By detecting non-symmetrical patterns in a posture sequence, we can conclude on a possibly pathological gait. The number of various postures in gait is limited. Ideally, we want to have each cycle stage to be assigned a correct corresponding label. The main interest for the gait is the configuration of the low-limbs. For motion analysis, the whole body posture is interesting. In the same moment, the configuration of the body is also important for gait up to a certain extent (i.e, the correct posture of the patient, or the use of arms for the balance). There are two approaches among researchers

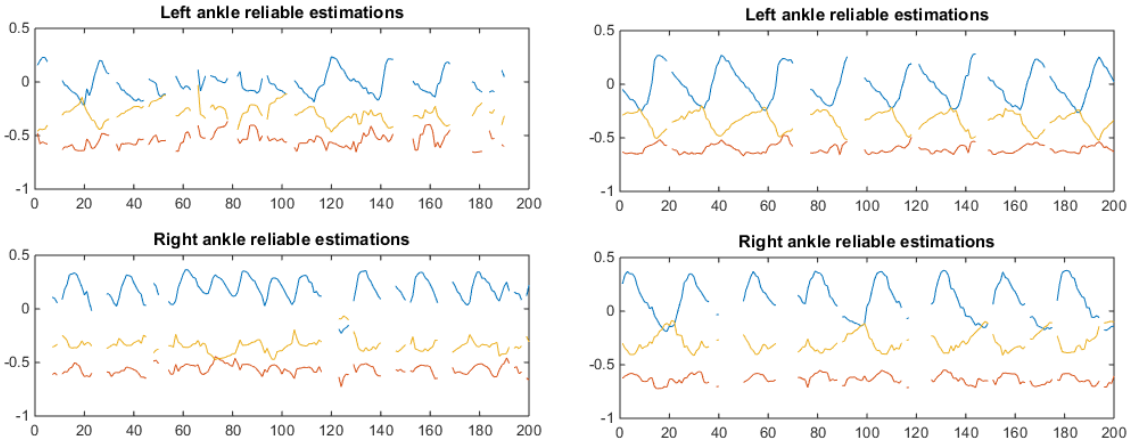


Figure 4.1: Visualization of the X(blue), Y(red), Z(yellow) ankle joint coordinates dynamics. Frames where status of joints was 'not tracked' are excluded. The ankle which is further from the sensor is not reliably estimated, there are many gaps and noisy data.

working on gait analysis: some take only information related to the low limbs [211, 168] and others prefer to use the whole skeleton/silhouette of a person [169, 250]. Not being sure which approach is better, we selected the following strategy. Before working on the gait key-postures, we selected a posture recognition dataset to evaluate the descriptor performance, since it contains more inherently different body configurations than a typical gait cycle, and therefore it should be easier to differentiate these postures. In addition, the resulting 3D descriptor is very general and can be used as an intermediate step in a great number of computer vision applications such as action recognition, gait analysis, smart homes, assessing the quality of sports actions, human-computer interaction and others, where posture estimation is an essential intermediate step. This work presents the descriptor in the context of action recognition, and the postures are estimated from frames of video sequences from MSR Action3D database.

4.2/ EXISTING POSTURE DESCRIPTORS

We start with a short review of the existing posture estimation methods. Most methods for human pose estimation are based on variations of a so called pictorial structures model, which represents human body configuration as a collection of connected rigid parts [140, 119, 43, 129]. To model an articulation, parts of the structure are parameterized by their spatial location and orientation. An example of such model is shown in Figure 4.2 a. Another example is the skeleton delivered by Kinect shown earlier in the pervious chapter.

Simple skeleton models are the most used by researchers. However, there is also a full 3D shape and body configuration model called SCAPE [41]. The model consists of pose deformation model that derives the non-rigid surface deformation as a function of the poses of the articulated skeleton and a separate model of variation based on body shape. Model parameters are learned from a human body scans dataset.

Holistic approaches [43][129][110] and middle-part [93] based methods form the other research direction in posture recognition. Holistic approaches aim to directly predict positions of body parts from image features without relying on an intermediate part-based representation. Part-based approaches first detect intermediate parts independently or with some constraints on body joints spatial relations.

Recently researchers significantly advanced posture recognition from natural images with the increasing popularity of machine learning based approaches [159, 170, 227, 252]. Cheron et al.

Figure 4.2: a) Visualization of the posture recognition approach by [129]. The poselets capture the anatomical configuration of the human in the input image, the representation is similar to the one used by Kinect. b) SKAPE model visualization. It describes the shape of the person and the posture. The posture is described by the articulations' position (shown by dots on the shape). The model produce 3D surface models with realistic shape deformations for different people in different poses.

[170] proposed a new Pose-based Convolutional Neural Network descriptor (P-CNN) for 2D action recognition. A pre-trained CNN learns the features corresponding to five pre-selected body parts based on quantized motion flow data for each frame. Chen and Ramanan [227] extend an estimated 2D model, using a neural network, to 3D using a simple Nearest Neighbor pose matching algorithm. A good review on recent advances in 3D articulated pose estimation is proposed by Sarafianos et al. [215]. Posture recognition is a part of action recognition, since actions can be modeled as a posture evaluation in time. Recent works on action recognition are based on CNNs [233, 236] and learn the features atomically, which leads to state-of-the-art results on available datasets. Guler et al. [252] train a deep neural network using a big manually annotated dataset, where each body is represented as a group of pre-defined segments.

Despite the significant progress made, full-body pose estimation from natural images remains a difficult and a largely unsolved problem due to numerous difficulties in real-life applications: the many degrees of freedom of the human body model, the variance in appearance, the changes in viewpoints, and lastly, an absence of data about an objects' shape. 3D data give a new important information which allows for improving posture recognition results. Depth-based pose estimation can be broadly categorized into two classes: generative and discriminative methods.

Generative approaches [163, 100] use a geometric or probabilistic human body model and estimate a pose by minimizing the distance between the human model and the input depth map. Human pose estimation is performed by optimizing the objective function for geometric model fitting by the means of variants of Iterative Closest Point [100] and graphical models [149] or pictorial structures [79]. A recent method by Wang et al. [219] uses several hand-crafted descriptors to recognize 5 distinct postures from the data obtained by the Kinect camera. Their algorithm is based on a simple 3D-2D projection method and the star skeleton technique. The final posture descriptor is composed of skeleton feature points together with a center of gravity. A pre-trained Learned Vector Quantization (LVQ) neural network is used for classification.

Discriminative approaches [130][194] perform classification on a pixel level and attempt to detect instances of body parts. Shotton et al. [130] trained a random forest classifier for body part segmentation from a single depth image and used Mean Shift [33] to estimate joint locations. Chang et al. [114] propose a fast random-forest-based human pose estimation method, where classifier is applied directly to pixels of the segmented human depth image. Jung et al. [194] used randomized regression trees and made their algorithm even faster by estimating the relative direction to each joint to avoid computationally demanding aggregating pixel-wise tree evaluations. The obtained skeleton data can later be used as the base for action recognition in videos as in recent Log-COV-Net method [226].

Most of the works on 3D pose estimation use a single depth camera. The most successful examples of single view pose estimation are [130][163][194][114] and most of them use randomized

trees and shape context features for pixel-wise classification which leads to real-time solutions.

Lately, multi-view depth image based posture recognition approaches acquired the attention of researchers [216][214]. The recent framework proposed by [216] uses several Kinect sensors and a deep CNN architecture. Multi-view scenarios allow to reconstruct 3D point clouds in the reference space. The authors use curriculum learning [58] to train the system on purely synthetic data. Curriculum learning modifies the order of the training procedure, gradually increasing the complexity of the instances, which hypothetically improves the convergence speed and the quality of the final local minima.

It is clear that the currently prevailing strategy is to use Machine Learning methods, specifically randomized trees [130][194][157], and a huge amount of training data. Modern posture recognition methods [130][216] have shown to be both effective and efficient in real-time posture estimation. Similar, for the following action recognition from videos, hand-crafted methods were overshadowed by deep learning based methods [236]. However, these methods are data-hungry, and there are not many 3D shape and posture datasets captured with an RGB-D sensor available.

This chapter introduces a new descriptor that estimates 3D human pose from a single point cloud. We are not attempting to outperform machine-learning based algorithms [130][194], but mostly propose an alternative, which does not require a priori human body model. In contrast to [219], we do not use a descriptor for a given posture but aim to use a general 3D point cloud structure. Unlike other popular descriptors [130][194] which use depth image features, our descriptor is based on a 3D structure and therefore can be used in a multi-camera scenario. We are aiming to use such a descriptor for gait key-postures recognition from a Point Cloud gait signature. Such signature is a collection of 3D frames, an example is shown in Figure 4.3. In order to be used in health assessment applications, we do not use an explicit human body model or other restrictions imposed in our algorithm.

4.3/ HUMAN POSTURE DESCRIPTOR DESIGN

OpenNI Framework [256] is used by many 3D cameras and provides the user with automatic body recognition and skeleton joints extraction functionality. The Kinect sensor also provides this functionality. Therefore, we are not addressing the task of background subtraction in our work and assume that it is a prior step. For this task, the data from an RGB-D camera, where the human is located and the background is subtracted, were used to test the proposed descriptor.

In this chapter we propose a handcrafted compact and discriminative descriptor for a single point cloud. The descriptor is designed to capture a point cloud configuration using a pre-selected greed structure. The most similar descriptor to ours is the Space-Time Occupancy Patterns method proposed by Vieira et al. [110] for the task of action recognition. Similar to this work, we propose to divide the 3D space by a regular grid and base our descriptor on spatial occupancy information. However, in [110] researchers compute the final descriptor vector by re-assigning weights based on cells where motion occurred. We are concentrated on a description of each static frame in order to recognize the posture in it. Other differences include the method of 3D space partitioning and descriptor cell initialization. Our partitioning is inspired by the 3D partitioning for human recognition from 3D point clouds proposed in [200]. Vieira et al. specifically design their method for video sequences, taking the time dimension into account. We assume that every initial frame posture is more important and temporal information can be encoded later in the process depending on the specific application. For the gait analysis and action recognition, a Hidden Markov Model can be coupled with a descriptor to capture the temporal information.

To construct our descriptor for each depth map video frame, we perform the following steps. First, the 2D-3D transformation is done to obtain a point cloud in 3D space from a depth map. We use a standard equation for basic geometric transformations:

$$X = Z * \frac{(j - c_x)}{f_x}; \qquad Y = Z * \frac{(i - c_y)}{f_y}; \qquad Z = z \tag{4.1}$$

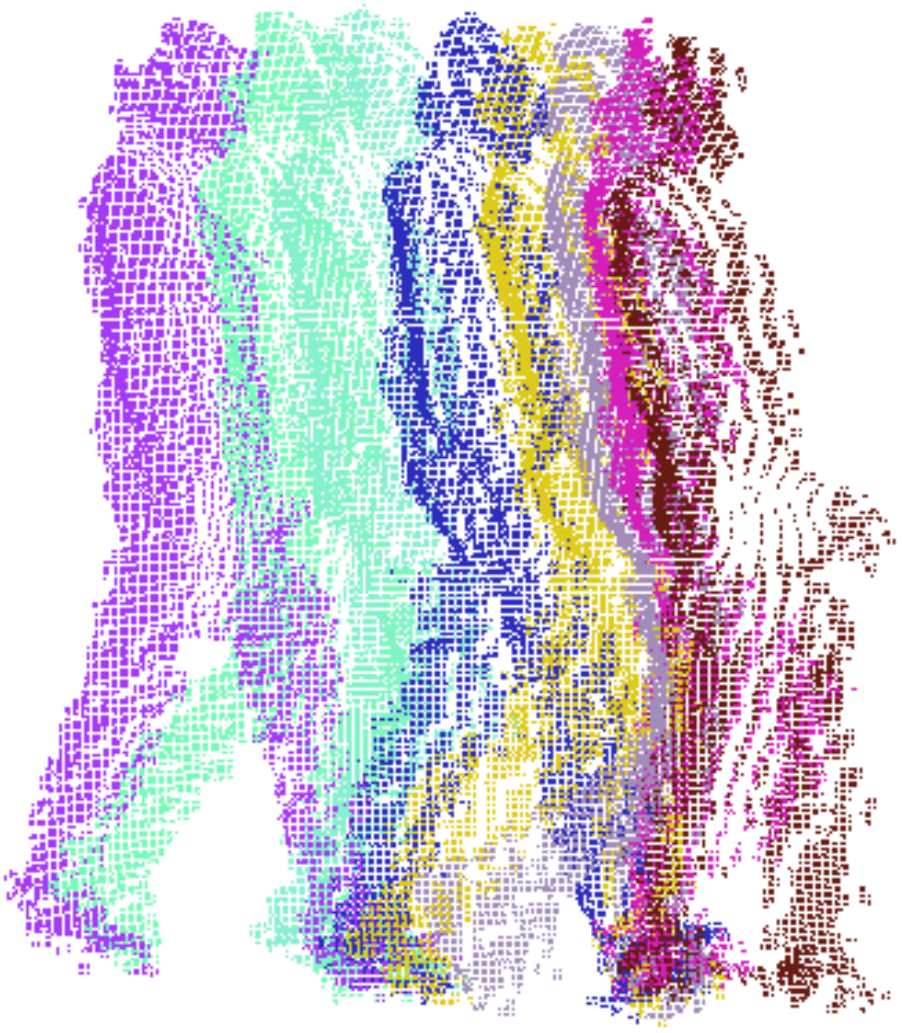


Figure 4.3: Point Clouds Gait Signature captured by a single Kinect v.1. Each frame is colored differently, each third frame of a gait cycle is used for visualization to make them easily distinguishable.

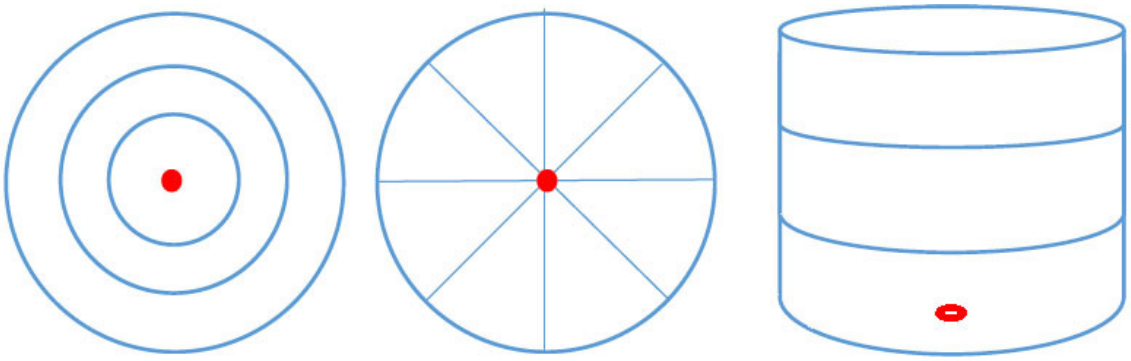


Figure 4.4: Descriptor spatial partitioning: 3 circles, 8 sectors, 3 sections. Projected center of gravity is shown in red.

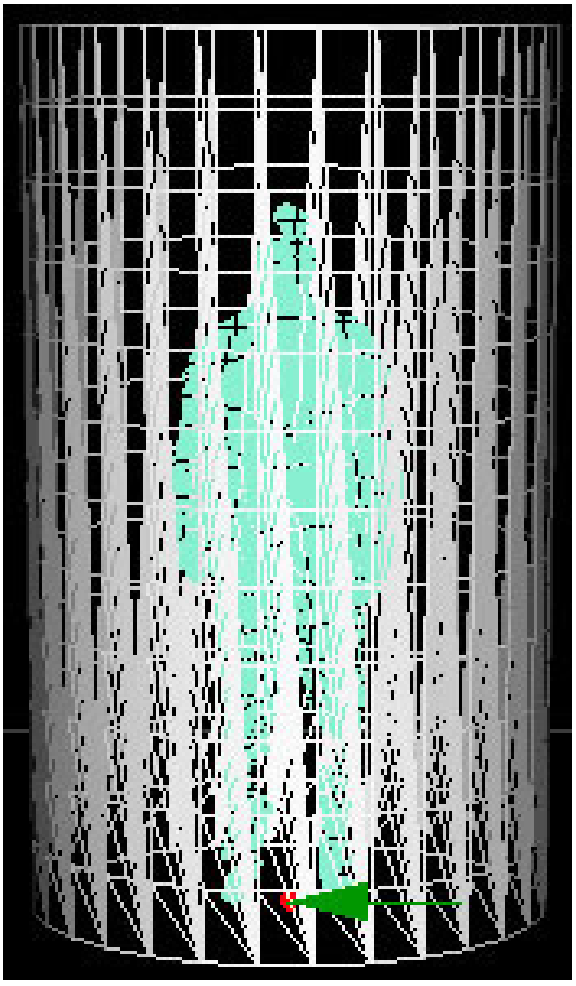


Figure 4.5: 3D spatial partitioning in 12 sections. Projected center of gravity is shown in red, fixed point view direction is shown by a green arrow.

where X, Y, Z are the point coordinates in 3D, j and i are the pixel coordinates, and c_x, c_y, f_x and f_y are the intrinsic matrix parameters obtained by a calibration of the Kinect camera. The intrinsic matrix and its parameters are explained in details in the previous chapter. Then the 3D spatial partitioning is performed. The center of gravity in 3D is calculated and projected to the ground plane:

$$C(X,Y,Z) = \frac{\sum_1^n (X,Y,Z)}{n} \tag{4.2}$$

where n is the total number of points in the point cloud. The ground plane can be calculated by 3 points, or, as it was done in this work, we assume that the camera direction is perpendicular to the floor, and then we can use XZ axis to estimate the plane orientation and the lowest point of the point cloud to locate it in 3D space. A 3D cylinder of varying dimensions with a base center in the computed centroid projection defines the space partitioning limits. The height and radius of the cylinder vary to adjust for the height of a person. The data about human body proportions ratio is used. A height of a person is estimated, simply via the minimum and maximum value calculated for the first static point cloud of a video sequence, corresponding to one action performed by one person. To have an equal grid for all frames of a video sequence, the normal is fixed based on the viewing point. This allow to start the partitioning in sectors from the same position for each video frame.

The visualization of the descriptor parameters is shown in Figure 4.4. For this work, we use only

a uniform space subdividing scheme and the cylinder volume partitioning is then performed as:

$$V = 2\pi RH \tag{4.3}$$

$$r_n = \frac{R}{n_c}; \qquad h_n = \frac{H}{n_h}; \qquad s_{angle} = \frac{360}{n_s} \tag{4.4}$$

where R and H are the fixed radius and height, r_n, h_n and s_{angle} are the circle and height intervals and the angle corresponding to each sector.

Figure 4.5 shows an example of 3D partitioning for one of the frames from MSR3D dataset. The only parameters of the descriptor are the number of sections, the number of sectors and the number of circles.

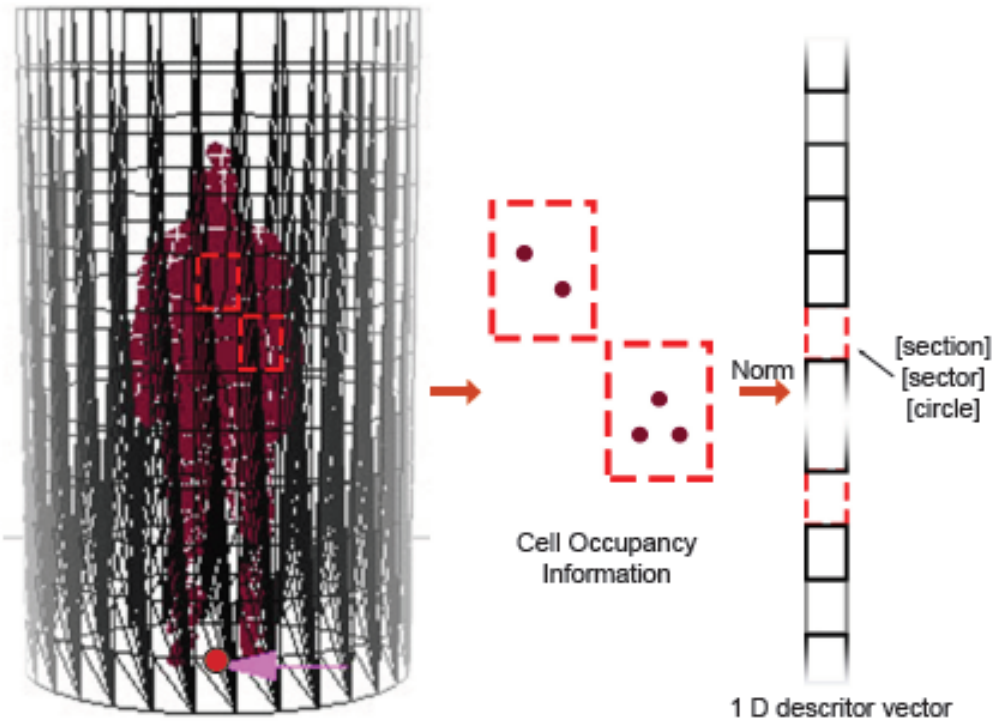


Figure 4.6: The point cloud structure is captured by the means of a modified 3D regular grid and the corresponding cells space occupancy information is then unwrapped into a 1D vector.

The final descriptor is obtained by calculating the number of points in each formed 3D cell i.e. the cell occupancy. The process is illustrated on Figure 4.6. Cell occupancy information is concatenated to a 1D vector. The descriptor is normalized by the total number of points in the point cloud in order to compensate for possible noise or shape differences. Occupancy descriptor proposed is available online: https://github.com/margokhokhlova/Occupancy_Decriptor.

Our descriptor design allows it to be used in a multiple camera views scenario to grant a more reliable and accurate pose description. For example, such partitioning was successfully employed earlier for human recognition from complete point clouds [200] based on histograms of normal orientations. To our knowledge, the subject of 3D data based human posture description is not widely researched, and only few artificial datasets are available. Generally, action recognition systems capture human movements using only one camera. However, the usage of a single camera is rather insufficient to ideally capture realistic movement of an object in occlusions. Additionally, a single view of objects has limited field of view of a single camera. The fact that the descriptor

supports the multi-camera setup is beneficial for the scenario, when complete movement of the person's body is to be evaluated. In such scenario the use of the skeletonization algorithms is not straight-forward. They are working on a single view, and two skeletons estimated from two Kinect sensors can have different configurations and postures, and are hard to align as it was shown in the previous chapter (an example can be found in Figure 3.17). An alternative approach is to calculate a 3D point cloud from multiple views and project it to the front-view to get a depth image which is a regular input for existing skeletonization algorithms [187]. With the descriptor proposed in this chapter, we tried to move forward the multi-view posture description from 3D point cloud data.

The descriptor code is developed in C++ and is available for a re-use for the research purposes https://github.com/margokhokhlova/Occupancy_Decriptor.

4.4/ TRAINING AND TESTING DATA

MSR Action3D Dataset [70] was selected to perform the experiments and evaluate the proposed descriptor. This is one of the most used RGB-D human action-detection and recognition datasets. It is also one of the first RGB-D datasets capturing motions (dated 2010) and it contains a big amount of different actions performed by different persons. It consists of 20 action types performed by 10 subjects 2 or 3 times. The actions are: high arm wave, horizontal arm wave, hammer, hand catch, forward punch, high throw, draw an x, draw tick, draw circle, hand clap, two hand wave, side-boxing, bend, forward kick, side kick, jogging, tennis swing, tennis serve, golf swing, pick up & throw. The resolution of the video is not very high, namely 320x240 and so is the frame rate, namely 15 fps. The data was recorded with a depth sensor similar to the Kinect device and contains color and depth video sequences. The sequences are pre-segmented for the background and foreground. An example of superimposed point clouds corresponding to 3 actions from MSR Action 3D dataset is shown in Figure 4.7. Skeleton joints data are also provided with a higher frame rate than the depth maps. However, many joints are wrongly estimated, as can be seen in Figure 4.8.

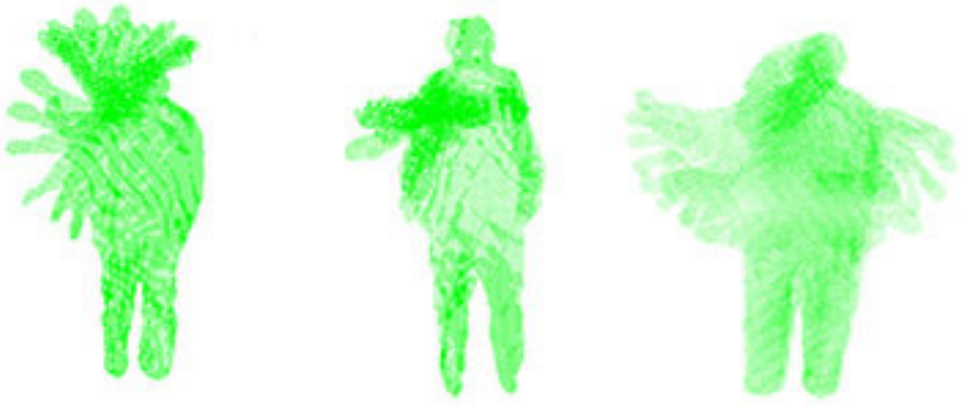


Figure 4.7: Three actions from MSR Action 3D dataset shown as point clouds: high arm wave, horizontal wave, golf swing.

For our experiments, we had to further manually segment the dataset into key postures in 3D. There is no accurate database with full body human poses as depth maps publicly available, despite several works where the features which represent the posture are learned from real and synthetic examples [130][67], neither the data nor the implementation of these methods are available. Recently, a new multi-kinect posture dataset was published [216], however, this one is huge and is not dedicated to the global pose estimation but body parts segmentation. Since we are not using any deep learning and proposing a hand-crafted descriptor, we considered that a well-

Table 4.1: Postures selected from the MSR3D dataset

	Posture	Training	Test
1	Staying relaxed	160	54
2	Forward Kick	102	50
3	Hand lifted 45°	29	13
4	Right hand up	137	64
5	Right hand to the left	80	71
6	Clap	59	25
7	Hands wide open	35	13
8	Pick from the ground	33	34
9	Half bend	80	53
10	Full bend	65	69
11	Right leg kick	60	40
12	Right leg kick on side	49	34
13	Throw from the back	134	78
14	Right hand up	42	28
15	Both hands left half bend	62	30
16	Both hands to the left half bend	62	30
17	Both hands to the right half bend	88	37
18	Throw from the front	54	33

known and widely used MSR Action 3D will be sufficient to perform the test and training to show the capabilities and limitations of our method.

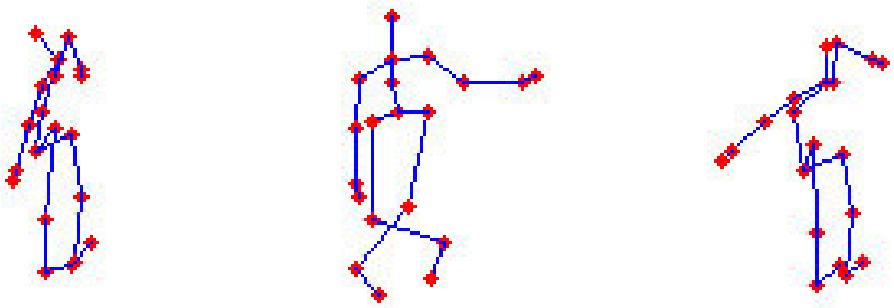


Figure 4.8: Examples of wrong skeleton estimation for MSR 3D dataset, actions 'High Arm Wave', 'Horizontal Arm Wave', 'Hammer'. A person is always facing the camera straight and his legs are not crossed.

In this work, we are aiming to perform a pose recognition without a skeleton aligning or human-body parts segmentation. The number of sequences for each action in MSR Action 3D dataset is between 27 and 30. We separated the data in a training and testing set, and selected between 3 to 7 key poses for each action. For the posture recognition test, 18 well distinguishable poses were selected. The resulting dataset structure is explained in Table 4.1. All the data acquired from person 3 were excluded from the dataset because half of the depth information was missing. Subjects 1-7 from the dataset are used for training and 8-10 for testing. The resulting dataset is not very big but corresponds to our goal to evaluate the descriptive capacities of the proposed solution.

4.5/ EXPERIMENTS

To validate our descriptor, we perform three series of experiments:

Unsupervised clustering of frames into k-postures in a video sequence: A visual evaluation of the descriptor performance, adapted to a video sequence scenario. Designed to demonstrate that body descriptor captures the posture differences. The interest of this test is to see if the descriptor can be used to automatically analyze video sequences, and correctly estimate frames with similar body configuration.

Posture recognition in one action sequence: static posture evaluation for a small number of various postures and small training data. This test is designed in order to see if the descriptor can be used to detect a particular human posture when it significantly differs from the other ones.

Posture recognition for a set of postures: Classification of a set of various static postures. The overall descriptor capabilities evaluation.

The average estimated time for the descriptor calculation (with the 2D-3D transformation performed beforehand) is 0.2 μs on a Intel C602 machine, which is compatible to the time of the extraction of feature vectors in [219].

4.5.1/ UNSUPERVISED K-MEANS CLUSTERING

A simplistic way to compare any two pose descriptors is to calculate an Euclidean distance between them. At first, we observed the dynamics of the distance changes on all the frames of a single video sequence from the dataset. Figure 4.9 visualizes the distance computations for the sequence 'Horizontal Wave' of MSR Action 3D dataset. There are 5 distinctive postures in this action. The result shows that there is a small distance between similar postures (i.e. postures from consequent frames, frames in the beginning and the end of the sequence corresponding to the same 'neutral' posture). To exploit this trend further, a simple test with K-means is performed, which shows that the descriptor captures the posture difference well. Automatic key positions were obtained by performing the K-means clustering for 3 video sequences when one person is performing an action 3 times. In MSR dataset, each person performs the action slightly different (i.e sometimes an action performed several time, arm is under a different angle to the body, etc). It is hard to capture all these differences with few data available. The optimal number of basis K was estimated using the elbow method [120]. Figure 4.10 shows the results of this experiment. The K-NN descriptor performs the clustering of all the input frames, assigning each a corresponding cluster. Qualitative visual analysis shows that automatically detected poses correspond well with the 5 most different poses in the action 'Horizontal Wave' selected manually. In Figure 4.10 the curves on top show the succession of postures detected. Each dot is the frame and the corresponding assigned cluster label (from 1 to 5). Since the curves are similar, the descriptor is able to capture the different postures successfully. The first sequence is slightly different in the beginning since in reality the test subject doesn't start with the neutral posture as in the sequence 2 and 3, the sequence is filmed slightly different. These tests work well for each person performing a single action multiple times, but the test for the whole data gives worse results, probably due to the fact that the neutral posture is dominant in the dataset and people tend to perform similar actions differently. The whole sequences test was designed by just running the k-means algorithm on the sample set of all frames for a given action for all test subjects, and then visualizing the obtained clusters for each sequence separately. As the result, we obtain more intermediate clusters which do not correspond precisely to any of the key-postures. Nevertheless, the obtained results are interesting enough to continue the tests and try to evaluate complete posture recognition based on the proposed descriptor.

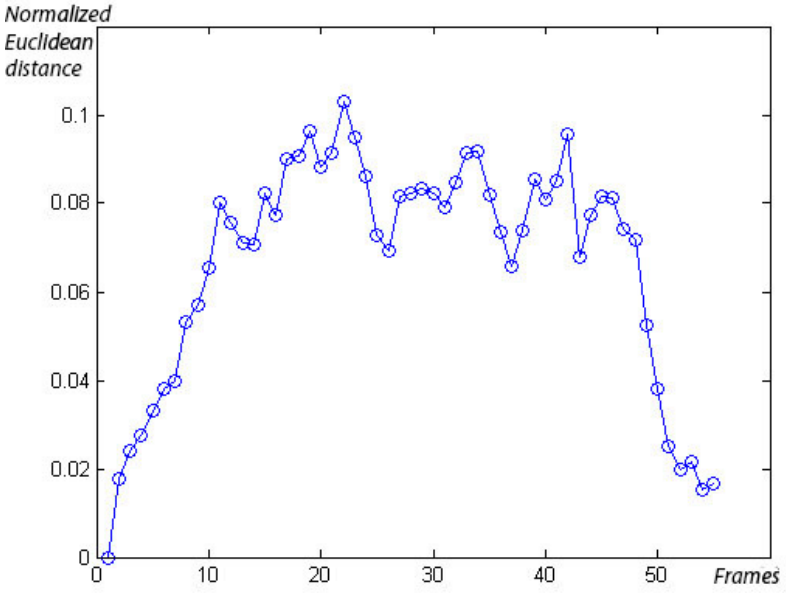


Figure 4.9: Pairwise descriptor distance for all frames of one action from MSR 3D dataset. The video sequence starts and finishes by the same posture. The distance between consequent frames is smaller and distinct 'key' positions can be viewed as peaks of the graph.

4.5.2/ Single Performance Action

A Support Vector Machine (SVM) classifier was trained, One vs All, in order to classify the postures, followed by 3-fold cross validation. One vs all model train one classifier per class in total N classifiers. For class i it assumes i-labels as positive and the rest as negative. This scheme gives us an unbalanced dataset, however, it should work for very different postures. During cross-validation, the SVM completes these steps:

1. Randomly partition the data into K sets of equal size
2. Train an SVM classifier on $K-1$ of the sets
3. Repeat steps 1-2 K times
4. Combine generalization statistics for each fold.

The SVM used in this work comes from CVSVMModel functionality in Matlab. Since we use each time a binary classification, the model with a linear SVM kernel is used.

We included recall and F-measure parameters along with precision in order to evaluate a possibility to use the descriptor in a scenario where accurate retrieval of all postures is essential. For the best results, we are interested in both high Precision and Recall values. The F-measure, recall and precision were used to evaluate the performance of the classifier. F-measure was proposed by [32] to evaluate the optimal segmentation result. F-measure is a trade-off between Precision, the probability that an above threshold pixel is on a true boundary, and Recall, the probability that a true boundary pixel is detected, calculated with the following formula:

$$F = \frac{2(Precision \times Recall)}{Precision + Recall}. \tag{4.5}$$

The results for each posture recognition for the action 'Horizontal Wave' are summarized in Table 4.2. This sequence can be manually separated in 5 extreme postures. Train and test data for this

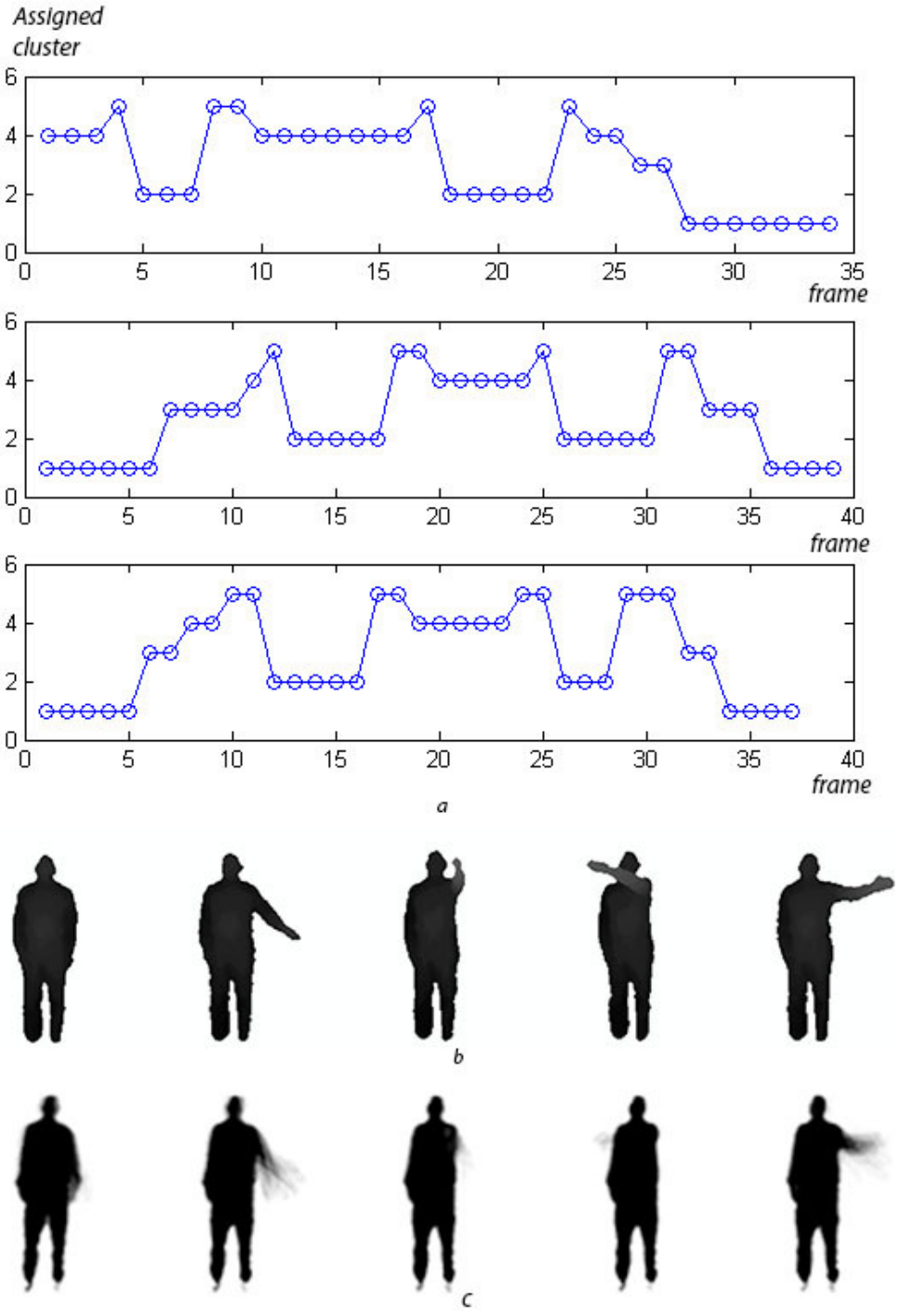


Figure 4.10: a) Three video sequences are shown as a succession of cluster centers. In first sequence person starts to perform the action sooner than in sequence 2 and 3; b) 5 key postures of the action 'Horizontal Wave' (selected manually); c) 5 clusters obtained automatically. Pixel values are averaged: the darker the color is, the more is the occurrence.

sequence were segmented manually according to the scheme introduced in the previous section. This simple test shows excellent results in terms of precision for all but one posture. This scenario

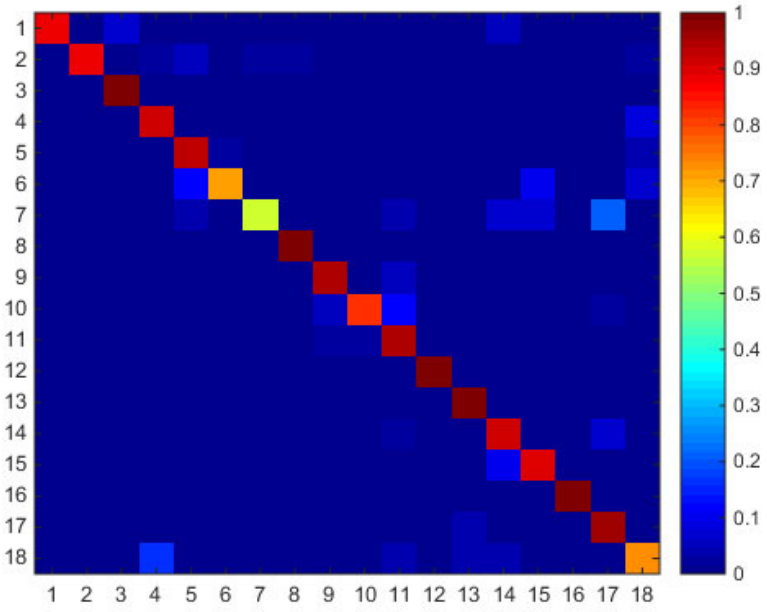


Figure 4.11: Confusion matrix for the SVM-based classification shows good results for all postures but one. The postures enumeration can be found in 4.1.

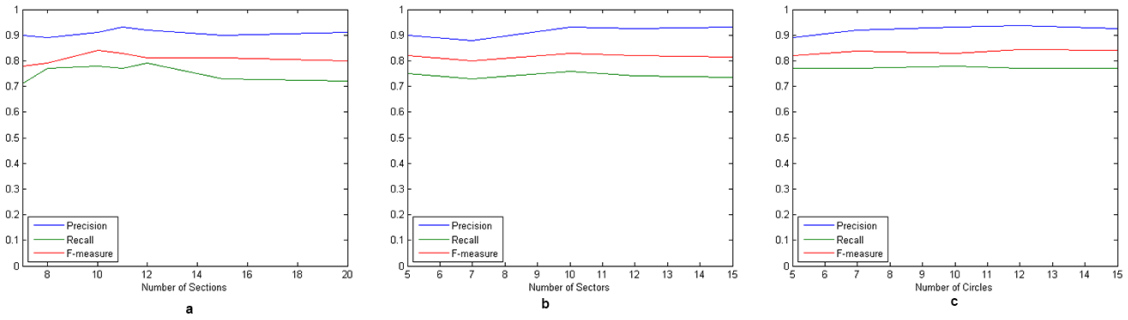


Figure 4.12: Tuning of the parameters. Precision, recall and F-measure curves for a) the number of section varies, sectors and circles fixed to 10; b) the number of sectors varies, sections and circles fixed to 10; the number of circles varies, sections and sectors are fixed to 10.

simulates the possibility to use the algorithm when a small set of postures is present, for example, key-events in gait. However, when evaluating F-measure, only the most different from one another postures have the score above 85: initial posture with arms down, arm in the left side and kick.

4.5.3/ Set Retrieval Performance

The test for a single action posture estimation shows good results, hence we conducted an extended version of this test containing a bigger number of various postures. A full test for 18 postures was performed with an SVM. Feature vectors of the selected postures were used for training and testing. Figure 4.11 shows the confusion matrix for the classes obtained by the SVM. The descriptor parameters (number of sections, circles and sectors) were tuned for the best performance. We obtained the best results with 12 sections, 10 circles and 10 sectors, corresponding average precision is 0.94. The parameter tuning is straight-forward and shows that the different

Table 4.2: Classification results for 5 postures of the action 'Horizontal Wave' show good results in terms of precision.

Posture	initial	arm 45°	kick	arm front left	arm right
Precision	0.94	0.81	1	1	1
Recall	1	0.8	0.76	1	0.71
F-measure	0.97	0.82	0.86	1	0.83

parameter combinations do not have much of an effect on the performance. The main observation is that for postures selected the most important parameter is the number of sections which helps to separate the volume by vertical planes. Different combinations of parameters can give slightly better or worse results in terms of precision, recall and F-measure. Corresponding curves obtained for different parameters are shown in Figure 4.12.

The results show good performance in terms of precision which is excellent for simple postures. Our results are comparable with the results of [219] where authors are using only 5 distinct postures: standing, sitting, stooping, kneeling and lying. Of these, several postures are similar to ours, plus we are aiming at more complex and varied postures. The original dataset of [219] is not available, but we also performed a test with just 3 very different postures and a similar amount of training and testing data. As before, the training data and test data are formed from different subjects. Our postures are: staying, right-hand up, bending. The corresponding numbers of training and testing images are: 384/125, 246/125, and 98/103. With this small dataset we obtain excellent results in term of precision and recall, all the tests are assigned correctly. Our results and the results from [219] cannot be directly compared due to the absence of the dataset used by the authors, but this test gives an idea about the descriptor capabilities. Wang et al. test their posture recognition method on 80-100 depth images taken each for 8 persons. The recognition rate is also very high, with some minor errors (for example, for the first person the recognition rate is: 79/80, 99/100, 80/80, 80/80, 79/80). It should be mentioned, that Wang et al. [219] use same subjects for testing and training, which is probably easier as we have shown in our tests from the previous section.

4.6/ CONCLUSIONS

This chapter introduces descriptor designed for human posture recognition from Point Cloud data based on the surface points spatial distribution. The introduced descriptor works well for capturing the 3D spatial arrangement of a point cloud structure. Experiments show that our method achieves competitive results compared to current hand crafted state of the art descriptors. Learned or trained descriptors may give superior performance but critically depend on the availability of large amounts of labeled data. Secondly, these architectures don't generalize outside their initial domain. Our algorithm is a simple and elegant solution, when joint information is not available or unreliable.

Example targeted applications include action recognition and gait analysis. For the latter, the descriptor may be deployed for cycle event or symmetry detection and evaluation [168]. Another possibility is to be able to divide a video along the time axis using posture information in the case of misalignment. Detected postures can be used to temporally align the data or as key-words describing the action. Other applications include video surveillance, video indexing/browsing, recognition of gestures, human-computer interfacing or analysis of sport-events. In real world, human acts essentially in 3D space. Therefore, a 3D representation is more informative than the

analysis of 2D movements carried out in the image plane, which is only a projection of the actual motion.

There is a number of open issues. The descriptor is noise-sensitive, which becomes more apparent if part of the depth data is missing. Secondly, the Euclidean distance metric between two descriptor vectors currently excludes 3D spatial information. Semantically different postures can thus result in descriptor vectors that are near similar. The last issue, being fast to compute, the descriptor still requires more storage space than a skeleton representation.

Future work is to address these issues next to developing an application for real-time gait cycle event recognition.

This chapter shows that body pose may be adequately represented without joint estimation using the data from a 3D camera. The proposed descriptor can be used exclusively, or as an advantageous addition to tradition skeleton-joints estimation methods. We crafted this solution in order to avoid using the skeleton representations and inaccurate algorithm in the Kinect v.1 SDK.

Finally, the descriptor can be interesting to work with complete point clouds obtained with stereo Kinects setup described earlier in this work in Chapter 3. However, it does not outperforms the skeleton representation for intra-persons posture recognition using the Kinect v.2, where the skeleton representation is more reliable. Therefore, when our laboratory changed the hardware, we opted for the skeleton representation of the human posture available in the Kinect v.2, along with a single sensor statically placed in front of the patient. Our experiments with a multi-setup proved it to be a difficult solution for a gait assessment platform, and with a single Kinect the ML based skeletonization algorithm based on a higher quality data from depth sensor is a better solution. The descriptor proposed here was therefore not developed further. In the next chapter of this dissertation we face the problem differently, and present our experiments with gait assessment using skeleton data from a Kinect v.2 sensor.

5

SKELETON BASED GAIT CLASSIFIER

This chapter is dedicated to a complete gait classification system using the Kinect v.2 sensor. Previously, we explored the gait parameters in Chapter 2, the parameters, and characteristics of modern 3D sensors, including the possibilities of different setups in Chapter 3. Chapter 4 described the posture descriptor based on 3D point clouds.

During the work on the first parts of this thesis, we changed the Kinect v.1 camera for the Kinect v.2, which apart from incremental updates, uses Time-of-flight instead of structured light for its depth estimation. The Kinect v.2 sensor gives more reliable skeleton data and was used in gait analysis and classification scenarios by other researchers. We also discovered the optimal conditions under which the existing skeletonization algorithms perform best. Therefore, at this moment, we are ready to propose a final algorithm for the gait assessment task. In the contrast with the algorithm presented in the previous chapter, here we use skeleton data and not the point clouds.

As stated earlier, we address the patients with a prosthesis. While assessing amputee gait, it is important to be aware of normal gait parameters and how normal gait in the amputee is affected. There may be deviations when an amputee will adapt to compensate for the prosthesis, muscle weakness or tightening, lack of balance or lack of flexibility. These deviations [21] create an altered gait pattern and it is important that these are recognized, as successful rehabilitation of the gait means to apply corrective measures to compensate for or eliminate these deviations. Such deviations vary depending on the amputation level and prosthesis type. They mostly affect the kinematic parameters of the gait, especially knee flexion and rotation [255], listed in Table A.1 in Chapter 2.

We decided to exploit the Transtibial gait deviations in order to perform pathological gait classification. This decision is made due to three reflections. First, kinematic parameters are the ones used by Proteor as shown in Annex A to evaluate the gait. Proteor [8] is the company specializing in prostheses and medical corsets, and it provides this thesis with clinical and industrial insights. Second, gait deviations affect one another. For example, an absent knee flexion will affect the hip flexion. So from kinematics, we can observe some transfemoral deviations, such as abducted gait, circumduction, and prosthetic instability. Third, recent methods assessing the gait use 2D kinematic gait parameters successfully [211].

Earlier in this work, Section 3.2.2 presented the procedure to calculate knee and hip flexion angles from Kinect v.2 orientations. We are using this procedure to calculate the kinematic gait parameters used for gait description in this chapter.

First, the task of clinical gait assessment is described in section 5.1. Second, the multi-model gait symmetry database acquired in our laboratory is described in 5.2. Third, we present a new covariance-based binary classification method of normal versus abnormal gait in 5.3. The covariance features are based on the kinematic gait parameters for knee and hip skeleton joints. We focus on the binary gait classification, but also assess a multi class option in order to compare the results with the ML-based classification. Finally, in 5.4 We present LSTMs-based compound Gait Model, which is taking into account kinematic gait features dynamics in order to perform an unsupervised gait assessment.

Contents

5.1/ GAIT ASSESSMENT

Gait assessment is a task of gait observation for the presence of deviations traditionally performed by clinicians. Accurate reliable knowledge of gait characteristics at a given time, and even more importantly, monitoring and evaluating them over time, will enable early diagnosis of diseases and their complications and help to find the best treatment. In the case of patients with a prosthesis, gait analysis can help to find a better fitting and restore a more optimal gait for the patient.

Traditional gait assessment methods provide subjective measurements and do not guarantee accuracy, repeatability, or reproducibility, which adversely affect the diagnosis, supervision and treatment of gait pathology.

The first gait assessment methods date back to the early nineties and use standard conventional cameras. Based on video data, several protocols for gait assessment were proposed. They commonly specify the list of gait characteristics which should be assessed during the diagnostic in a form of a check list. For example, the two commonly used are:

- **Gait Abnormality Rating Scale (GARS) [18]** is a video-based analysis of 16 human gait characteristics. The GARS includes five general categories, four categories for the lower limbs and seven for the trunk, head, and upper limbs. The protocol can also be used as the screening tool to identify patients at risk for injury from falls.

- **Extra-Laboratory Gait Assessment Method (ELGAM) [17]** is a method to evaluate gait in the home or community, mostly targeting the identification of risk factors for falls among the elderly. The parameters studied include step length, speed, initial gait style, ability to turn the head while walking and static balance. Researchers link low speed (under 0.5 m/s), short steps, difficulty turning the head and lack of balance to unstable gait.

Research works dedicated to gait analysis often elaborate custom protocols and groups of parameters to consider. Many works concentrate on the estimation of particular gait parameters from 2D video sequence [132, 64]. Some works are also present in 2D based gait recognition [45, 38]. However, quickly researchers realized the limitations of such methods and turned their attention towards multi-camera setups [133]. At first, the calibrated setup of multiple passive sensors allowed to restore the 3D coordinates of the scene, however, this direction was mostly abandoned when off-the-shelf depth sensors appeared. Recently, gait analysis research is predominantly performed with RGB-D cameras or wearable sensors. The protocols used vary, however, mostly researchers use statically placed Kinect and person walking a given distance [169, 250, 153].

Earlier studies have shown that the Kinect sensor can be exploited to examine gait. The validation of the Kinect v.2 sensor was earlier performed in 3.1.4 section of this thesis. The latest binary gait classification was reviewed in 2.3.1. Gait assessment methods from skeleton data is presented in 2.3.2.

Usually, to validate their algorithms, researchers use small custom datasets acquired in their facilities. The datasets overview can be found in Chapter 2, Section 2.1.5. The realized database review shows, that Rocha et al. [156] used just 3 training and 3 testing subjects and the data is not available. Paiement et al. [153] collected data from 20 subjects and shared them for the following research. Chaaraoui et al. [169] gathered data from 7 actors imitating 2 gait pathologies. The data are available online. Devanne et al. [199] used two earlier datasets [153, 169]. Nguyen et al. [211, 251] used earlier data from VICON and also created a new dataset with 8 subjects simulating gait deviations on a treadmill with the use of Kinect v.2 sensor. Li et al. [250] used a small custom dataset which they do not share.

5.2/ New Gait Symmetry Database MMGS

Recent gait assessment algorithms [250, 211, 169] report accurate classification results. However, the used benchmarks are either very small with only 56 sequences in [169], contain lots of noise [153], or are taken in special conditions [251], such as with a use of a treadmill.

The collection of patient data requires many permissions and a special laboratory setup. In the same moment, the research teams working on computer vision-based motion analysis do not commonly have access to clinical environment. This results in a situation where there are very few benchmark datasets publicly available in the gait assessment field. Earlier in this work we present the procurable gait datasets in section 2.1.5.

The absence of data is a big issue in gait clinical studies. Normal gait is more prevalent, since we can use data from the gait recognition domain, where less restrictions apply. The amount of publicly available gait data is small compared to the number of gait studies that have been performed over the years. The data that is available generally suffers from limitations such as few subjects, few gait cycles, highly clinical, no raw data, lack of meta data, non-standard formats, and restrictive licensing [183]. However, there are very few examples of the abnormal gait publicly available for a reuse. We collected the information about the available gait datasets acquired with a RGB-D camera devices and shared it with other users. It should be noted, that there exist more databases collected with other devices, such as 2D cameras, golden-standard systems, WS and others. A work [183] reviews many of these databases, so we don't include them in the current work dedicated to Kinect sensor.

The biggest multi-modal pathological gait database among all to our knowledge is the recent one by [251]. However, Nguyen et al. do not provide the orientations and joint state data. In addition, a treadmill slightly affects kinematic parameters [102], so this data cannot be used in the more common solid surface walking scenario. To continue with the promising direction of gait assessment, more data is required. This data should satisfy the multi-modal criteria, and be captured in normal conditions.

To meet the criteria specified, we had to construct our own dataset. A database of normal and pathological gait examples was collected in our laboratory. The goal of this database creation was to create a benchmark for gait classification algorithms, focusing on gait deviations characteristic for prosthetic patients. It is a multi-modal gait database, which can be used for the gait symmetry analysis, or MMGS for short. We used a single Kinect v.2 sensor placed in front of the subject, and each person walked towards the camera. The setup is shown in Figure 3.5 (left) in chapter 3, and was selected after the initial trials described in chapter 3. Each walk was executed between 5 and 7 times. A total of 22 persons participated in the experiment. We found 3 females and 19 male participants, with a mean age of 31 years and a standard deviation of 8. None of the participants was previously diagnosed with a gait affecting disease. Detailed information about subjects who participated can be found in Table A.6 in the Annex A.

The software to acquire the data from Kinect was written on the base project KINECT for Windows SDK programming [5]. The acquisition protocol consisted of three steps. First, each person was instructed to walk normally with a self-assigned comfortable speed from a starting line towards the camera. Second, we added a padding sole into the right shoe of each patient. The height of the padding material is 7 cm (see Figure 5.1). Similar sole padding to simulate a pathology was already performed in [251]. Padding allows us to simulate limping gait. Third, we asked the person to not bend the right knee during walking as it was done previously by [169]. This test is designed to simulated gait problems, which can be due to the wearing of a prosthesis or characteristic for recovery after a fracture. Second and third parts of the experiment were performed after an initial time, when the person walked around and practiced the simulation. The first normal trial was executed immediately after a person was instructed.

We stored the following four types of data:

- Depth maps of the scene.

Figure 5.1: Sole padding used in this work. The height is equal to 7 cm. We placed the padding into the right shoe.

- Binary masks with the silhouettes.
- Skeleton joints 3D coordinates (X, Y, Z) and their state (i.e., tracked, inferred, not tracked).
- Bone orientations (w, x, y, z).

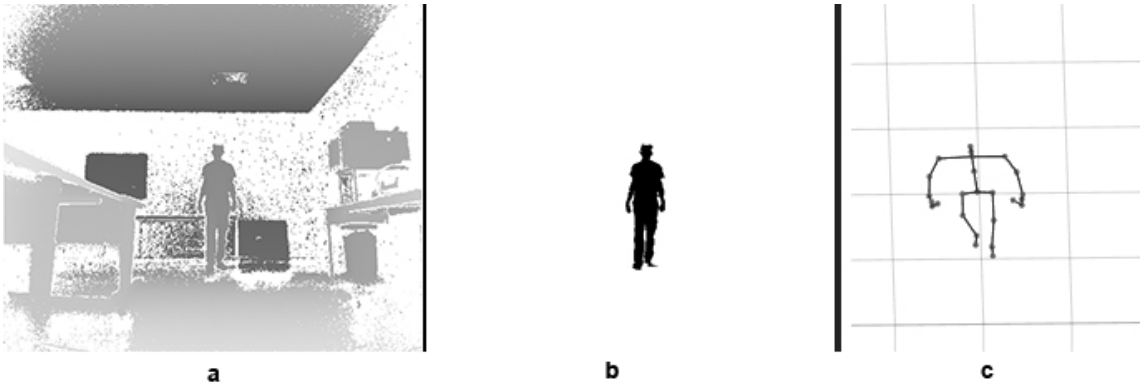


Figure 5.2: Database contents. a) depth image inverted for visualization purposes. b) user mask. c) skeleton data.

The MMGS data is available online: https://github.com/margokhokhlova/LSTM_gait_model, the file $data\_base\_joints$. The data in plain text format can be freely downloaded and used for research purposes.

5.3/ BINARY GAIT CLASSIFICATION

With the prior work on kinematic parameters computation described in Chapter 3 and MMGS dataset presented in the previous section 5.2, we are now ready to experiment with different gait descriptors.

First, we aimed for the binary gait classification task similar to [211, 250]. We are interested in simply evaluating the patient's gait in order to tell if it has a normal or pathological pattern. In other words, we seek the means to create a normal gait model.

Figure 5.3 shows the principal steps of the proposed binary gait assessment algorithm. We use skeleton joints and orientations from a Kinect v.2 sensor to compute kinematic gait parameters. Then covariance-based features are used to build a normal gait statistical model and to train a

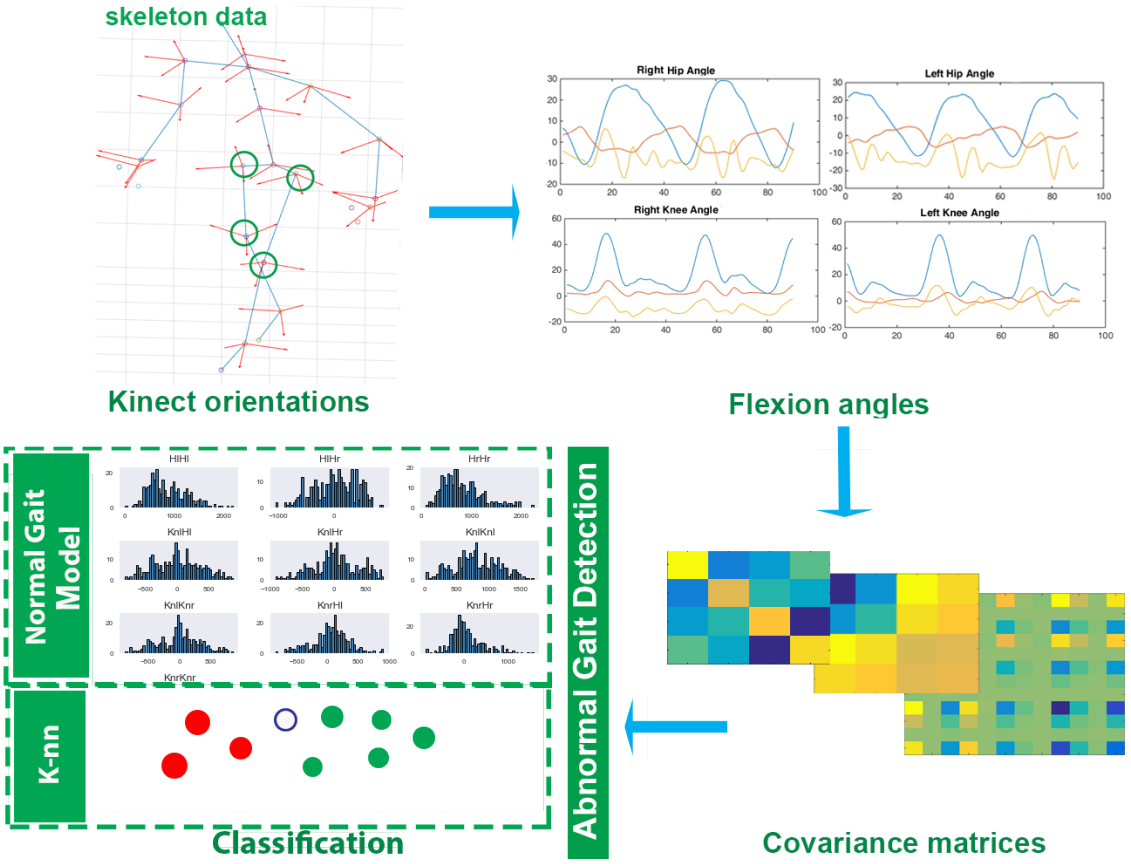


Figure 5.3: The framework of the proposed gait assessment method. Follow the arrows: Kinect orientations are modified to match the VICON system data and X, Y, Z angles are calculated for hip and knee joint (shown by green circles). The resulting angles are filtered. The covariance matrices are computed from flexion angles. The final classification is performed using a normal statistical model and a K-NN classfier.

K-NN classifier. Lastly, this model is used to detect a pathological gait. Our method is inspired by the works [211, 250]. In our work we propose to use compact covariance features based on low-limbs flexion angle data only. We choose to use angles over joints directly because they are more robust and vary less between individuals. The main novelties of the proposed algorithm are the following:

- We selected to use only low-limbs joint data and excluded the ankle data as the least reliable, in contrast to the methods using all 25 joints [250, 169] or all low-limb joints [211]. This arrangement was done to fit the specifics of the Kinect v.2 skeleton data reliability better.
- We propose to use flexion and rotation dynamics calculated from orientation data, validating our features with a comparison with a golden-standard MOCAP system.
- We use a covariance descriptor based on angles and not on the joints data directly.

5.3.1/ Covariance-based Descriptor

The two works that are most closely related to the proposed were presented by Ngueyen et al. [211] and Li et al. [250]. We extend the kinematic parameters used by [211] and propose to encode them with a covariance matrix representation.

Covariance is a measure of the extent to which corresponding elements from two sets of ordered data move in the same direction. Covariance matrices are discriminative features for action recognition from skeletons as shown in [226]. Lately, they were also successfully exploited for gait assessment in [250]. One of the advantages of the covariance matrices is the fact that they provide a feature representation, which is independent of the cycle period. It allows to obtain a gait feature with constant size, independently of the speed variations in a subjects' walking manner. The latter holds true when the matrix is calculated over a significant number of frames.

Flexion angles are known and widely used features for gait analysis in clinics and hospitals. We combine the compactness and representativeness of the covariance matrices with the highly relevant flexion angles features to propose a new gait assessment method. A covariance matrix summarizes the relations between the hip and knee flexion angles for a gait sequence. Using the covariance features, we aim to describe the symmetry of the gait based on lower limbs flexion angles.

We adopted traditional variance-covariance matrices (Equation 5.1) and covariance matrices (Equation 5.2) as described in [226] for our tests.

$$cov = \frac{1}{(T-1)} \sum_{t=1}^{T} (P_a - \mu)(P_a - \mu)^T, \tag{5.1}$$

$$cov = \frac{1}{(T-1)} P_a (\frac{1}{T} I_T - 1_T) P_a^T, \tag{5.2}$$

where μ is the mean of P_a, T is the number of frames, I_T is the identity matrix, $1T$ is the $T \times T$ unity matrix and P_a contains Knee and Hip angles coordinates for all the frames of a given sequence.

$$P_a = \begin{bmatrix} Kl_1 & Kr_1 & Hl_1 & Hr_1 \\ \vdots & \vdots & \vdots & \vdots \\ Kl_T & Kr_T & Hl_T & Hr_T. \end{bmatrix} \tag{5.3}$$

Kl_i, Kr_i are knee Cardan angles and Hl_i and Hr_i are hip Cardan angles for each frame i. Note that the obtained matrices should not be further normalized since the flexion angles are already within the same scale.

The classical covariance matrices obtained by the equation 5.1 are well-known and widely used features, generalizing the notion of variance to multiple dimensions. The covariance matrix calculated by the equation 5.2 is a symmetric and positive definite matrix. The authors don't specifically explain the design of their matrix, however, by the formula given we can see that instead of substracting the mean value, the matrix diagonal values (representing variance) will be normalized by the number of frames, (because $\frac{1}{T} I_T - 1_T$ will be a TxT matrix with -1 everywhere but the diagonal, which is $\frac{1}{T} - 1$). This matrix has shown good results for action recognition experiment performed in [226].

To summarize, we propose to use the knee and hip flexion covariance as a gait symmetry feature. In the following, they are simply called covariance features.

5.3.2/ Experiments with Covariance Flexion Descriptor

To evaluate our proposed covariance-based gait features, we first use the existing data from other researchers for the our tests. We performed a serie of experiments to validate the proposed approach. The main goal is to evaluate the possibility of the covariance features to be used in a binary normal/abnormal gait classification task. The experiment design is conditioned by the nature of the datasets. Cross-dataset evaluations are made in order to show that our method

can be generalized towards different data. Due to the different purposes of the datasets used, we had to elaborate different testing protocols for the cases when a) there are pathological data in training and testing part of the dataset; b) there are pathological data in the testing part only. We use statistical normal gait modeling when the training part of the dataset contains only one type of data (i.e normal gait samples), and a K-NN classifier otherwise. K-NN is a non-parametric classification method, which classifies an object by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. A Gaussian Mixture model (GMM) corresponds to the mixture distribution that represents the probability distribution of observations of the overall examples. Normal gait model was used since for one of the databases [169] the testing scheme suggests to learn from normal gait data only, so it is impossible to use K-NN. We use SVM in one of the tests, because initially we started to experiment with a single database [169] to verify if the our covariance matrices can be used for gait classification for small 'toy' example. The obtained results are interesting, so we provide the results of the simple trial using an SVM in this work. A K-NN classifier was selected because of the small number of testing and training data available in the pathological gait datasets. It was easier to use a method which does not require 'learning' from the data. The K-NN was therefore selected and used for the main tests.

5.3.2.1/ DATA USED

The absence of data is a big issue in gait clinical studies. Normal gait is more prevalent, since we can use data from the gait recognition domain. Nevertheless, there are very few examples of the abnormal gait publicly available. However, even if the dataset is shared, quaternions data saved along with the joints coordinates is not a common practice. Initially, we use the following skeleton datasets in our experiments:

- UPCV Gait K2 [204] contains data from 30 persons performing a normal walking routine 10 times. It was acquired with a Kinect v.2.
- The DAI dataset [169] contains 7 actors performing normal and abnormal gait. It was acquired with a Kinect v.2.
- The SPHERE-Walking2015 dataset [153] contains normal and abnormal gait sequences from a Kinect v.1 sensor. There are 21 normal sequence and 20 sequences of patients with Parkinson and Stroke diseases.
- The walking gait dataset [251] was recently proposed by Nguyen et al. It has been established to enable comparative studies on gait analysis, especially the problems of gait index estimation and abnormal gait detection. The dataset includes 9 normal gaits and 8 simulated abnormal (asymmetric) ones performed by 9 individuals on a treadmill. Abnormal gaits were simulated by attaching a weight to a foot and padding a sole. It was acquired with a Kinect v.2.
- CHU Dijon dataset used in Chapter 3. Contains 22 normal sequences acquired by a Kinect v.2 camera in CHU Dijon.

The datasets proposed in [169] and [251] perfectly suit the proposed algorithm specifics, since actors imitate different anomalies affecting the symmetry of the gait. However, the number of sequences is small. We enlarged the normal dataset with data coming from a gait recognition dataset [204] captured with the Kinect v.2, assuming that the people do not have a pathological gait, and by adding the Sphere dataset and the data acquired in our lab. The latter containing 22 normal walks performed by 2 healthy individuals and were earlier used in our experiments in Chapter 3. We provide the results obtained on each dataset separately and also perform some experiments on a mixed dataset as was previously done by [211]. Information about the employed datasets is grouped in Table 5.1.

Table 5.1: Datasets used in this work

Dataset	Normal seq	Abnormal seq	Normal cycles	Abnormal cycles	Data	Comment
DAI [169]	28	28	28	55	25 joints	Normal/abnormal sequences performed by actors
UPCV Gait K2 [204]	300	na	1555	na	25 joints and 20 orientations	Gait recognition, 30 subjects, 10 trials
SPHERE-Walking2015 [153]	21	20	65	162	15 joints	Normal, PD and stroke patients
Walking gait dataset [251]	9	72	17047	1405	25 joints, point clouds	A treadmill is used, padding sole and attaching weight to simulate the pathological gait
CHU Dijon	22	n/a	64	na	25 joints and orientations	Normal subjects, clinical conditions

We pre-process the data from all datasets in the following way. We apply a low-pass filter to the joints and quaternions data. A standard normalization is then done by subtracting the center of the spine joint from other joints for each frame.

5.3.2.2/ TESTS

We performed series of experiments to validate the proposed approach. The experiment design is conditioned by the nature of the datasets. Cross-dataset evaluations are made in order to show that our method can be generalized towards different data. In all reported tests we used a covariance matrice for 2 low-limbs angles coming from the left and right feet. The angles were calculated using orientations for the datasets, where orientations data was available, and using joints when they were not. In this case, a simple geometric cosine rule was used for the triangles hip-knee-ankle and mid spine-hip-knee. In most parts of the tests we use only Z-flexion angles to compute a 4×4 covariance matrix. However, we get better classification results using flexion, abduction and rotation information on the datasets when quaternions data are available to calculate the parameters with the method presented in Chapter 3. We provide a comparison of the results with external/internal rotation angles and abduction/adduction angles. We compare our method with state-of-the-art methods [211, 169, 199] on the available datasets and perform a generalization of the results on custom dataset similar to [250]. In the following each subsection is dedicated to the experiments with a particular dataset.

In order to use a covariance matrix as a feature in a classifier, a natural choice consists in vectorizing it in order to process this quantity as a vector and then use any vector-based classification algorithms. We also took this approach in the current work. Covariance matrices are symmetric, so we exclude the values below the diagonal in our final descriptor. An example of a single covariance matrix and the unrolled covariance matrices from Z angles flexion calculated for all sequences of the DAI dataset is shown in Figure 5.4. In this case, the initial size of the covariance matrix is 4x4, and the diagonal and upper triangle values form a final 10 dimensional descriptor.

5.3.3/ COVARIANCE FEATURE SELECTION BASED ON DAI DATASET

The Dai dataset contains just 7 persons. The first initial test we performed was to see which covariance matrix works better with angle data. Since the number of sequences is very small, we constructed the following experiment. We trained an SVM on all persons but one and then used the SVM to assign labels for the sequences for the remaining person. The whole sequences were used to calculate the covariance feature, which means that for the test we had just 4 normal and 4 abnormal sequences. The SVM used CVSVMModel functionality in Matlab is used. Since we use each time a binary classification, the model with a linear SVM kernel is used. We tested the matrices calculated by equations 5.2, 5.1, and also tried to normalize matrices calculated by

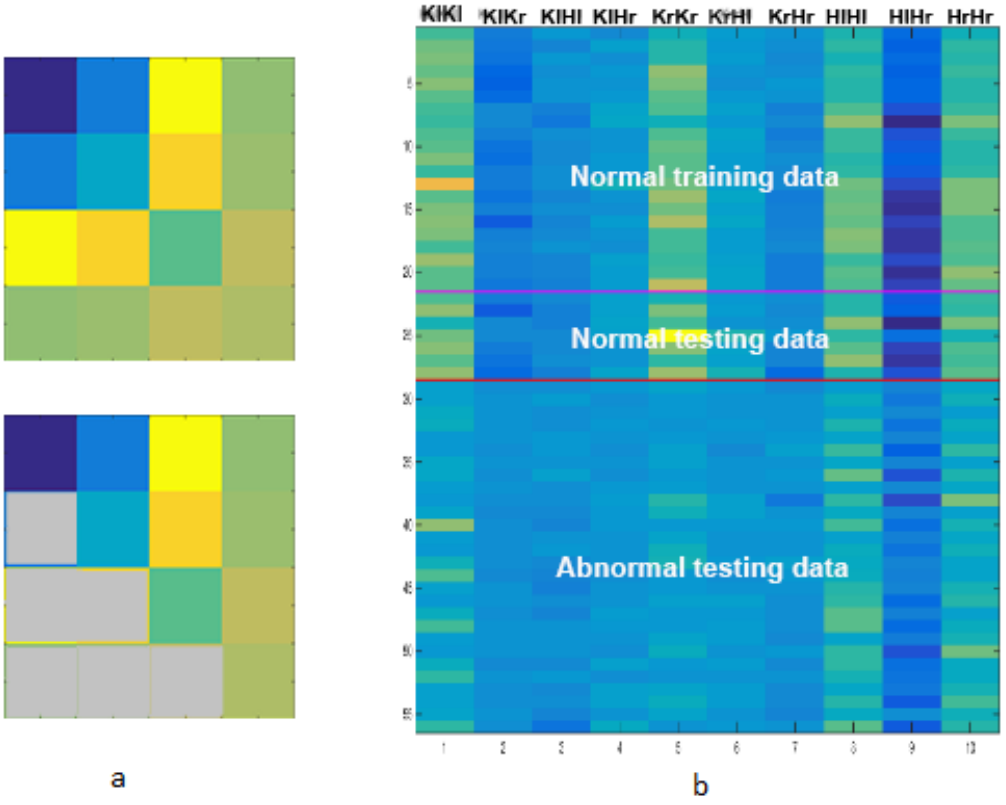


Figure 5.4: a) An example of the covariance matrix and its unique elements. b) Covariance matrices unwrapped for all sequences from Dai dataset. Lines divide the dataset into training and testing data as advised by [169]: 21 normal learning sequences (3 from each actor) & 35 testing sequences (all remaining). The matrices are colored-coded for the visualization purposes. Even visually the normal and abnormal gait descriptor values are different, having less positive (represented by yellow values) in abnormal sequences.

the equation 5.2 by the diagonal value. The selected normalizing is the one commonly used, when the matrix is normalized such that each element of it's diagonal is equal to 1 in order to compensate for possible difference in energy between two signals. So, the covariance matrix is then scaled and represents only the relation between the different flexion angles independently of each particular signal amplitude.

The tests result show the following trends. This test showed us that using abduction, rotation and flexion angles for a person recogntion (correspond to ψ, χ, ϕ and are named XYZ in the following) is better than using a single flexion angle in one plane. The experiments with using different types of covariance matrices (normalized and non-normalized) show the superiority of the Cardan angles over a single flexion angles by 5-12 percent. The results are grouped in Table 5.2. The formula used in [226] to calculate covariance matrices for action recognition works worse than a standard covariance matrix in the context of our data. The normalization of the variance decreases the accuracy when the standard covariance matrix is normalized. Therefore, in the following tests we use the standard covariance matrices without normalization.

5.3.4/ The Normal Gait Model

The DAI dataset [169] design requires one to use only the normal examples for training and then detect the outliers (i.e., pathological gait sequences) during the test. We developed a normal gait

Table 5.2: F1 score for each person and average F-score for all persons.

Metod	Alex	Javi	Jero	Jose	Juanmo	Mario	Rafa	F-score average
Z rotation angles, CovMat Eq.5.2	0.44	0.66	0.58	0.84	0.48	0.48	0.83	0.62
XYZ rotaion angles, CovMat Eq. 5.2	0.75	0.59	0.84	0.63	0.41	0.56	0.84	0.66
Z rotation angles, norm cov matrix Eq. 5.2	0.7	0.56	0.41	0.47	0.51	0.9	0.42	0.57
XYZ rotaion angles, norm cov matrix Eq. 5.1	0.83	0.88	0.64	0.73	0.73	0.75	0.77	0.76
Z rotation angles, non-norm cov matrix Eq. 5.1	0.83	0.88	0.64	0.73	0.73	0.75	0.77	0.76
XYZ rotaion angles, non-norm cov matrix Eq. 5.1	0.91	0.77	1	0.77	0.9	0.89	0.98	0.88

model based on the training data. We choose the Gaussian distribution to model our data.

The authors specify 2 scenarios for testing. We also suggest a third scenario scheme.

- The first scenario uses all the actors, learning the from the normal gait sequences and then assessing 4 pathological and 1 normal sequences for each actor.
- The second scenario uses the same division scheme but leaves 3 unseen actors for the test.
- The third scenario we propose, uses both normal and abnormal sequences from the first three actors to train a simple SVM model which is then tested on the resting four actors.

Table 5.3: Normal gait model results, DAI [169]

method	scenario 1	scenario 2	scenario 3
[169]	0.98	0.85	na
[199]	0.98	**0.96**	na
ours	0.98	0.88	1.0

Table 5.4: Different features, DAI dataset [169]

angles	scenario 1	scenario 2	scenario 3
Z	0.90	0.86	**1**
Z, Y	**0.98**	**0.88**	0.94
Z, Y, X	0.68	0.60	0.87

We represent normal poses by their probability mass function (pmf) $f(cov)$. We obtain this pmf from covariance features from the traning data. Let X be a discrete random variable with range $RX = x_1, x_2, x_3, ...x_k$ (our training normal examples). The pmf function is then:

$$PX(x_k) = P(X = x_k) \tag{5.4}$$

The example is shown in Annex A.3. Then we evaluate each training sample by the model to obtain a negative log likelihood $(nlogL)$ as the measure of similarity. The threshold for the normal sequence is estimated as a $\mu \pm 3\sigma$, where μ is the mean and σ is the standard deviation of the normal $nlogL$ data. For the test set, all the sequences with the log-likelihood according to the normal model higher than this threshold were considered as abnormal.

Results for the test schemes proposed by [169] are summarized in Table 5.3.

Finally we explored the possibility to use different combinations of the angles to find the most discriminate feature. Results in terms of accuracy are presented in Table 5.4. We used flexion only, flexion and abduction and flexion, abduction and rotation angles. The same procedure was

established to build the model, with one distinct difference. Since our covariance matrices are becoming bigger, we use the eigen-values to compare two matrices as proposed by [101]. This allows us to reduce the size of our features. The results shows great potential for the use of flexion and abduction angles for symmetry assessment. However, the rotation angles are not very robust and lead to a decrease in overall accuracy. Our results are comparable with the results published by [199, 169]. However, we believe our method to be more general and to work well when more data is available.

5.3.5/ K-NN Classification on Walking dataset

The next group of experiments was performed on a unit dataset created by Nguyen et al. [251].

We selected a non-parametric learning algorithm K-NN for the classification performed later in this section. K-NN does not make any assumptions on the underlying data distribution, is fast and easy to use and can give high accuracy.

We used the division scheme proposed by the authors where 5 subjects are used for training and the others for testing. Each sequence has 1200 frames, so to increase the number of samples, we use the sliding window approach to segment the data. Similar window-based signal segmentation was used in gait assessment researchers [250, 211]. We opted to use non-overlapping windows, in order to have independent data for the experiment. Training and testing data are segmented with window size of T frames. Earlier we estimated that the cycle length in this dataset is between 15 and 25 frames. The optimal window size T estimated in our tests is equal to 90 frames. The following results are provided using T = 90. For each sequences from the dataset we then have 13 unique gait samples.

Table 5.5: Results on Walking dataset [251] with K-NN

method	F-measure	precision	recall	accuracy
[211]	0.79	0.93	0.70	0.68
ours	**0.86**	0.76	**0.99**	**0.84**
joint cov by Eq 5.1	0.82	**0.95**	0.73	0.79
joint cov by Eq 5.2	0.69	0.55	0.90	0.13

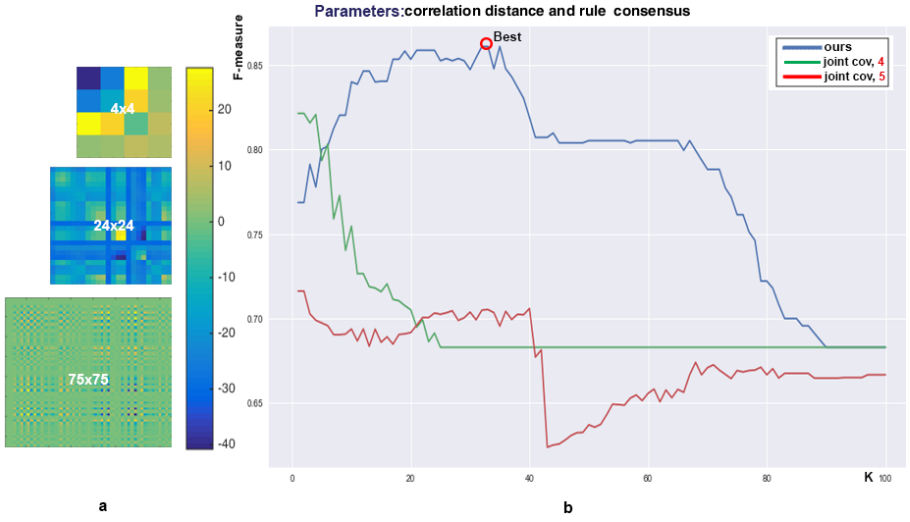


Figure 5.5: a) Color-coded covariance matrices. b) Corresponding F-measure curves for results summarized in Table 5.5. The covariance angles-based descriptor shows better results than skeleton-based covariance matrices.

Since our goal is to detect abnormal samples and the dataset contains more abnormal (positive)

than normal (negative) examples, we construct the abnormal part of our test in the following way. We divide the dataset such that the total number of positive examples is equal to the number of the negative examples. The positive example set is equally divided to contain an identical number of sequences from each pathology. The pathological sequences are assigned randomly from the whole set. This test set construction procedure is performed 20 times in order to cover the dataset.

Each time we calculate the precision, recall and F-measure based on the number of k neighbors in the K-NN algorithm. The correlation distance is used and the concensus model are employed as the parameters of the K-NN algorithm. Correlation is calculated as one minus the sample correlation between points (treated as sequences of values), concensus rule is a parameter of the voting process, when the final label assignment requires a consensus, as opposed to majority rule in the voting process.

We also tested the eigenvalue-based distance between matrices as [250] and Riemannian manifolds using the formula 5.5, but it was outperformed by the Euclidean and Correlation distance.

$$D = \sqrt{\sum \log \lambda(M_1, M_2))^2} \tag{5.5}$$

Where $\lambda(M_1, M_2)$ is an the vector containing the generalized eigenvalues of covariance matrices M_1 and M_2.

Then we take the mean value of all the trials as the final F-measure value, and report corresponding precision, recall and accuracy. The results are summarized in Table 5.5. Our algorithm outperforms the method [211], which works on gait cycles, with $K = 32$ to maximize the F-measure at 0.86.

We think that our assessment pipeline allows to evaluate the algorithm performance fairly, since the number of positive and negative examples is equal in the testing set. However, we also made an assessment experiment in the same way as [211]. The researchers ignored the class distribution, i.e., prevalence of the abnormal data to the normal one with a rate of 8/9 versus 1/9, and use the F-measure, precision, recall (sensitivity) and specificity assigned for the pathological class. Definitions for the F-measure, precision and recall may be found in Chapter 4. Specificity can be calculated as:

$$Specificity = \frac{TN}{(TN + FP)}. \tag{5.6}$$

It should be noted that a more standard procedure is to use a validation set to estimate the best number of nearest neighbors. Nevertheless, the methods we compare with [250, 211, 169], do not use a validation set. We assessed this possibility, but in this case the training and validation sets would contain only 5 persons, and the number of data items is simply too small to estimate the optimal k correctly. In practise, we tried to use 1 or 2 persons as a validation set to estimate an optimal K based on the best F-measure which resulted in low values. This phenomenon occurs when most the validation examples are assigned to a pathological class.

We thus took the best parameters obtained in the previous test: window size = 90 frames, K = 32.

Table 5.6 shows the results obtained along with the ones reported by Nguyen et al. [211] for the dataset [251]. We also compared the angle-based covariance matrices with standard joint based matrices as in [250] calculated by Equation 5.1 on 24 joints, and the covariance features proposed in [226] calculated by Equation 5.2. In the first case, the final feature is 24×24 and in the second 75×75. We use the same window size as for our angles based data, thus equaling 90 frames. The Optimal number of K is set to 2 and just 1 neighbor correspondingly

Table 5.6: Results on 4 test subjects from [251] with K-NN

method	F-measure	precision	recall	specificity
[211]	0.791	0.933	0.686	**0.679**
ours	**0.981**	**0.991**	**0.958**	0.654

for the joint-based covariance matrices. Results are presented in Table 5.5. Our simple feature outperforms the algorithm [211] and joint-based matrices. Resulting F-measure curves and covariance matrices are shown in Figure 5.5.

5.3.6/ Cross-datasets Analysis

In order to show that the proposed approach can be also applied to entire normal/abnormal sequences and is not limited to a given dataset, we arranged a simple test. The data from [204] was used to train a normal gait model based on flexion, rotation and abduction angles data. The distribution of each covariance feature follows a multivariate normal model. All the sequences from the DAI gait dataset were used as test data. In this test we used each gait sequence as a sample. We evaluated the normal negative loglikelihood for each test sequence and then compared the distribution of the values. The distribution is shown on Figure 5.6. It can be seen that the normal sequences can be easily separated from the abnormal ones based on the likelihood value. The test shows the generality and descriptiveness of the proposed gait correlation feature. It also confirms that accurate cycle segmentation is not necessary for this gait assessment task.

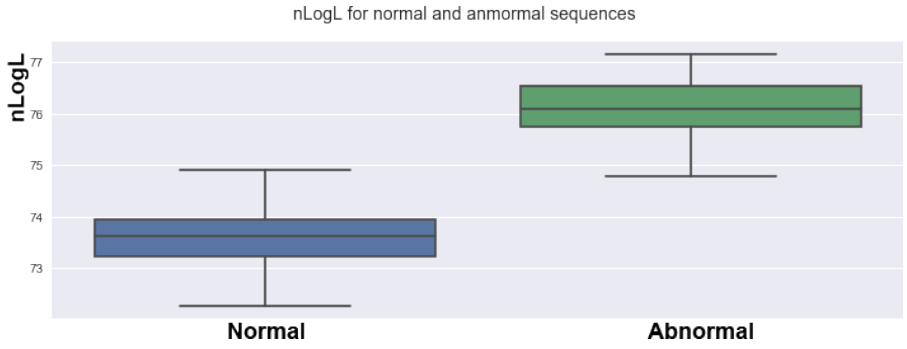


Figure 5.6: The $nlogL$ for normal/abnormal gait sequences from [169] by the $pdf(cov)$ build on the [204] dataset. The resulting distribution and confidence intervals show that the data are easily separable.

We also evaluated the possibility to combine the DAI and Sphere datasets together with our K-NN classifier. This time the data was pre-segmented in cycles. The segmentation was necessarily due to the fact that the sphere joints are extremely noisy, and our correlation matrix design is affected by that noise.

We added 64 normal cycles acquired in our lab to the test data. The resulting set contains 162 abnormal and 104 normal cycles. The Sphere dataset contains the gaits of patients with Parkinson disease and stroke. We aim to see if the proposed method can be used to detect this type of gait pathology and compare it with the algorithm in [250], where researchers use a Covariance-based approach to detect Hemiplegia and Parkinson gaits on a proprietary dataset. We used the DAI dataset [169] as a training dataset for our K-NN classifier and the Sphere dataset for the test. The test set contains an equal number of normal and abnormal cycles as in the previous experiment, and the same parameters for the K-NN are used. The results of this test are summarized in Table 5.7. The best parameters were obtained with K=12. We have lower accuracy than reported in [250]. This experiment shows the limits of our method, which is more suited to detect gait asymmetry and not able to detect 'static' abnormalities characterizing Parkinson or stroke patients.

Table 5.7: Generalization on data [169] [153]

F-measure	precision	recall
0.76	0.60	0.96

5.3.7/ COVARIANCE DESCRIPTOR TEST ON THE NEW DATABASE

Initially, we started to work with the available datasets to compare our results with the previous research. Later we collected our MMGS dataset which captures more sequences and we use it in our experiments. This section is dedicated to the group of experiments with the covariance gait feature and MMGS dataset.

5.3.7.1/ WHOLE SEQUENCES ANALYSIS

To start the work with the MMGS dataset, we evaluated the algorithm described in 5.3.1 using new data. With our new database, we seek to make a multi-class gait assessment, so we tested the algorithm on the newly acquired data using a new testing scheme. We do not compare the result with other binary gait assessment methods anymore, but aim to validate the proposed covariance feature in the new scenario. The whole sequence was used for the first test, without prior segmentation. Note that the number of gait cycles in a sequence in our dataset varies, and is between 1 and 3.

Our dataset is class balanced, so we decided to use the F1 mean score for three classes as the main characteristic to evaluate the data. We report the precision and recall as well. The other difference between this test and the earlier ones, is that we used the training and validation set to estimate the optimal number of k, which was not possible with the smaller and unbalanced datasets used earlier [251, 169, 153]. There are 22 persons in our dataset, so we use 11 persons for training, 5 persons to validate the optimal number of k in K-NN algorithm, and 6 persons for test. In each run of the algorithm, the training, validating and testing data are assigned in a random fashion. In order to report results independent on the data partitioning, we perform the test 25 times and evaluate the mean F1 score, precision and recall and their standard deviations. Finally, we test three combinations of flexion angles: X, Y, Z , X, Z, and Y, Z. The covariance matrices are calculated using Eq. 5.2. The results are shown in Table 5.8. They show that speed does not provide additional discriminative new information for our covariance-based gait feature. The main information is captured by the covariance matrix calculated from Z, Y (flexion and rotation) angles.

Similar to [250] we decided to check if speed information is beneficial for the targeted symmetry gait classification. We used the speed information from corresponding flexion angles data, by calculating the difference between two consequent frames. The final speed descriptor was calculated as a covariance matrix for four speed variations for hip and knee flexion angles. We then simply concatenate the speed matrices with angle data for the final feature flexion-speed matrix.

Although speed is an important characteristic of gait, and can be an important feature for some pathology detection, with our data it does not improve the classification. It may therefore, safely be dropped from the final aggregated feature vector.

Table 5.8: Results of the multi-class gait assessement

angles / mean (std)	precision knee	precision normal	precision padding	recall knee	recall normal	recall padding	mean F-score
Z, Y angles	0.86(0.11)	0.72(0.13)	0.77(0.12)	0.90(0.07)	0.72(0.08)	0.76(0.10)	0.78(0.05)
Z, X angles	0.65 (0.08)	0.45 (0.13)	0.52 (0.11)	0.64 (0.10)	0.46 (0.066)	0.53 (0.09)	0.54 (0.06)
X, Y, Z angles	0.86 (0.11)	0.70 (0.16)	0.78 (0.16)	0.91 (0.08)	0.73 (0.12)	0.75 (0.09)	0.78(0.06)
Z, Y angles + speed	0.82 (0.13)	0.71 (0.11)	0.76 (0.13)	0.89 (0.07)	0.71 (0.09)	0.7325 (0.09)	0.76(0.05)
Z, X angles + speed	0.84 (0.12)	0.70 (0.11)	0.73 (0.14)	0.88 (0.08)	0.67 (0.06)	0.74 (0.07)	0.75 (0.04)
Z, X, Y angles + speed	0.87(0.11)	0.73(0.14)	0.75 (0.15)	0.90 (0.08)	0.72 (0.10)	0.77 (0.08)	0.78(0.05)

Table 5.9: Binary gait assessment for gait cycles with a K-NN

Data	Precision (std)		Recall(std)		F-measure(std)		F1mean
	normal	abnormal	normal	abnormal	normal	abnormal	
our	0.850(0.095)	0.766(0.073)	0.647(0.070)	0.916(0.048)	0.730 (0.055)	0.830 (0.042)	0.7831 (0.047)
[169]	0.609(0.076)	0.856(0.084)	0.6124(0.090)	0.8661(0.073)	0.602(0.115)	0.860(0.028)	0.743(0.057)
[251] alg. [211]	1.00(0)	0.909(0.016)	0.778(0.243)	0.604(0.084)	0.841(0.188)	0.720(0.062)	0.812(0.125)

5.3.7.2/ CYCLES ANALYSIS

As shown by the previous experiments, our covariance based descriptor does not require the segmentation of a gait sequence in cycles. However, it is a common pre-processing step in skeleton-based gait assessment algorithms [211, 156, 153]. Therefore, in this section the results for pre-segmented cycles are reported. In addition, we later use pre-segmented cycles in a sequential gait model described in Section 5.4, so we use the same pre-processing to compare the data.

The number of cycles in the dataset after segmentation is equal to 548 cycles.

We perform the selection of the persons for training and testing data randomly and repeat the procedure N times to see the variance of results depending on the dataset distribution. The validation part of the set is used to estimate the optimal number of k in the K-NN algorithm. The optimal K is selected as the one maximizing the mean F1 score. In this test N = 200.

The DAI [169] dataset was used to test our approach, as the only normal/abnormal gait dataset containing orientation information. The data are segmented in cycles. This gives us the following partitioning:

- Testing part: 86 cycles from which 7 normal and 79 are pathological.
- Training part: 21 cycles, all normal.

We use the resulting dataset composed of gait cycles for out tests. It gives us in total 86 abnormal and 28 normal sequences.

Our custom dataset is used as a training and validation set. We use 16 persons for training and 6 to validate the optimal number of k in K-NN algorithm.

The results for our dataset and the DAI dataset are summarized in Table 5.9. We also provide the results reported for the algorithm of Nguyen et al. [211] on the dataset [251] for a comparison. We use the same evaluation scheme as in [211].

Our results are comparable with the results reported in [211].

Finally, we evaluated the possibility of our algorithm to perform a multi-class gait assessment. This is done in order to evaluate the possibility to use the proposed method, and compare the results with the later gait model introduced in the next section. Table 5.10 provides with the results. They demonstrate that our covariance-based features can be used in a more complex classification task with a similar F-measure value that we got for the binary gait classification.

5.3.8/ CONCLUSION ON THE KINEMATIC NCOVARIANCE FEATURES

Based on our experiments, we propose to use the following algorithm for gait kinematic covariance-based gait analysis for the case when orientation data from Kinect are available:

Table 5.10: Multi-class assessment for gait cycles with a k-NN

Data	Precision (std)			Recall(std)			F-measure(std)			F1mean
	knee	norm	limp	knee	norm	limp	knee	norm	limp	
our	0.86(0.11)	0.72(0.13)	0.79(0.13)	0.90(0.06)	0.73(0.09)	0.76 (0.09)	0.88(0.07)	0.71 (0.08)	0.76(0.07)	0.781 (0.05)

Data: Hip and joint orientations from Kinect v.2
Result: Covariance matrices

- Calculate Flexion angles in 3D
- Exclude the Abduction angles
- Calculate covariance matrices using the equation 5.1 to obtain an $8x8$ feature

Algorithm 1: Final Kinematic covariance features calculation

We recommend to use correlation distance and consensus rule as the parameters of the K-NN classifier.

Overall, the proposed covariance flexion gait features paired with a simple K-NN classifier show the performance close to the state-of-the-art methods. However, the temporal information is an important part of the information about gait. Therefore, in the next section we present a way to use a sequence model using kinematic features to evaluate gait.

5.4/ SEQUENCE GAIT MODEL

Lately, Neural Networks-based methods became ubiquitous in Computer Vision. They are so wide-spread, that the researchers who don't use it are often questioned why. In this section we select an appropriate ML-based model for our gait assessment task and provide the results obtained. The Sequence model is the model which works with sequence data i.e. the data in which the temporal component is important. When looking at the gait data, we see that many commonly used by clinicians parameters are temporal ones. In addition, gait can naturally be segmented in repetitive sub-parts called gait cycles. It is then interesting to use the temporal information for the gait assessment tool.

5.4.1/ LSTMS

Earlier in this thesis in Section 2.2.5.2 we presented gait models which can preserve temporal information in a gait sequence. Such models are popular because this technique is effective in describing the transition of human posture states during a gait cycle.

LSTM [23] recurrent neural networks (RNN) can be used to learn the features for gait recognition or classification and there were several attempts to use them with conventional video sequences [201], radar data [235], and data from depth sensors [223]. RNN-based approaches have achieved outstanding performance on action recognition [223, 246]. Zhu et al [223] use an LSTM network for significant co-occurrence feature minings from the skeleton joint positions for a task of action recognition. Zhang [246] select a set of simple geometric features to train a sequential model for action recognition. LSTMs were used in biometrics-related research. Feng et al [201] use an LSTM model for gait recognition. The features used are also learned from static frames with a convolutional neural network (CNN).

Only several recent methods adapted Recurrent Neural Networks for gait pathology detection. Liu et al. [208] proposed a Deep Rehabilitation Gait Learning (DRGL) for modeling the knee joints of

the lower-limb exoskeleton, which leverages a simple one-layer LSTM to learn the inherent spatial-temporal correlations of gait features. Researchers sought to predict the abnormal knee joint trajectories based on the other joints. The method results were evaluated visually, by comparing the predicted and actual trajectory. Feng et al. [201] trained an LSTM model on human joints data to characterize gait. An LSTM was used to model temporal parameters of the gait sequence. The hidden activation values of the Neural Network represented the final gait feature. Researchers used 2D data, however the method can be easily extended to 3D. Zhao et al. [254] adopted an LSTM net to diagnose Neurodegenerative diseases (ND). The model was trained and tested using temporal data that was recorded by force-sensitive resistors including time series, such as stride interval and swing interval. Researchers indicated the results allow to diagnose particular ND.

5.4.1.1/ LSTM Cell Structure

LSTM is a specific recursive neural network with a set of memory blocks, which are concatenated. Each block contains inputs, outputs, forget gate units, and one or more self-connected memory cells. There several slightly different LSTM architectures, so here we introduce the mostly used one. The block diagram of the LSTM memory block is shown in Figure 5.7.

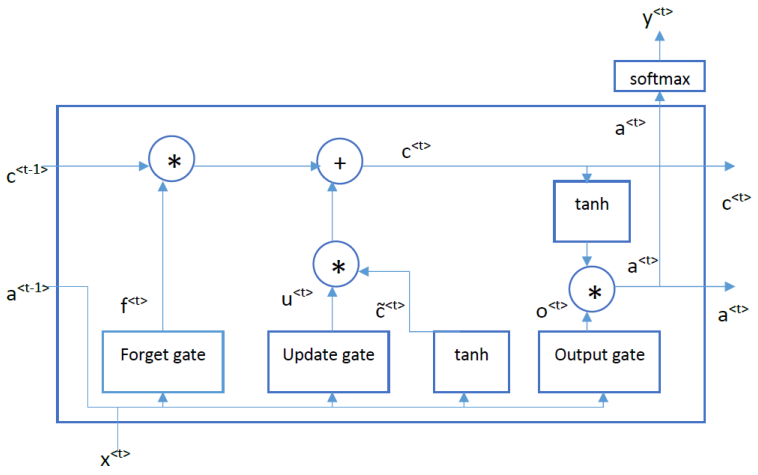


Figure 5.7: LSTM unit as introduced in [258].

An LSTM cell c and the candidate value for updating it, $\widetilde{c}^{<t>}$ are design in the following way. The three gates can perform update (Γ_u), forget (Γ_f) operations for the cells and give the output ((Γ_o)):

$$\begin{aligned}
\widetilde{c}^{<t>} &= tanh(W_c[a^{<t-1>,x^{<t>}}] + b_c) \\
\Gamma_u &= \sigma(W_u[a^{<t-1>}, x^{<t>}] + b_u) \\
\Gamma_f &= \sigma(W_f[a^{<t-1>}, x^{<t>}] + b_f) \\
\Gamma_o &= (W_o[a^{<t-1>}, x^{<t>}] + b_o) \\
c^{<t>} &= \Gamma_u * \widetilde{c}^{<t>} + \Gamma_f * c^{<t-1>} \\
a^{<t>} &= \Gamma_o * c^{<t>}
\end{aligned} \tag{5.7}$$

Where b - biases, a - activations, W - weights, $tanh$ - activation function.

In practise, LSTMs are considered to be more powerful and general than Gated Recurrent Unit (GRU) models. Because the state of the LSTM is regulated much more delicately than that of the basic RNN, the LSTM cell can learn very long term relations in the data. Multiple LSTM cells can be stacked for more expressive power.

5.4.2/ LSTM-BASED GAIT MODEL

In this section, we describe the proposed Compound LSTM network, exploring several architectural variations.

Single LSTM net: Gait data is a multidimensional time-domain sequence. In gait sequences, the joint angles of current time is related with the previous angles, future angles and also other joints angles. To capture these complex relations, we use a bi-directional LSTM gait model. Such model allows to increase the amount of input information available to the network i.e. future and past dependencies, at the cost of having more parameters that need to be estimated.

The optimal architecture and the choice for the hyperparameters of the network depend on the specific problem. We adopt a double-layer LSTM network. The proposed network is constructed by inheriting many insights from a recent Sequence Models course [258]. The Two-layer LSTM network was earlier adapted for gait analysis in [254].

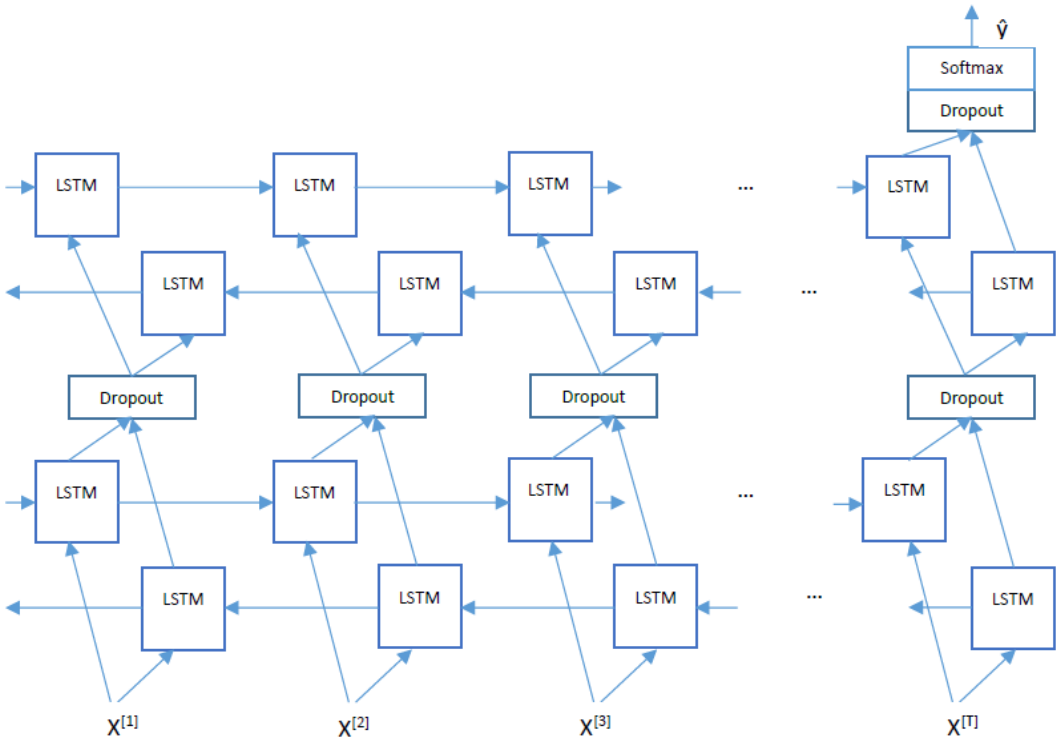


Figure 5.8: Single bi-directional LSTM model architecture with two layers combined with a dropout layer, and a softmax layer that gives the pathology predictions.

The main architecture used for a single LSTM model is shown in Figure 5.8. We use a bi-directional LSTM, which means that it captures the temporal relations from both previous and next states of the model. In Figure 5.8 this fact is shown by two rows of LSTM, in which one part have forward connections (left to right), and other backward connections (right to left). Such architecture can be called an acyclic graph. The final prediction that depends on both forward and backward activations.

After some prior testing, we set the number of LSTM hidden units to be 64. This results in a model with 140,163 parameters to train. With performed initial test to select the hyper-parameters.

The Adam optimizer [147] is used to control the learning rate, with the parameters: learning rate=0.0002, beta1=0.9, beta2=0.999, epsilon=1e-08, no decay. We also use the standard mini-batches technique for training, the mini-batch size being set to 21. No batch normalization is used.

Each of our architectures was trained end-to-end using a stopping criteria and an adaptive decreasing learning rate, based on the validation set performance. We were training to maximize the categorical entropy cross-validation loss on the validation data. Only 40 epochs (a single step in training a neural network) were used, since we don't have a very big dataset and input features are low-dimensional. The used loss function was categorical cross-entropy.

Compound LSTM model: A single LSTM model in general gives good results. However, we discovered that the performance varies depending on the training/validation data fold. In order to achieve better results, a compound model was designed, in which 5 separate LSMT models are trained on different training/validation partitions, and final label for the test set is assigned based on an elaborated handcrafted weighting scheme. The scheme weights the assigned probability for each model based on its validation accuracy in the following way. First each of the models is trained on different training/validation sets. Then each of the 5 models independently assigns probability to the test set. The final probability of a class is the average of 5 single LSTM weighted by their correspondent validation accuracy. The final label is then the one the most likely (maximizing the $p(x)$). The described model architecture is shown in Figure 5.9. The 5 models softmax weighted output example is shown in Figure 5.9.

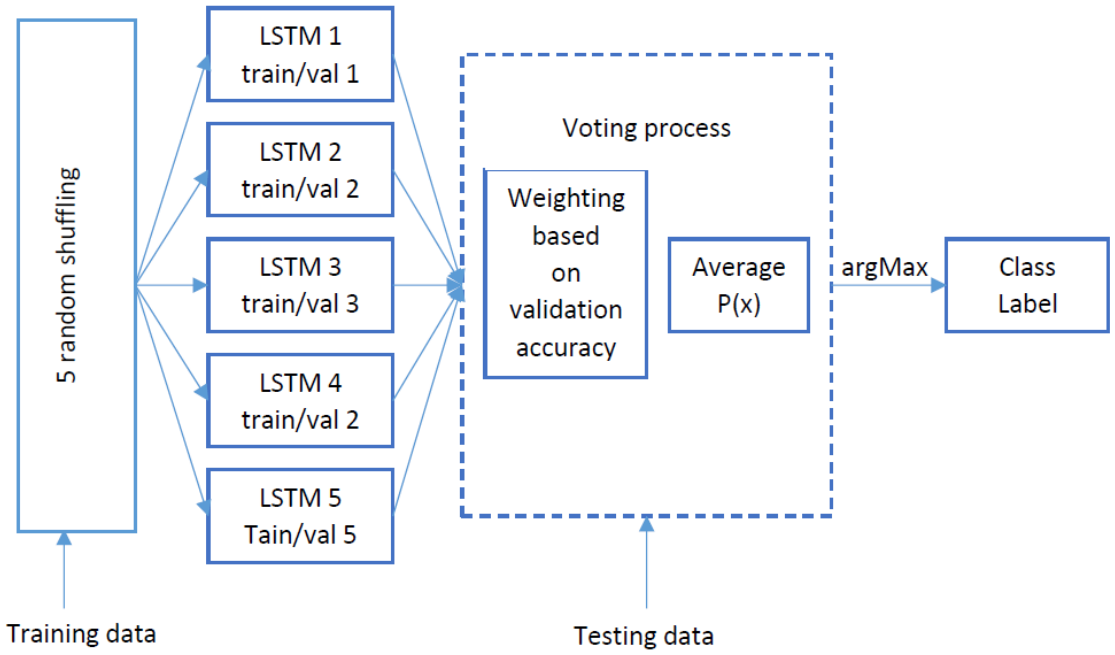


Figure 5.9: Proposed compound LSTM model architecture.

.

The above presented models were implemented in Python using the Keras Library.

5.4.3/ Experiment Protocol

This section describes the experiments we make to validate the sequence model proposed. In this section, we only use the MMGS dataset as it contains the biggest amount of training data.

5.4.3.1/ Preprocessing of the Data

We pre-processed the data in the following way. First we exclude all the joint measurements where the joint status returned by the Kinect SDK is $not\ tracked$ or $inferred$. This allows us to remove the data, where ML algorithms failed to estimate the joint's position. We apply a low-pass filter to the quaternions data. Filtered quaternions are then used to calculated low-limb kinematic parameters as demonstrated in Section 3.2.

Then we segment the gait sequences into cycles. Such segmentation allows us to easier compare gait features. The segmentation done is similar to the one made in [211], but in this work it is based on the right hip Z flexion angle peaks. After the segmentation, we obtain in total 548 gait cycles. We sub-sample each cycle to have the same number of frames. For our tests we used a fixed cycle size equal 30 frames and linear interpolation technique. In the following experiments, we use the Rotation and Flexion gait angles as features for our model, as this combination has shown the best results in terms of accuracy in comparison with the use of covariance matrices from angles and skeleton joints directly.

5.4.3.2/ Data Division

In order to train and evaluate the model, we need to divide the dataset into training, validation and testing data. This is done in the following way. We use 11 persons for training and 5 persons for validation, which leaves 6 persons for testing.

Previously, the same scheme was used with $K-NN$ in section 5.3.7.2 so we can compare the performance of the two classification algorithms. This division scheme guarantees the generalization of the algorithm for the data of new, unseen subjects. We perform the selection of the persons for training and testing data randomly and repeat the procedure several times, each time training a new LSTM model, in order to show it's independence of the particular data partitioning. For the final result, such random partitioning was performed 20 times. The exact details can be found in Table 5.14.

5.4.4/ Results & Discussions

Our results can be found in Table 5.11. A single LSTM model actually gives good results. However, we have noticed that depending on the persons distribution in training/validation data, we have possible deviations of about 3%. A compound model architecture was proposed in order to decrease the variance of each model and its dependency on the test partitioning. Training/validation partitioning affects the results even more, so we generalize by making 23 trials for the compound model, which means each single LSTM was trained 115 times.

Table 5.11 provides the mean results values. Detailed result examples and corresponding data partitioning are in 5.14. We handle the problem of variance by using a compound model in this task. However, we also have a bias between training and validation results, so more training data will be beneficial for the model.

The confusion matrix is shown in Figure 5.10 b.

We performed exhaustive testing to check if the resulting score of the algorithm does not depend on the random fold. However, to make it easier to compare with our results in the Future, table 5.12

metric	Single LSTM			Compound Model		
	mean	min	max	mean	min	max
test accuracy (%)	**76**	68	94	**80**	75	84

Table 5.11: Single and Compound LSTM accuracy on 115 (5x23) and 23 trials.

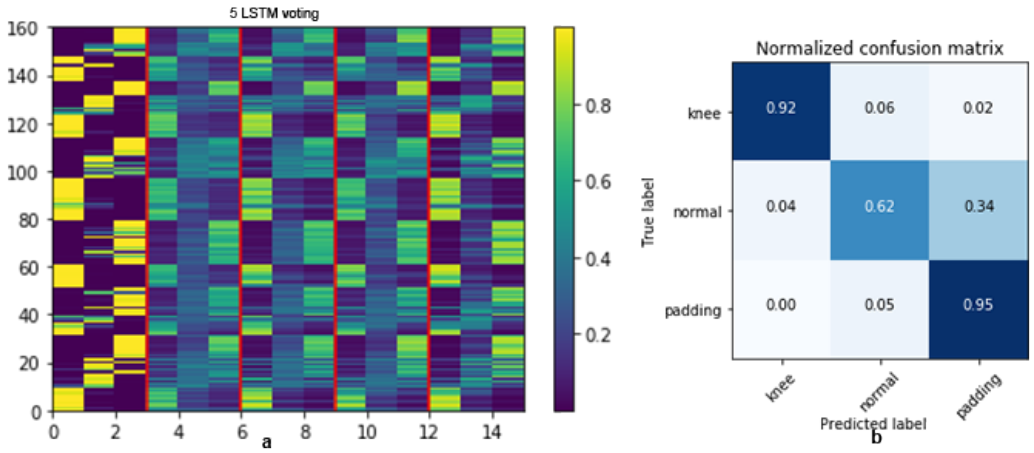


Figure 5.10: a) Five LSTM models Softmax outputs: each LSTM output is a 3x1 probability vector, color codes the probability value. b) Final normalized confusion matrix for 3 classes. Two pathological gaits are detected by the system with a high accuracy.

provides the single compound model data, with the training/validating/testing subjects specified. The confusion matrix can be found in Figure 5.11.

Finally, we compared the performance of the skeleton joints and the covariance matrices presented earlier and calculated using the algorithm 1. The fixed partitioning used in the same specified in Table 5.12. The results are summarized in 5.13. Intially, we planned to use the covariance matrices as features for our Gait Model, however, the direct use of the flexion angles turned to be more representive. Both new features, however, outperform the direct use of Kinect skeleton joints' positions.

From these results, we can conclude the following. Overall, the proposed system is comparable to the accuracy reported by other studies in the domain [211, 250, 153]. Moreover, we use more data than [153, 250] and assign three classes and not a binary pathology detection as [211]. Hip and flexion angles work extremely well for recognizing the knee-bending problem. The gait pathology simulated by padding is harder to separate from the normal walk data. We see a solution in a search for a reliable way to estimate ankle kinematics from the point cloud data directly.

The system accuracy is not high enough for clinical studies, but the results are interesting. Although the researchers are used to dealing with confusion matrices, the clinicians usually look for parameters to evaluate the performance. Table 5.15 reports the sensitivity and specificity of the system for two pathological classes. To make it easier for future comparisons, we report the results for the earlier model. The performance and data partitioning for the model were presented in Table 5.12. Please note that our mean data for 23 trials from Figure 5.10 shows similar trends, even more pronounced in case of two pathological classes.

There are a number of factors to why we don't use the joints or quaternions data directly in order to learn the features using our model. The main reason is the fact that we were looking for particular features, which correspond to the ones traditionally measured in clinical gait analysis. Secondly, having more hand-engineered components generally allows a RNN system to learn with less data. Our dataset is small, hence, it is better to use really meaningful features. Our algorithm's knowledge is then supported by human insight. In the future, with more data, automatic feature

Table 5.12: Accuracy on the particular data partitioning

persons	acc, %	M1	M2	M3	M4	M5	Compound model
Train + val: 2,3,4,5,6,7,8,10,11,12, 13,14,16,17,19,22	val	85	82	81	84	79	n/a
Test: 1,9,15,18,20,21	test	83	0.76	0.74	0.82	0.76	82

Table 5.13: Comparison with skeleton and covariance features on the particular data partitioning

features	final model accuracy, %
24 skeleton joints from Kinect v.2	48
Kinematic covariance matrices	75
Kinematic angles	82

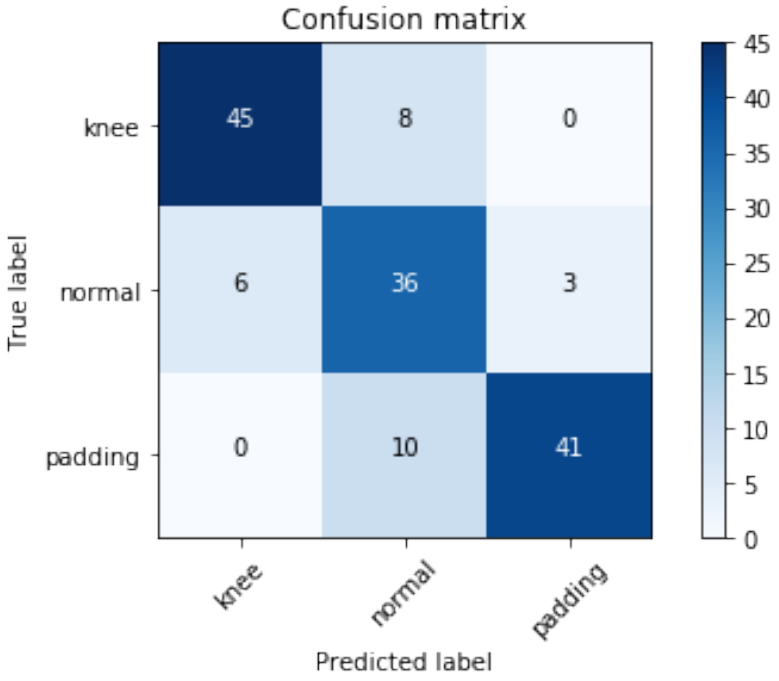


Figure 5.11: Non-normalized confusion matrix for a particular data partitioning.

learning from the initial skeleton data can be performed.

5.5/ CONCLUSION

It has been shown that the Kinect v.2 sensor has potential for abnormal gait detection. However, the number of publicly available datasets is very small which makes it difficult to generalize the results and compare the existing algorithms.

In this chapter we propose a gait assessment database for the use in clinical gait assessment research. Our proposed database captures more individuals than the existing ones, and provides data of different modalities.

Then we proposed to use kinematic gait parameters for two scenarios: abnormal gait detection, and particular pathology classification.

We used kinematic parameters of the low-limbs joints in two experiment setups. First, we used a covariance descriptor for binary gait assessment paired with a simple K-NN classifier. Second, we proposed a more complex ML-based model based on kinematic parameters calculated for gait cycles.

With the datasets used, we obtained state of the arts results. A direct comparison of results was not always possible because no dataset is publicly provided by the authors. The other difficulties are the different frameworks used by authors. We tried to adopt some in order to make the comparison of the results easier.

The results obtained show the potential of the proposed kinematic features for the assessment of the pathological gait. We obtained the best results with the detection of limping/knee rigidity pathology using a compound LSTM model. The machine learning based method proved to be better than a simple K-NN classifier.

The result of this work is a complete gait assessment platform: we propose an acquisition system design, an algorithm which tackles the problem of low-limb kinematic symmetry description, and normal/abnormal gait modeling via a LSTM Network. Experimental results demonstrate the effectiveness of the proposed approach in the context of abnormal gait detection and classification.

Table 5.14: Examples of the concrete single LSTMs and compound model accuracy on particular data partitioning.

train\test	validation	train acc, %	val acc,%	test acc,%	compound model test acc, %
1,2,3,6,7,8,9,10,	1,6,10,17,18	55	96	90	
11,13,14,15,17,	1,8,10,11,13	80	88	77	
18,19,21	1,3,6,10,11	83	84	77	80
\	9,13,17,19,21	57	80	72	
4,5,12,16,20,22	2,8,14,17,18	87	75	76	
1,2,4,5,8,10,11,	1,4,17, 18,21	88	72	86	
13,14,15,16,17,	10,11,16,18,20	87	78	82	
18,19,20,21	13,14,15,16,21	84	70	85	84
\	10,11,15,17,19	77	83	82	
3,6,7,9,12,22	1. 5. 15. 16. 19.	84	84	75	
2,3,5,6, 8,9,10,11,	5,6, 9,11,21	83	72	80	
12,13,16,17,	8,9,10,13,16	80	92	78	
18,20,21,22	2,5,11,18,21	78	78	75	**77**
\	6,12,13,20,22	85	72	73	
1,4,7,14,15,19	3,5,6,8,20	86	81	75	

Table 5.15: Accuracy on a particular data partitioning

category	TP	FP	TN	FN	sensitivity, %	specificity, %
knee (bending)	45	8	6	90	88	92
limping (padding)	41	10	3	95	93	90

6

Conclusion

Gait affects the quality of life of an individual, and deviations from a normal gait pattern can indicate a particular disease. It is therefore important to detect the deviation, preferably in the early stage of its development. At the same time, the gait of each person is unique, and is therefore often used for person recognition scenarios. An ideal gait assessment method has to deal with these two constraints. We strived to develop a clinical gait analysis system, which will not carry a significant financial burden for hospitals and practitioners. There are golden standard gait evaluation devices such as VICON and different GF platforms on the market already. However, their price reaches thousands of dollars per device. Consequently, only big hospitals or specialized centers can afford such a system. In addition, such devices require a long setup time for each patient, which makes gait analysis a time-consuming procedure.

Therefore, we choose to work with a more accessible device delivering human skeleton joint data - the Microsoft Kinect sensor. The acquisition setup selected in this work is also very simple and doesn't require any additional equipment, although we consider the possibility to deploy a common treadmill.

Our work was performed in collaboration with Dijon-based company Proteor, which creates prosthesis and medical corsets for patients. The company seeks a method allowing to perform gait acquisition and objective gait analysis on its premises.

The two main contributions of this dissertation are the two algorithms presented in Chapters 4 and 5. First is 3D human shape descriptor, and the second is skeleton-based gait classification model.

One works with 3D data and captures the point cloud arrangement to describe human posture. The other one is skeleton data-based and assign a label to the patients' gait. Based on our experiments, we conclude that our 3D posture descriptor can be used to classify static human postures from point cloud data. It is simple, fast to compute and generalizes, so it can be used in different application domains.

The descriptor for gait classification shows comparable or better than state-of-the-art results on several available datasets. We paired our descriptor with a simple k-NN classifier, and with a more sophisticated LSTM-based model. The deep-learning method slightly outperforms the k-NN classifier-based method, and can probably be improved once more training and validation data become available.

The 3D posture algorithm can be used directly on point cloud data when skeleton information provided by the depth camera is not reliable. On the contrary, our gait assessment model is based on skeleton data and shows good results when the patient is optimally placed in front of the sensor and all joints are tracked.

Based on the experiments described in Chapter 3, we conclude that a setup with a Kinect v.2 or an RGB-D camera with similar characteristics combined with a treadmill has potential for clinical gait assessment. A treadmill can slightly affect the gait pattern, but will overcome a range problem when deploying common RGB-D sensors. Needless to say, there are more expensive and accu-

rate cameras on the market, but their usage will increase the final price of a system by an order of magnitude.

Relying on the research work performed, we are confident that the Microsoft Kinect v.2 can be exploited for gait assessment. Low-limb kinematics calculated from the Kinect data can be used to classify gait deviations affecting gait symmetry: limping, rigidity and others. 3D information from the Kinect camera can be used to classify human postures. However, further work is required in order to increase the accuracy of the system.

6.1/ ADVANCEMENT OF THE THESIS

The main contributions of this work are the following. We realized a detailed literature review on the subject. The research problem was approached from two point of view. Firstly as a gait assessment exploiting clinically-used gait parameters. Secondly as a human motion description from 3D data. We proposed a study of the state-of-start in motion description from 3D data. We summarized different studies related to gait assessment and discerned the most used gait characteristics.

A significant part of our work was dedicated to the selection of the optimal gait acquisition setup. We experimented with single and multi-kinect setups for the use of point cloud and skeleton-based algorithms. We tested different hardware solutions and selected the most appropriate for a clinical gait analysis tool. We evaluated the possibility to make different acquisition setups using the Kinect v.2 sensor in order to obtain maximum performance and accuracy. We also performed several data acquisitions in our laboratory, on the Proteor premises and in the CHU hospital of Dijon. Part of this data were used in the experiments with the algorithms proposed in chapters 4 and 5.4.

We experimented both with 3D data and skeleton data before selecting the final gait assessment algorithm design and features. First, we worked with point cloud data directly. A simple 3D posture descriptor was proposed, and the possibility to assess motion flow from 3D data was reviewed. Our 3D posture descriptor captures the configuration of the human body by using a specially designed and adapted grid. Several tests using data from the MSR Action 3D dataset, were performed to validate the descriptor. Experimental results show that the proposed descriptor can capture a static point cloud structure and distinguish different human postures without the use of a big amount of data or deep learning-based techniques.

Our next contribution is dedicated to the use of skeleton data in gait research. Pathological gait multi-modal dataset was acquired in our laboratory using a Kinect v.2 device. This dataset simulates asymmetric gait and has the biggest number of subjects in comparison to other available normal/abnormal gait datasets.

We adopted a method to estimate kinematic parameters from orientations data, which to our knowledge was not used in gait Kinect-based gait research previously. We assessed the reliability of the low-limbs gait kinematics given by the Kinect v.2 sensor comparing the data acquired with a golden-standard used in medical analysis MOCAP and Vicon. For this task, a custom dataset was collected in the CHU Dijon.

Several descriptors based on low-limbs kinematics were developed and tested on the available gait data. We addressed the problem of binary gait classification and multi-label classification. Finally, an LSTM-based gait model capable of detecting limping and rigidity was proposed.

6.2/ LIMITATIONS

This thesis proposes several findings in gait analysis based on 3D data. The main difficulty for this thesis was to propose gait assessment methods which would be able to detect abnormal patterns

and deal with the variance in normal gait patterns. The other difficulty is the absence of test data. Currently, machine learning based methods are widely used in all computer vision applications. In this work, we also adopted such a learning framework, which has shown quite good results, but even the new dataset we acquired does not give the possibility to train very deep architectures. Similar, the EU law restrictions did not allow us to acquire a database of pathological gait from real patients, although we spent a significant amount of time designing an acquisition platform to be used in clinics. This part is presented in Chapter 3. Unfortunately, the planned acquisition site was not able to receive a special registration, allowing us to gather data from real patients.

This limitation made us to direct the work for this thesis into the new direction. Initially, we wanted to use the Motion Flow to describe human movements, and try to learn abnormal pattern from examples. However, in the final work we searched for hand-crafted features for gait evaluation, which require less training and testing data.

6.3/ PERSPECTIVES

This thesis was a first exploratory step in the gait analysis project realized in the Le2i laboratory. Previously, the main focus of the research of the project team was 3D information from static point clouds. The initial plan was to build this research on the previous findings. However, gait and dynamic depth data turned out to be quite different and required a different approach. This thesis explored several possibilities to assess patients' gait using a Kinect v.2 based acquisition system.

The features we choose correspond to the data used by clinicians. Therefore, a priori, the targeted precision should be comparable to human evaluation. At the moment, the results are less accurate, probably due to erroneous skeleton estimations. The Kinect sensor cannot accurately capture subtle joint movements, so further work on human skeleton extraction is needed. There is a great need for a reliable open source skeleton joints estimation method in the human movement analysis domain. Recently, researchers mostly concentrated on posture recognition using 2D data [252]. We believe that the combination of 3D data and machine learning can help to further improve results further.

At present, Microsoft has decided to discontinue the Kinect sensor production. Although inconvenient, it does not affect the results of our work. 3D sensors are a big part of the sensors market, with new and better solutions proposed regularly. However, with the discontinuation of the popular Microsoft Kinect sensors, another consumer grade camera delivering depth data, will need to be adapted. The possible choice for today is the ZED camera, which uses a passive sensor stereo technology to triangulate 3D points from 2 views. The distance range of this camera might also solve the problem of the small acquisition distance range supported by the Kinect sensor. There are probably new improved depth sensors yet to come based on the latest research results [234].

In the last years, 3D shape completion methods from point clouds have shown good results [237] and even 2D images [196]. Adaptation of such a technique can lead to very interesting results in gait assessment studies. Ideally, a patients model can be acquired prior to the walking test. At a later stage, this will make it be easier to estimate the posture more accurately in a temporal sequence and evaluate the motion of different segments. The possible model to be used is the SCAPE model [41] based on which many newer models were built. The model can be trained on a huge human shape dataset, such as the Civilian American European Surface Anthropometric Resource (CAESAR).

Regarding the work realized in this Ph.D thesis, the direct continuation is the installation of the sensor in the motion assessment laboratory to work with real patients data. Once the system is deployed, we can gather more patients data to improve the proposed assessment algorithm and perform further research. If a treadmill is used, there are several straight-forward improvements to our algorithm, such as averaging cycles in order to compensate for erroneous skeleton estimations.

We selected relevant gait parameters, given our main goal to assess the gait of prosthetic patients.

For a particular disease, other features may be used, or even learned from patients data if a sufficiently large dataset will be captured. It is essential to identify and better understand abnormal gait patterns linked to disorders and to treat them.

Future research should put significant energy in ascertaining higher definition body shape models from 3D data that enable to track subtle changes in skeleton configuration and muscular tissue. To illustrate the gap, the special effect movie industry already attains the sought resolution be it via synthetic means.

BIBLIOGRAPHY

[1] **Camera calibration with opencv**. https://docs.opencv.org/2.4/doc/tutorials/calib3d/camera_calibration/camera_calibration.html. Accessed: 2017-23-06.

[2] **Definition of gait**. http://www.medicinenet.com/script/main/art.asp?articlekey=3533. Accessed: 01.12.2015.

[3] **Gait. research on different biometric modalities.** http://biometrics.derawi.com/?page_id=38. Accessed: 01.12.2015.

[4] **Kinect for windows sdk 2.0**. https://www.microsoft.com. Accessed: 2017-23-06.

[5] **Kinect for windows sdk programming - kinect for windows v2 sensor supported version**. http://www.shuwasystem.co.jp/products/7980html/4395.htmls, note = Accessed: 2017-04-06, organization =Shuwa System Co.,Ltd.

[6] **Openkinect**. https://openkinect.org. Accessed: 2017-05-05.

[7] **Optotrac optical measurement**. https://www.ndigital.com/msci/products/. Accessed: 2017-09-05.

[8] **Proteor**. http://www.proteor.com. Accessed: 2018-09-01.

[9] **Radial distortion correction**. http://www.uni-koeln.de. Accessed: 2017-01-07.

[10] **Rare Ltd home page, an xbox game studio**. http://www.rare.co.uk/. Accessed: 2015-10-30.

[11] **Stereo calibration using c++ and opencv**. http://sourishghosh.com/2016/stereo-calibration-cpp-opencv/. Accessed: 2017-23-02.

[12] **Understanding quaternions**. http://www.chrobotics.com/library/understanding-quaternions. Accessed: 2018-04-09.

[13] BLAND, J. M., ALTMAN, D. G., AND OTHERS. **Statistical methods for assessing agreement between two methods of clinical measurement**. *lancet 1*, 8476 (1986), 307–310.

[14] TINETTI, M. E. **Performance-oriented assessment of mobility problems in elderly patients**. *Journal of the American Geriatrics Society 34*, 2 (1986), 119–126.

[15] TSAI, R. Y. **An efficeint and accurate camera calibration technique for 3d machine vision**. *CVPR, 1986* (1986).

[16] RABINER, L. R. **A tutorial on hidden markov models and selected applications in speech recognition**. *Proceedings of the IEEE 77*, 2 (1989), 257–286.

[17] FRIED, A., CWIKEL, J., RING, H., AND GALINSKY, D. **Elgam—extra-laboratory gait assessment method: Identification of risk factors for falls among the elderly at home**. *International disability studies 12*, 4 (1990), 161–164.

[18] WOLFSON, L., WHIPPLE, R., AMERMAN, P., AND TOBIN, J. N. **Gait assessment in the elderly: a gait abnormality rating scale and its relation to falls**. *Journal of gerontology 45*, 1 (1990), M12–M19.

[19] PODSIADLO, D., AND RICHARDSON, S. **The timed "up & go": a test of basic functional mobility for frail elderly persons**. *Journal of the American geriatrics Society 39*, 2 (1991), 142–148.

[20] BESL, P. J., MCKAY, N. D., AND OTHERS. **A method for registration of 3-d shapes**. *IEEE Transactions on pattern analysis and machine intelligence 14*, 2 (1992), 239–256.

[21] JAEGERS, S. M., ARENDZEN, J. H., AND DE JONGH, H. J. **Prosthetic gait of unilateral transfemoral amputees: a kinematic study**. *Archives of physical medicine and rehabilitation 76*, 8 (1995), 736–743.

[22] MCGRAW, K. O., AND WONG, S. P. **Forming inferences about some intraclass correlation coefficients.** *Psychological methods 1*, 1 (1996), 30.

[23] HOCHREITER, S., AND SCHMIDHUBER, J. **Long short-term memory**. *Neural computation 9*, 8 (1997), 1735–1780.

[24] SCHÖLKOPF, B., SMOLA, A., AND MÜLLER, K.-R. **Nonlinear component analysis as a kernel eigenvalue problem**. *Neural computation 10*, 5 (1998), 1299–1319.

[25] ANKERST, M., KASTENMÜLLER, G., KRIEGEL, H.-P., AND SEIDL, T. **3d shape histograms for similarity search and classification in spatial databases**. In *Advances in Spatial Databases* (1999), Springer, pp. 207–226.

[26] JOHNSON, A. E., AND HEBERT, M. **Using spin images for efficient object recognition in cluttered 3d scenes**. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 5 (1999), 433–449.

[27] HE, Q., AND DEBRUNNER, C. **Individual recognition from periodic activity using hidden markov models**. In *Human Motion, 2000. Proceedings. Workshop on* (2000), IEEE, pp. 47–52.

[28] MATSAS, A., TAYLOR, N., AND MCBURNEY, H. **Knee joint kinematics from familiarised treadmill walking can be generalised to overground walking in young unimpaired subjects**. *Gait & posture 11*, 1 (2000), 46–53.

[29] ZHANG, Z. **A flexible new technique for camera calibration**. *IEEE Transactions on pattern analysis and machine intelligence 22*, 11 (2000), 1330–1334.

[30] BOBICK, A. F., AND DAVIS, J. W. **The recognition of human movement using temporal templates**. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*, 3 (2001), 257–267.

[31] DINGWELL, J., CUSUMANO, J., CAVANAGH, P., AND STERNAD, D. **Local dynamic stability versus kinematic variability of continuous overground and treadmill walking**. *Journal of biomechanical engineering 123*, 1 (2001), 27–32.

[32] MARTIN, D., FOWLKES, C., TAL, D., AND MALIK, J. **A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics**. In *Proc. 8th Int'l Conf. Computer Vision* (July 2001), vol. 2, pp. 416–423.

[33] COMANICIU, D., AND MEER, P. **Mean shift: A robust approach toward feature space analysis**. *IEEE Transactions on pattern analysis and machine intelligence 24*, 5 (2002), 603–619.

[34] KALE, A., RAJAGOPALAN, A., CUNTOOR, N., AND KRUGER, V. **Gait-based recognition of humans using continuous hmms**. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on* (2002), IEEE, pp. 336–341.

[35] HARTLEY, R., AND ZISSERMAN, A. **Multiple view geometry in computer vision**. Cambridge university press, 2003.

[36] HOBART, J., RIAZI, A., LAMPING, D., FITZPATRICK, R., AND THOMPSON, A. **Measuring the impact of ms on walking ability the 12-item ms walking scale (msws-12)**. *Neurology 60*, 1 (2003), 31–36.

[37] NOVOTNI, M., AND KLEIN, R. **3d zernike descriptors for content based shape retrieval**. In *Proceedings of the eighth ACM symposium on Solid modeling and applications* (2003), ACM, pp. 216–225.

[38] WANG, L., TAN, T., NING, H., AND HU, W. **Silhouette analysis-based gait recognition for human identification**. *IEEE transactions on pattern analysis and machine intelligence 25*, 12 (2003), 1505–1518.

[39] WHITNEY, S., WRISLEY, D., AND FURMAN, J. **Concurrent validity of the berg balance scale and the dynamic gait index in people with vestibular dysfunction**. *Physiotherapy Research International 8*, 4 (2003), 178–186.

[40] ZIVKOVIC, Z. **Improved adaptive gaussian mixture model for background subtraction**. In *Proceedings of the 17th International Conference on Pattern Recognition* (2004), vol. 2, IEEE, pp. 28–31.

[41] ANGUELOV, D., SRINIVASAN, P., KOLLER, D., THRUN, S., RODGERS, J., AND DAVIS, J. **Scape: shape completion and animation of people**. In *ACM transactions on graphics (TOG)* (2005), vol. 24, ACM, pp. 408–416.

[42] GELFAND, N., MITRA, N. J., GUIBAS, L. J., AND POTTMANN, H. **Robust global registration**. In *Symposium on geometry processing* (2005), vol. 2, p. 5.

[43] AGARWAL, A., AND TRIGGS, B. **Recovering 3d human pose from monocular images**. *IEEE transactions on pattern analysis and machine intelligence 28*, 1 (2006), 44–58.

[44] BARAK, Y., WAGENAAR, R. C., AND HOLT, K. G. **Gait characteristics of elderly people with a history of falls: a dynamic approach**. *Physical therapy 86*, 11 (2006), 1501–1510.

[45] MAN, J., AND BHANU, B. **Individual recognition using gait energy image**. *IEEE transactions on pattern analysis and machine intelligence 28*, 2 (2006), 316–322.

[46] STROBL, K., SEPP, W., FUCHS, S., PAREDES, C., AND ARBTER, K. **Camera calibration toolbox for matlab**. *Pasadena, CA* (2006).

[47] HILLMAN, S. J., HAZLEWOOD, M. E., SCHWARTZ, M. H., VAN DER LINDEN, M. L., AND ROBB, J. E. **Correlation of the edinburgh gait score with the gillette gait index, the gillette functional assessment questionnaire, and dimensionless speed**. *Journal of Pediatric Orthopaedics 27*, 1 (2007), 7–11.

[48] JONSDOTTIR, J., AND CATTANEO, D. **Reliability and validity of the dynamic gait index in persons with chronic stroke**. *Archives of physical medicine and rehabilitation 88*, 11 (2007), 1410–1415.

[49] LEARDINI, A., SAWACHA, Z., PAOLINI, G., INGROSSO, S., NATIVO, R., AND BENEDETTI, M. G. **A new anatomically based protocol for gait analysis in children**. *Gait & posture 26*, 4 (2007), 560–571.

[50] MITRA, S., AND ACHARYA, T. **Gesture recognition: A survey**. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 37*, 3 (2007), 311–324.

[51] RILEY, P. O., PAOLINI, G., DELLA CROCE, U., PAYLO, K. W., AND KERRIGAN, D. C. **A kinematic and kinetic comparison of overground and treadmill walking in healthy subjects**. *Gait & posture 26*, 1 (2007), 17–24.

[52] SCOVANNER, P., ALI, S., AND SHAH, M. **A 3-dimensional sift descriptor and its application to action recognition**. In *ACM Proceedings of the 15th international conference on Multimedia* (2007), pp. 357–360.

[53] WREN, T. A., DO, K. P., HARA, R., DOREY, F. J., KAY, R. M., AND OTSUKA, N. Y. **Gillette gait index as a gait analysis summary measure: comparison with qualitative visual assessments of overall gait**. *Journal of Pediatric Orthopaedics 27*, 7 (2007), 765–768.

[54] BEDER, C., AND KOCH, R. **Calibration of focal length and 3d pose based on the reflectance and depth image of a planar object**. *International Journal of Intelligent Systems Technologies and Applications 5*, 3-4 (2008), 285–294.

[55] BRADSKI, G., AND KAEHLER, A. **Learning OpenCV: Computer vision with the OpenCV library**. " O'Reilly Media, Inc.", 2008.

[56] KLASER, A., MARSZAŁEK, M., AND SCHMID, C. **A spatio-temporal descriptor based on 3d-gradients**. In *British Machine Vision Conference-BMVC* (2008), pp. 275–1.

[57] SCHWARTZ, M. H., AND ROZUMALSKI, A. **The gait deviation index: a new comprehensive index of gait pathology**. *Gait & posture 28*, 3 (2008), 351–357.

[58] BENGIO, Y., LOURADOUR, J., COLLOBERT, R., AND WESTON, J. **Curriculum learning**. In *Proceedings of the 26th annual international conference on machine learning* (2009), ACM, pp. 41–48.

[59] BO, L., AND SMINCHISESCU, C. **Efficient match kernel between sets of features for visual recognition**. In *Advances in neural information processing systems* (2009), pp. 135–143.

[60] GRAVES, A., AND SCHMIDHUBER, J. **Offline handwriting recognition with multidimensional recurrent neural networks**. In *Advances in neural information processing systems* (2009), pp. 545–552.

[61] RUSU, R. B. **Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments**. PhD thesis, Computer Science department, Technische Universitaet Muenchen, Germany, October 2009.

[62] SORKINE, O. **Least-squares rigid motion using svd**. *Technical notes 120*, 3 (2009), 52.

[63] TOSRANON, P., SANPANICH, A., BUNLUECHOKCHAI, C., AND PINTAVIROOJ, C. **Gaussian curvature-based geometric invariance**. In *IEEE. 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology* (2009), vol. 2, pp. 1124–1127.

[64] WANG, F., STONE, E., DAI, W., SKUBIC, M., AND KELLER, J. **Gait analysis and validation using voxel data**. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE* (2009), IEEE, pp. 6127–6130.

[65] ZHONG, Y. **Intrinsic shape signatures: A shape descriptor for 3d object recognition**. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on* (2009), IEEE, pp. 689–696.

[66] BLANCO, J.-L. **A tutorial on se (3) transformation parameterizations and on-manifold optimization**. *University of Malaga, Tech. Rep 3* (2010).

[67] GANAPATHI, V., PLAGEMANN, C., KOLLER, D., AND THRUN, S. **Real time motion capture using a single time-of-flight camera**. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2010), pp. 755–762.

[68] GU, J., DING, X., WANG, S., AND WU, Y. **Action and gait recognition from recovered 3-d human joints**. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 40*, 4 (2010), 1021–1033.

[69] HESS, R. J., BRACH, J. S., PIVA, S. R., AND VANSWEARINGEN, J. M. **Walking skill can be assessed in older adults: validity of the figure-of-8 walk test**. *Physical therapy 90*, 1 (2010), 89–99.

[70] LI, W., ZHANG, Z., AND LIU, Z. **Action recognition based on a bag of 3d points**. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2010), pp. 9–14.

[71] MOLLOY, M., MCDOWELL, B., KERR, C., AND COSGROVE, A. **Further evidence of validity of the gait deviation index**. *Gait & posture 31*, 4 (2010), 479–482.

[72] PATTERSON, K. K., GAGE, W. H., BROOKS, D., BLACK, S. E., AND MCILROY, W. E. **Evaluation of gait symmetry after stroke: a comparison of current methods and recommendations for standardization**. *Gait & posture 31*, 2 (2010), 241–246.

[73] PERRONNIN, F., SÁNCHEZ, J., AND MENSINK, T. **Improving the fisher kernel for large-scale image classification**. *Computer Vision–ECCV 2010* (2010), 143–156.

[74] POPPE, R. **A survey on vision-based human action recognition**. *Image and vision computing 28*, 6 (2010), 976–990.

[75] STEDER, B., RUSU, R. B., KONOLIGE, K., AND BURGARD, W. **Narf: 3d range image features for object recognition**. In *Workshop on Defining and Solving Realistic Perception Problems in Personal Robotics at the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (2010), vol. 44.

[76] SURYANARAYAN, P., SUBRAMANIAN, A., AND MANDALAPU, D. **Dynamic hand pose recognition using depth data**. In *20th IEEE International Conference on Pattern Recognition (ICPR)* (2010), pp. 3105–3108.

[77] WANG, Y., LI, Y., AND ZHENG, J. **A camera calibration technique based on opencv**. In *Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference on* (2010), IEEE, pp. 403–406.

[78] ZHANG, J., AND GONG, S. **Action categorization with modified hidden conditional random field**. *Pattern Recognition 43*, 1 (2010), 197–203.

[79] CHARLES, J., AND EVERINGHAM, M. **Learning shape models for monocular human pose estimation from the microsoft xbox kinect**. In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops),* (2011), pp. 1202–1208.

[80] COATES, A., AND NG, A. Y. **The importance of encoding versus training with sparse coding and vector quantization**. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)* (2011), pp. 921–928.

[81] HADFIELD, S., AND BOWDEN, R. **Kinecting the dots: Particle based scene flow from depth sensors**. In *IEEE International Conference on Computer Vision (ICCV)* (2011), pp. 2290–2295.

[82] HU, M., WANG, Y., ZHANG, Z., AND ZHANG, D. **Multi-view multi-stance gait identification**. In *Image Processing (ICIP), 2011 18th IEEE International Conference on* (2011), IEEE, pp. 541–544.

[83] KAUTZ, S. A., BOWDEN, M. G., CLARK, D. J., AND NEPTUNE, R. R. **Comparison of motor control deficits during treadmill and overground walking poststroke**. *Neurorehabilitation and neural repair 25*, 8 (2011), 756–765.

[84] REN, Z., YUAN, J., AND ZHANG, Z. **Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera**. In *Proceedings of the 19th ACM international conference on Multimedia* (2011), pp. 1093–1096.

[85] RUSU, R. B., AND COUSINS, S. **3d is here: Point cloud library (pcl)**. In *IEEE International Conference on Robotics and Automation (ICRA)* (2011), pp. 1–4.

[86] SAGAWA, Y., TURCOT, K., ARMAND, S., THEVENON, A., VUILLERME, N., AND WATELAIN, E. **Biomechanics and physiological parameters during gait in lower-limb amputees: a systematic review**. *Gait & posture 33*, 4 (2011), 511–526.

[87] SHOTTON, J., FITZGIBBON, A., COOK, M., SHARP, T., FINOCCHIO, M., MOORE, R., KIPMAN, A., AND BLAKE, A. **Real-time human pose recognition in parts from single depth images**. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (2011), Ieee, pp. 1297–1304.

[88] SIVAPALAN, S., CHEN, D., DENMAN, S., SRIDHARAN, S., AND FOOKES, C. **Gait energy volumes and frontal gait recognition using depth images**. In *IEEE International Joint Conference on Biometrics (IJCB)* (2011), pp. 1–6.

[89] STONE, E., AND SKUBIC, M. **Evaluation of an inexpensive depth camera for in-home gait assessment**. *Journal of Ambient Intelligence and Smart Environments 3*, 4 (2011), 349–361.

[90] SUNG, J., PONCE, C., SELMAN, B., AND SAXENA, A. **Human activity detection from rgbd images.** *Conference on Plan, Activity, and Intent Recognition 64* (2011).

[91] WEINLAND, D., RONFARD, R., AND BOYER, E. **A survey of vision-based methods for action representation, segmentation and recognition**. *Computer Vision and Image Understanding 115*, 2 (2011), 224–241.

[92] WREN, T. A., GORTON, G. E., OUNPUU, S., AND TUCKER, C. A. **Efficacy of clinical gait analysis: A systematic review**. *Gait & posture 34*, 2 (2011), 149–153.

[93] YANG, Y., AND RAMANAN, D. **Articulated pose estimation with flexible mixtures-of-parts**. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2011), pp. 1385–1392.

[94] CHENG, Z., QIN, L., YE, Y., HUANG, Q., AND TIAN, Q. **Human daily action analysis with multi-view and color-depth data**. In *European Conference on Computer Vision* (2012), Springer, pp. 52–61.

[95] DESTELLE, F., ROUDET, C., NEVEU, M., AND DIPANDA, A. **Toward a real-time tracking of dense point-sampled geometry**. In *Image Processing (ICIP), 2012 19th IEEE International Conference on* (2012), IEEE, pp. 381–384.

[96] ESSMAEEL, K., GALLO, L., DAMIANI, E., DE PIETRO, G., AND DIPANDÀ, A. **Temporal denoising of kinect depth data**. In *Signal Image Technology and Internet Based Systems (SITIS), 2012 Eighth International Conference on* (2012), IEEE, pp. 47–52.

[97] FAION, F., FRIEDBERGER, S., ZEA, A., AND HANEBECK, U. D. **Intelligent sensor-scheduling for multi-kinect-tracking**. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on* (2012), IEEE, pp. 3993–3999.

[98] FERN'NDEZ-BAENA, A., SUSIN, A., AND LLIGADAS, X. **Biomechanical validation of upper-body and lower-body joint movements of kinect motion capture data for rehabilitation treatments**. In *Intelligent Networking and Collaborative Systems (INCoS), 2012 4th International Conference on* (2012), IEEE, pp. 656–661.

[99] GABEL, M., GILAD-BACHRACH, R., RENSHAW, E., AND SCHUSTER, A. **Full body gait analysis with kinect**. In *Engineering in Medicine and Biology Society (EMBC), Annual International Conference of the IEEE* (2012), IEEE, pp. 1964–1967.

[100] GANAPATHI, V., PLAGEMANN, C., KOLLER, D., AND THRUN, S. **Real-time human pose tracking from range data**. In *European conference on computer vision* (2012), Springer, pp. 738–751.

[101] GARCIA, C. **A simple procedure for the comparison of covariance matrices**. *BMC evolutionary biology 12*, 1 (2012), 222.

[102] GATES, D. H., DARTER, B. J., DINGWELL, J. B., AND WILKEN, J. M. **Comparison of walking overground and in a computer assisted rehabilitation environment (caren) in individuals with and without transtibial amputation**. *Journal of neuroengineering and rehabilitation 9*, 1 (2012), 81.

[103] HERRERA, D., KANNALA, J., AND HEIKKILÄ, J. **Joint depth and color camera calibration with distortion correction**. *IEEE Transactions on Pattern Analysis and Machine Intelligence 34*, 10 (2012), 2058–2064.

[104] HINTERSTOISSER, S., LEPETIT, V., ILIC, S., HOLZER, S., BRADSKI, G. R., KONOLIGE, K., AND NAVAB, N. **Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes.** In *ACCV (1)* (2012), pp. 548–562.

[105] HODT-BILLINGTON, C. **Measures of symmetry in gait**. *Methodological principles and clinical choices [dissertation].[Bergen (Norway)]: University of Bergen* (2012).

[106] HOFMANN, M., BACHMANN, S., AND RIGOLL, G. **2.5 d gait biometrics using the depth gradient histogram energy image**. In *IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems* (2012), pp. 399–403.

[107] KOLAWOLE, A., AND TAVAKKOLI, A. **A novel gait recognition system based on hidden markov models**. *Advances in Visual Computing* (2012), 125–134.

[108] KURAKIN, A., ZHANG, Z., AND LIU, Z. **A real time system for dynamic hand gesture recognition with a depth sensor**. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European* (2012), IEEE, pp. 1975–1979.

[109] TAO, W., LIU, T., ZHENG, R., AND FENG, H. **Gait analysis using wearable sensors**. *Sensors 12*, 2 (2012), 2255–2283.

[110] VIEIRA, A. W., NASCIMENTO, E. R., OLIVEIRA, G. L., LIU, Z., AND CAMPOS, M. F. **Stop: Space-time occupancy patterns for 3d action recognition from depth map sequences**. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Springer, 2012, pp. 252–259.

[111] WANG, J., LIU, Z., WU, Y., AND YUAN, J. **Mining actionlet ensemble for action recognition with depth cameras**. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2012), pp. 1290–1297.

[112] XIA, L., CHEN, C.-C., AND AGGARWAL, J. **View invariant human action recognition using histograms of 3d joints**. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (2012), pp. 20–27.

[113] YANG, X., ZHANG, C., AND TIAN, Y. **Recognizing actions using depth motion maps-based histograms of oriented gradients**. In *Proceedings of the 20th ACM international conference on Multimedia* (2012), pp. 1057–1060.

[114] CHANG, J. Y., AND NAM, S. W. **Fast random-forest-based human pose estimation using a multi-scale and cascade approach**. *ETRI Journal 35*, 6 (2013), 949–959.

[115] CLARK, R. A., BOWER, K. J., MENTIPLAY, B. F., PATERSON, K., AND PUA, Y.-H. **Concurrent validity of the microsoft kinect for assessment of spatiotemporal gait variables**. *Journal of biomechanics 46*, 15 (2013), 2722–2725.

[116] DEVANNE, M., WANNOUS, H., BERRETTI, S., PALA, P., DAOUDI, M., AND DEL BIMBO, A. **Space-time pose representation for 3d human action recognition**. In *International Conference on Image Analysis and Processing* (2013), Springer, pp. 456–464.

[117] GIANARIA, E., BALOSSINO, N., GRANGETTO, M., AND LUCENTEFORTE, M. **Gait characterization using dynamic skeleton acquisition**. In *Multimedia Signal Processing (MMSP), 2013 IEEE 15th International Workshop on* (2013), IEEE, pp. 440–445.

[118] HU, M., WANG, Y., ZHANG, Z., ZHANG, D., AND LITTLE, J. J. **Incremental learning for video-based gait recognition with lbp flow**. *IEEE transactions on cybernetics 43*, 1 (2013), 77–89.

[119] JHUANG, H., GALL, J., ZUFFI, S., SCHMID, C., AND BLACK, M. J. **Towards understanding action recognition**. In *Proceedings of the IEEE international conference on computer vision* (2013), pp. 3192–3199.

[120] KODINARIYA, T. M., AND MAKWANA, P. R. **Review on determining number of cluster in k-means clustering**. *International Journal 1*, 6 (2013), 90–95.

[121] KOPPULA, H. S., GUPTA, R., AND SAXENA, A. **Learning human activities and object affordances from rgb-d videos**. *The International Journal of Robotics Research 32*, 8 (2013), 951–970.

[122] LIU, L., AND SHAO, L. **Learning discriminative representations from rgb-d video data.** In *IJCAI* (2013), vol. 1, p. 3.

[123] MILOVANOVIC, M., MINOVIC, M., AND STARCEVIC, D. **Walking in colors: human gait recognition using kinect and cbir**. *IEEE MultiMedia 20*, 4 (2013), 28–36.

[124] MUNARO, M., BALLIN, G., MICHIELETTO, S., AND MENEGATTI, E. **3d flow estimation for human action recognition from colored point clouds**. *Biologically Inspired Cognitive Architectures 5* (2013), 42–51.

[125] MUNARO, M., MICHIELETTO, S., AND MENEGATTI, E. **An evaluation of 3d motion flow and 3d pose estimation for human action recognition**. In *RSS Workshops: RGB-D: Advanced Reasoning with Depth Cameras* (2013).

[126] NEGIN, F., ÖZDEMIR, F., AKGÜL, C. B., YÜKSEL, K. A., AND ERÇIL, A. **A decision forest based feature selection framework for action recognition from rgb-depth cameras**. In *International Conference Image Analysis and Recognition* (2013), Springer, pp. 648–657.

[127] OFLI, F., CHAUDHRY, R., KURILLO, G., VIDAL, R., AND BAJCSY, R. **Berkeley mhad: A comprehensive multimodal human action database**. In *IEEE Workshop on Applications of Computer Vision (WACV)* (2013), pp. 53–60.

[128] OREIFEJ, O., AND LIU, Z. **Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2013), pp. 716–723.

[129] PISHCHULIN, L., ANDRILUKA, M., GEHLER, P., AND SCHIELE, B. **Poselet conditioned pictorial structures**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2013), pp. 588–595.

[130] SHOTTON, J., SHARP, T., KIPMAN, A., FITZGIBBON, A., FINOCCHIO, M., BLAKE, A., COOK, M., AND MOORE, R. **Real-time human pose recognition in parts from single depth images**. *Communications of the ACM 56*, 1 (2013), 116–124.

[131] TOMBARI, F. **Keypoints and features**. In *CGLibs Conference in Pisa* (2013), pp. 303–312.

[132] UGBOLUE, U. C., PAPI, E., KALIARNTAS, K. T., KERR, A., EARL, L., POMEROY, V. M., AND ROWE, P. J. **The evaluation of an inexpensive, 2d, video based gait assessment system for clinical use**. *Gait & posture 38*, 3 (2013), 483–489.

[133] WANG, F., STONE, E., SKUBIC, M., KELLER, J. M., ABBOTT, C., AND RANTZ, M. **Toward a passive low-cost in-home gait assessment system for older adults**. *IEEE journal of Biomedical and Health Informatics 17*, 2 (2013), 346–355.

[134] WANG, H., KLÄSER, A., SCHMID, C., AND LIU, C.-L. **Dense trajectories and motion boundary descriptors for action recognition**. *International journal of computer vision 103*, 1 (2013), 60–79.

[135] YE, M., ZHANG, Q., WANG, L., ZHU, J., YANG, R., AND GALL, J. **A survey on human motion analysis from depth data**. In *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*. Springer, 2013, pp. 149–187.

[136] ZHENG, J., AND JIANG, Z. **Learning view-invariant sparse representations for cross-view action recognition**. In *Proceedings of the IEEE International Conference on Computer Vision* (2013), pp. 3176–3183.

[137] ARAUJO, R., AND ANDERSSON, V. **Kinect gait biometry dataset - data from 164 individuals walking in front of a x-box 360 kinect sensor**, 07 2014.

[138] CESERACCIU, E., SAWACHA, Z., AND COBELLI, C. **Comparison of markerless and marker-based motion capture technologies through simultaneous data collection during gait: proof of concept**. *PloS one 9*, 3 (2014), e87640.

[139] CHATTOPADHYAY, P., SURAL, S., AND MUKHERJEE, J. **Frontal gait recognition from incomplete sequences using rgb-d camera**. *IEEE Transactions on Information Forensics and Security 9*, 11 (2014), 1843–1856.

[140] CHEN, X., AND YUILLE, A. L. **Articulated pose estimation by a graphical model with image dependent pairwise relations**. In *Advances in Neural Information Processing Systems* (2014), pp. 1736–1744.

[141] CIRUJEDA, P., AND BINEFA, X. **4dcov: a nested covariance descriptor of spatio-temporal features for gesture recognition in depth sequences**. In *2nd IEEE International Conference on 3D Vision* (2014), vol. 1, pp. 657–664.

[142] FILIPE, S., AND ALEXANDRE, L. A. **A comparative evaluation of 3d keypoint detectors in a rgb-d object dataset**. In *Computer Vision Theory and Applications (VISAPP), 2014 International Conference on* (2014), vol. 1, IEEE, pp. 476–483.

[143] GALNA, B., BARRY, G., JACKSON, D., MHIRIPIRI, D., OLIVIER, P., AND ROCHESTER, L. **Accuracy of the microsoft kinect sensor for measuring movement in people with parkinson's disease**. *Gait & posture 39*, 4 (2014), 1062–1068.

[144] HADFIELD, S., LEBEDA, K., AND BOWDEN, R. **Natural action recognition using invariant 3d motion encoding**. In *Computer Vision–ECCV 2014*. Springer, 2014, pp. 758–771.

[145] HOFMANN, M., GEIGER, J., BACHMANN, S., SCHULLER, B., AND RIGOLL, G. **The tum gait from audio, image and depth (gaid) database: Multimodal recognition of subjects and traits**. *Journal of Visual Communication and Image Representation 25*, 1 (2014), 195–206.

[146] JIANG, S., WANG, Y., ZHANG, Y., AND SUN, J. **Real time gait recognition system based on kinect skeleton feature**. In *Asian Conference on Computer Vision* (2014), Springer, pp. 46–57.

[147] KINGMA, D. P., AND BA, J. **Adam: A method for stochastic optimization**. *arXiv preprint arXiv:1412.6980* (2014).

[148] KWOLEK, B., KRZESZOWSKI, T., MICHALCZUK, A., AND JOSINSKI, H. **3d gait recognition using spatio-temporal motion descriptors**. In *Intelligent Information and Database Systems*. Springer, 2014, pp. 595–604.

[149] LI, S., LIU, Z.-Q., AND CHAN, A. B. **Heterogeneous multi-task learning for human pose estimation with deep convolutional neural network**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2014), pp. 482–489.

[150] MURO-DE-LA HERRAN, A., GARCIA-ZAPIRAIN, B., AND MENDEZ-ZORRILLA, A. **Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications**. *Sensors 14*, 2 (2014), 3362–3394.

[151] OHN-BAR, E., AND TRIVEDI, M. M. **Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations**. *IEEE transactions on intelligent transportation systems 15*, 6 (2014), 2368–2377.

[152] PADILLA-LÓPEZ, J. R., CHAARAOUI, A. A., AND FLÓREZ-REVUELTA, F. **A discussion on the validation tests employed to compare human action recognition methods using the msr action3d dataset**. *arXiv preprint arXiv:1407.7390* (2014).

[153] PAIEMENT, A., TAO, L., HANNUNA, S., CAMPLANI, M., DAMEN, D., AND MIRMEHDI, M. **Online quality assessment of human movement from skeleton data**. In *British Machine Vision Conference* (2014), BMVA press, pp. 153–166.

[154] PAOLINI, G., PERUZZI, A., MIRELMAN, A., CEREATTI, A., GAUKRODGER, S., HAUSDORFF, J. M., AND DELLA CROCE, U. **Validation of a method for real time foot position and orientation tracking with microsoft kinect technology for use in virtual reality and treadmill based gait training programs**. *IEEE Transactions on Neural Systems and Rehabilitation Engineering 22*, 5 (2014), 997–1002.

[155] PFISTER, A., WEST, A. M., BRONNER, S., AND NOAH, J. A. **Comparative abilities of microsoft kinect and vicon 3d motion capture for gait analysis**. *Journal of medical engineering & technology 38*, 5 (2014), 274–280.

[156] ROCHA, A. P., CHOUPINA, H., FERNANDES, J. M., ROSAS, M. J., VAZ, R., AND CUNHA, J. P. S. **Parkinson's disease assessment based on gait analysis using an innovative rgb-d camera system**. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2014), IEEE, pp. 3126–3129.

[157] TANG, D., JIN CHANG, H., TEJANI, A., AND KIM, T.-K. **Latent regression forest: Structured estimation of 3d articulated hand posture**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 3786–3793.

[158] TANG, J., LUO, J., TJAHJADI, T., AND GAO, Y. **2.5 d multi-view gait recognition based on point cloud registration**. *Sensors 14*, 4 (2014), 6124–6143.

[159] TOMPSON, J. J., JAIN, A., LECUN, Y., AND BREGLER, C. **Joint training of a convolutional network and a graphical model for human pose estimation**. In *Advances in neural information processing systems* (2014), pp. 1799–1807.

[160] VEMULAPALLI, R., ARRATE, F., AND CHELLAPPA, R. **Human action recognition by representing 3d skeletons as points in a lie group**. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2014), pp. 588–595.

[161] XIAO, Y., ZHAO, G., YUAN, J., AND THALMANN, D. **Activity recognition in unconstrained rgb-d video using 3d trajectories**. In *ACM SIGGRAPH Asia Autonomous Virtual Humans and Social Robot for Telepresence* (2014), p. 4.

[162] YANG, X., AND TIAN, Y. **Super normal vector for activity recognition using depth sequences**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 804–811.

[163] YE, M., AND YANG, R. **Real-time simultaneous pose and shape estimation for articulated objects using a single depth camera**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 2345–2352.

[164] AHMED, F., PAUL, P. P., AND GAVRILOVA, M. L. **Dtw-based kernel and rank-level fusion for 3d gait recognition using kinect**. *The Visual Computer 31*, 6-8 (2015), 915–924.

[165] ALOTAIBI, M., AND MAHMOOD, A. **Automatic real time gait recognition based on spatiotemporal templates**. In *IEEE Systems, Applications and Technology Conference (LISAT)* (2015), pp. 1–5.

[166] ANDERSSON, V. O., AND DE ARAÚJO, R. M. **Person identification using anthropometric and gait data from kinect sensor.** In *AAAI* (2015), pp. 425–431.

[167] AUVINET, E., MULTON, F., AUBIN, C.-E., MEUNIER, J., AND RAISON, M. **Detection of gait cycles in treadmill walking using a kinect**. *Gait & posture 41*, 2 (2015), 722–725.

[168] AUVINET, E., MULTON, F., AND MEUNIER, J. **New lower-limb gait asymmetry indices based on a depth camera**. *Sensors 15*, 3 (2015), 4605–4623.

[169] CHAARAOUI, A. A., PADILLA-LÓPEZ, J. R., AND FLÓREZ-REVUELTA, F. **Abnormal gait detection with rgb-d devices using joint motion history features**. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on* (2015), vol. 7, IEEE, pp. 1–6.

[170] CHÉRON, G., LAPTEV, I., AND SCHMID, C. **P-cnn: Pose-based cnn features for action recognition**. In *Proceedings of the IEEE international conference on computer vision* (2015), pp. 3218–3226.

[171] CIPPITELLI, E., GASPARRINI, S., SPINSANTE, S., AND GAMBI, E. **Kinect as a tool for gait analysis: validation of a real-time joint extraction algorithm working in side view**. *Sensors 15*, 1 (2015), 1417–1434.

[172] CLARK, R. A., VERNON, S., MENTIPLAY, B. F., MILLER, K. J., MCGINLEY, J. L., PUA, Y. H., PATERSON, K., AND BOWER, K. J. **Instrumenting gait assessment using the kinect in people living with stroke: reliability and association with balance tests**. *Journal of neuroengineering and rehabilitation 12*, 1 (2015), 1.

[173] DING, W., LIU, K., CHENG, F., AND ZHANG, J. **Stfc: spatio-temporal feature chain for skeleton-based human action recognition**. *Journal of Visual Communication and Image Representation 26* (2015), 329–337.

[174] ESSMAEEL, K., MIGNIOT, C., AND DIPANDA, A. **Une nouvelle approche de classification de personnes à partir d'une plate-forme multi-kinect**. In *Journées francophones des jeunes chercheurs en vision par ordinateur* (2015).

[175] GEERSE, D. J., COOLEN, B. H., AND ROERDINK, M. **Kinematic validation of a multi-kinect v2 instrumented 10-meter walkway for quantitative gait assessments**. *PloS one 10*, 10 (2015).

[176] JAIMEZ, M., SOUIAI, M., GONZALEZ-JIMENEZ, J., AND CREMERS, D. **A primal-dual framework for real-time dense rgb-d scene flow**. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on* (2015), IEEE, pp. 98–104.

[177] JAIN, A. K., AND ROSS, A. **Bridging the gap: From biometrics to forensics**. *To appear in Philosophical Transactions of The Royal Society B* (2015), 2.

[178] KASTANIOTIS, D., THEODORAKOPOULOS, I., THEOHARATOS, C., ECONOMOU, G., AND FOTOPOULOS, S. **A framework for gait-based recognition using kinect**. *Pattern Recognition Letters 68* (2015), 327–335.

[179] KONG, Y., SATARBOROUJENI, B., AND FU, Y. **Hierarchical 3d kernel descriptors for action recognition using depth sequences**. In *11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (2015), vol. 1, pp. 1–6.

[180] LEIGHTLEY, D., YAP, M. H., COULSON, J., BARNOUIN, Y., AND MCPHEE, J. S. **Benchmarking human motion analysis using kinect one: an open source dataset**. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2015 Asia-Pacific* (2015), IEEE, pp. 1–7.

[181] LIM, C. D., CHENG, C.-Y., WANG, C.-M., CHAO, Y., AND FU, L.-C. **Depth image based gait tracking and analysis via robotic walker**. In *IEEE International Conference on Robotics and Automation (ICRA)* (2015), pp. 5916–5921.

[182] MENTIPLAY, B. F., PERRATON, L. G., BOWER, K. J., PUA, Y.-H., MCGAW, R., HEYWOOD, S., AND CLARK, R. A. **Gait assessment using the microsoft xbox one kinect: Concurrent validity and inter-day reliability of spatiotemporal and kinematic variables**. *Journal of biomechanics 48*, 10 (2015), 2166–2170.

[183] MOORE, J. K., HNAT, S. K., AND VAN DEN BOGERT, A. J. **An elaborate data set on human gait and the effect of mechanical perturbations**. *PeerJ 3* (2015), e918.

[184] OHN-BAR, E., AND TRIVEDI, M. M. **A comparative study of color and depth features for hand gesture recognition in naturalistic driving settings**. In *IEEE Intelligent Vehicles Symposium (IV)* (2015), pp. 845–850.

[185] PAILLARD, T., AND NOÉ, F. **Techniques and methods for testing the postural function in healthy and pathological subjects**. *BioMed research international 2015* (2015).

[186] PAPAGEORGIOU, X. S., CHALVATZAKI, G., TZAFESTAS, C. S., AND MARAGOS, P. **Hidden markov modeling of human pathological gait using laser range finder for an assisted living intelligent robotic walker**. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on* (2015), IEEE, pp. 6342–6347.

[187] PHONSING, P., LAOARPHAKUN, S., THAINIMIT, S., KEINPRASIT, R., AND KOIKE, Y. **Multi kinect cameras setup for skeleton based action recognition**. In *Information and Communication Technology for Embedded Systems (IC-ICTES), 2015 6th International Conference of* (2015), IEEE, pp. 1–6.

[188] RAUTARAY, S. S., AND AGRAWAL, A. **Vision based hand gesture recognition for human computer interaction: a survey**. *Artificial Intelligence Review 43*, 1 (2015), 1–54.

[189] SALBACH, N. M., O'BRIEN, K. K., BROOKS, D., IRVIN, E., MARTINO, R., TAKHAR, P., CHAN, S., AND HOWE, J.-A. **Reference values for standardized tests of walking speed and distance: a systematic review**. *Gait & posture 41*, 2 (2015), 341–360.

[190] SLOOT, L. H., HARLAAR, J., AND VAN DER KROGT, M. M. **Self-paced versus fixed speed walking and the effect of virtual reality in children with cerebral palsy**. *Gait & posture 42*, 4 (2015), 498–504.

[191] WANG, Q., KURILLO, G., OFLI, F., AND BAJCSY, R. **Evaluation of pose tracking accuracy in the first and second generations of microsoft kinect**. In *Healthcare Informatics (ICHI), 2015 International Conference on* (2015), IEEE, pp. 380–389.

[192] WOHLHART, P., AND LEPETIT, V. **Learning descriptors for object recognition and 3d pose estimation**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 3109–3118.

[193] WU, C., ZHANG, J., SAVARESE, S., AND SAXENA, A. **Watch-n-patch: Unsupervised understanding of actions and relations**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 4362–4370.

[194] YUB JUNG, H., LEE, S., SEOK HEO, Y., AND DONG YUN, I. **Random tree walk toward instantaneous 3d human pose estimation**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 2467–2474.

[195] ZHANG, H., REARDON, C., ZHANG, C., AND PARKER, L. E. **Adaptive human-centered representation for activity recognition of multiple individuals from 3d point cloud sequences**. In *IEEE International Conference on Robotics and Automation (ICRA)* (2015), pp. 1991–1998.

[196] BOGO, F., KANAZAWA, A., LASSNER, C., GEHLER, P., ROMERO, J., AND BLACK, M. J. **Keep it smpl: Automatic estimation of 3d human pose and shape from a single image**. In *European Conference on Computer Vision* (2016), Springer, pp. 561–578.

[197] CHEN, C., ZHANG, B., HOU, Z., JIANG, J., LIU, M., AND YANG, Y. **Action recognition from depth sequences using weighted fusion of 2d and 3d auto-correlation of gradients features**. *Multimedia Tools and Applications* (2016), 1–19.

[198] CHENG, H., DAI, Z., LIU, Z., AND ZHAO, Y. **An image-to-class dynamic time warping approach for both 3d static and trajectory hand gesture recognition**. *Pattern Recognition 55* (2016), 137–147.

[199] DEVANNE, M., WANNOUS, H., DAOUDI, M., BERRETTI, S., DEL BIMBO, A., AND PALA, P. **Learning shape variations of motion trajectories for gait analysis**. In *Pattern Recognition (ICPR), 2016 23rd International Conference on* (2016), IEEE, pp. 895–900.

[200] ESSMAEEL, K., MIGNIOT, C., AND DIPANDA, A. **3d descriptor for an oriented-human classification from complete point cloud.** In *VISIGRAPP (4: VISAPP)* (2016), pp. 353–360.

[201] FENG, Y., LI, Y., AND LUO, J. **Learning effective gait features using lstm**. In *23rd International Conference on Pattern Recognition (ICPR)* (2016), IEEE, pp. 325–330.

[202] FITERAU, M., FRIES, J., HALILAJ, E., SIRANART, N., BHOOSHAN, S., AND RÉ, C. **Similarity-based lstms for time series representation learning in the presence of structured covariates**. *29th Conference on Neural Information Processing Systems* (2016).

[203] IBAÑEZ, R., SORIA, A., TEYSEYRE, A. R., BERDUN, L., AND CAMPO, M. R. **A comparative study of machine learning techniques for gesture recognition using kinect**. In *Handbook of Research on Human-Computer Interfaces, Developments, and Applications*. IGI Global, 2016, pp. 1–22.

[204] KASTANIOTIS, D., THEODORAKOPOULOS, I., AND FOTOPOULOS, S. **Pose-based gait recognition with local gradient descriptors and hierarchically aggregated residuals**. *Journal of Electronic Imaging 25*, 6 (2016), 063019–063019.

[205] KHOKHLOVA, M., MIGNIOT, C., AND DIPANDA, A. **3d visual-based human motion descriptors: A review**. In *Signal-Image Technology & Internet-Based Systems (SITIS), 2016 12th International Conference on* (2016), IEEE, pp. 564–572.

[206] KOO, T. K., AND LI, M. Y. **A guideline of selecting and reporting intraclass correlation coefficients for reliability research**. *Journal of chiropractic medicine 15*, 2 (2016), 155–163.

[207] LAZZARINI, B. S. R., AND KATARAS, T. J. **Treadmill walking is not equivalent to overground walking for the study of walking smoothness and rhythmicity in older adults**. *Gait & posture 46* (2016), 42–46.

[208] LIU, D.-X., DU, W., WU, X., WANG, C., AND QIAO, Y. **Deep rehabilitation gait learning for modeling knee joints of lower-limb exoskeleton**. In *Robotics and Biomimetics (ROBIO), 2016 IEEE International Conference on* (2016), IEEE, pp. 1058–1063.

[209] MAHASSENI, B., AND TODOROVIC, S. **Regularizing long short term memory with 3d human-skeleton sequences for action recognition**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 3054–3062.

[210] MENG, M., DRIRA, H., DAOUDI, M., AND BOONAERT, J. **Detection of abnormal gait from skeleton data.** In *VISIGRAPP (3: VISAPP)* (2016), pp. 133–139.

[211] NGUYEN, T.-N., HUYNH, H.-H., AND MEUNIER, J. **Skeleton-based abnormal gait detection**. *Sensors 16*, 11 (2016), 1792.

[212] NORDIN, M. J., AND SAADOON, A. **A survey of gait recognition based on skeleton mode l for human identification**. *Research Journal of Applied Sciences, Engineering and Technology* (2016).

[213] OTTE, K., KAYSER, B., MANSOW-MODEL, S., VERREL, J., PAUL, F., BRANDT, A. U., AND SCHMITZ-HÜBSCH, T. **Accuracy and reliability of the kinect version 2 for clinical measurement of motor function**. *PloS one 11*, 11 (2016), e0166532.

[214] PENG, B., AND LUO, Z. **Multi-view 3d pose estimation from single depth images**. Tech. rep., Technical report, Stanford University, USA, Report, Course CS231n: Convolutional Neural Networks for Visual Recognition, 2016.

[215] SARAFIANOS, N., BOTEANU, B., IONESCU, B., AND KAKADIARIS, I. A. **3d human pose estimation: A review of the literature and analysis of covariates**. *Computer Vision and Image Understanding 152* (2016), 1–20.

[216] SHAFAEI, A., AND LITTLE, J. J. **Real-time human motion capture with multiple depth cameras**. In *IEEE 13th Conference on Computer and Robot Vision (CRV)* (2016), pp. 24–31.

[217] SPRINGER, S., AND YOGEV SELIGMANN, G. **Validity of the kinect for gait assessment: A focused review**. *Sensors 16*, 2 (2016), 194.

[218] SRIDHAR, S., MUELLER, F., ZOLLHÖFER, M., CASAS, D., OULASVIRTA, A., AND THEOBALT, C. **Real-time joint tracking of a hand manipulating an object from rgb-d input**. *arXiv preprint arXiv:1610.04889* (2016).

[219] WANG, W.-J., CHANG, J.-W., HAUNG, S.-F., AND WANG, R.-J. **Human posture recognition based on images captured by the kinect sensor**. *International Journal of Advanced Robotic Systems 13*, 2 (2016), 54.

[220] WEI, S.-E., RAMAKRISHNA, V., KANADE, T., AND SHEIKH, Y. **Convolutional pose machines**. In *CVPR* (2016).

[221] WU, D., PIGOU, L., KINDERMANS, P.-J., LE, N. D.-H., SHAO, L., DAMBRE, J., AND ODOBEZ, J.-M. **Deep dynamic neural networks for multimodal gesture segmentation and recognition**. *IEEE transactions on pattern analysis and machine intelligence 38*, 8 (2016), 1583–1597.

[222] ZHANG, H., AND PARKER LYNNE, E. **Code4d: color-depth local spatio-temporal features for human activity recognition from rgb-d videos**. *IEEE Transaction on Circuits Syst Video Technol 26*, 3 (2016), 541–555.

[223] ZHU, W., LAN, C., XING, J., ZENG, W., LI, Y., SHEN, L., XIE, X., AND OTHERS. **Co-occurrence feature learning for skeleton based action recognition using regularized deep lstm networks.** In *AAAI* (2016), vol. 2, p. 6.

[224] BELGHALI, M., CHASTAN, N., CIGNETTI, F., DAVENNE, D., AND DECKER, L. M. **Loss of gait control assessed by cognitive-motor dual-tasks: pros and cons in detecting people at risk of developing alzheimer's and parkinson's diseases**. *GeroScience* (2017), 1–25.

[225] CAO, Z., SIMON, T., WEI, S.-E., AND SHEIKH, Y. **Realtime multi-person 2d pose estimation using part affinity fields**. In *CVPR* (2017).

[226] CAVAZZA, J., MORERIO, P., AND MURINO, V. **When kernel methods meet feature learning: Log-covariance network for action recognition from skeletal data**. *arXiv preprint arXiv:1708.01022* (2017).

[227] CHEN, C.-H., AND RAMANAN, D. **3d human pose estimation= 2d pose estimation+ matching**. *Computer Vision and Pattern Recognition (CVPR)* (2017).

[228] ELTOUKHY, M., OH, J., KUENZE, C., AND SIGNORILE, J. **Improved kinect-based spatiotemporal and kinematic treadmill gait assessment**. *Gait & posture 51* (2017), 77–83.

[229] ELTOUKHY, M., OH, J., KUENZE, C., AND SIGNORILE, J. **Improved kinect-based spatiotemporal and kinematic treadmill gait assessment**.

[230] GAO, Z., LI, S., ZHU, Y., WANG, C., AND ZHANG, H. **Collaborative sparse representation leaning model for rgbd action recognition**. *Journal of Visual Communication and Image Representation* (2017).

[231] GEERSE, D., COOLEN, B., KOLIJN, D., AND ROERDINK, M. **Validation of foot placement locations from ankle data of a kinect v2 sensor**. *Sensors 17*, 10 (2017), 2301.

[232] GUESS, T. M., RAZU, S., JAHANDAR, A., SKUBIC, M., AND HUO, Z. **Comparison of 3d joint angles measured with the kinect 2.0 skeletal tracker versus a marker-based motion capture system**. *Journal of applied biomechanics 33*, 2 (2017), 176–181.

[233] HAN, Y., ZHANG, P., ZHUO, T., HUANG, W., AND ZHANG, Y. **Video action recognition based on deeper convolution networks with pair-wise frame motion concatenation**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2017), pp. 8–17.

[234] KADAMBI, A., AND RASKAR, R. **Rethinking machine vision time of flight with ghz heterodyning**. *IEEE Access 5* (2017), 26211–26223.

[235] KLARENBEEK, G., HARMANNY, R., AND CIFOLA, L. **Multi-target human gait classification using lstm recurrent neural networks applied to micro-doppler**. In *Radar Conference (EURAD), 2017 European* (2017), IEEE, pp. 167–170.

[236] LAN, Z., ZHU, Y., HAUPTMANN, A. G., AND NEWSAM, S. **Deep local video feature for action recognition**. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2017), pp. 1219–1225.

[237] LITANY, O., BRONSTEIN, A., BRONSTEIN, M., AND MAKADIA, A. **Deformable shape completion with graph convolutional autoencoders**. *arXiv preprint arXiv:1712.00268* (2017).

[238] MERRIAUX, P., DUPUIS, Y., BOUTTEAU, R., VASSEUR, P., AND SAVATIER, X. **A study of vicon system positioning performance**. *Sensors 17*, 7 (2017), 1591.

[239] PAPEGAAIJ, S., AND STEENBRINK, F. **Clinical gait analysis: Treadmill-based vs overground**.

[240] QI, C. R., SU, H., MO, K., AND GUIBAS, L. J. **Pointnet: Deep learning on point sets for 3d classification and segmentation**. *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE 1*, 2 (2017), 4.

[241] RAJAGOPALAN, R., LITVAN, I., AND JUNG, T.-P. **Fall prediction and prevention systems: recent trends, challenges, and future research directions**. *Sensors 17*, 11 (2017), 2509.

[242] ROBERTS, M., MONGEON, D., AND PRINCE, F. **Biomechanical parameters for gait analysis: a systematic review of healthy human gait**. *Physical Therapy and Rehabilitation 4*, 1 (2017), 6.

[243] SAHA, P. K., BORGEFORS, G., AND DI BAJA, G. S. **Skeletonization and its applications–a review**. In *Skeletonization*. Elsevier, 2017, pp. 3–42.

[244] SIMON, T., JOO, H., MATTHEWS, I., AND SHEIKH, Y. **Hand keypoint detection in single images using multiview bootstrapping**. In *CVPR* (2017).

[245] ZHANG, H., ZHONG, P., HE, J., AND XIA, C. **Combining depth-skeleton feature with sparse coding for action recognition**. *Neurocomputing 230* (2017), 417–426.

[246] ZHANG, S., LIU, X., AND XIAO, J. **On geometric features for skeleton-based action recognition using multilayer lstm networks**. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* (2017), IEEE, pp. 148–157.

[247] DRUMOND, R. R., MARQUES, B. A. D., VASCONCELOS, C. N., AND CLUA, E. **Peek - an lstm recurrent network for motion classification from sparse data**. In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 1: GRAPP,* (2018), INSTICC, SciTePress, pp. 215–222.

[248] KHOKHLOVA, MARGARITA, C. M., AND DIPANDA, A. **Human posture recognition and representation without skeleton data from 3d point cloud**. In *Visapp 18* (2018).

[249] KHOKHLOVA, M., MIGNIOT, C., AND DIPANDA, A. **Advances in description of 3d human motion**. *Multimedia Tools and Applications* (2018), 1–27.

[250] LI, Q., WANG, Y., SHARF, A., CAO, Y., TU, C., CHEN, B., AND YU, S. **Classification of gait anomalies from kinect**. *The Visual Computer 34*, 2 (2018), 229–241.

[251] NGUYEN, T.-N., AND MEUNIER, J. **Walking gait dataset: point clouds, skeletons and silhouettes**. Tech. Rep. 1379, DIRO, University of Montreal, April 2018.

[252] RIZA ALP GÜLER, NATALIA NEVEROVA, I. K. **Densepose: Dense human pose estimation in the wild**.

[253] SUN, J., WANG, Y., LI, J., WAN, W., CHENG, D., AND ZHANG, H. **View-invariant gait recognition based on kinect skeleton feature**. *Multimedia Tools and Applications* (2018), 1–27.

[254] ZHAO, A., QI, L., LI, J., DONG, J., AND YU, H. **Lstm for diagnosis of neurodegenerative diseases using gait data**. In *Ninth International Conference on Graphic and Image Processing (ICGIP 2017)* (2018), vol. 10615, International Society for Optics and Photonics, p. 106155B.

[255] ABBY CAIN, N. O. **Gait deviations in amputees**.

[256] CONSORTIUM, O., AND OTHERS. **Openni, the standard framework for 3d sensing**. URL as accessed on 2017-09-30.

[257] KHUMAN, R. **Gait: normal abnormal**.

[258] NG, A. **Sequence models**. https://fr.coursera.org/learn/nlp-sequence-models. Accessed: 2018-23-06.

List of Figures

List of Tables

GLOSSARY

CNN is a term from machine learning. A class of deep, feed-forward artificial neural networks, most commonly applied to analyzing visual imagery 33, 119

DTW is an algorithm for measuring similarity between two temporal sequences 41, 53

LSTM is a special case of recurrent neural network (RNN) 11, 33, 45–47, 49, 103, 119–123, 126

ML is an application of artificial intelligence that provides systems the ability to automatically learn and improve from experience without being explicitly programmed 66, 88, 102, 103, 119, 123, 126

MOCAP is a process of digital recording of people's or objects' movements 19, 57, 84

PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components 38

RNN is a type of advanced artificial neural network (ANN) that involves directed cycles in memory 119

SVM are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis 98, 100, 110, 111, 113

ACRONYMS

GMM Gaussian Mixture Model 38, 47, 110

GPU Graphics Processing Unit 64

GRF Ground-Force 23, 26

HMM Hidden Markov Model 33, 37, 42, 45–47, 49

IR Infra-Red 3, 7

K-NN K-Nearest Neighbors 51, 97, 108, 110, 114–119, 123, 126

LST Local Spatio-Temporal 32, 40, 48

NWS Non-Wearable Sensors 2, 3, 5, 18, 19

ROI Region of Interest 29

SDK Software Development Kit 19, 64, 67, 70, 75, 84, 88, 102, 123, 150

TOF Time of Flight 3, 7, 62

WS Wearable Sensors 2, 3, 5, 19, 24, 106

I

ANNEXES

A

First Chapter of Annexes

A.1/ Gait Deviations in Amputees

Common deviations for prosthesis patients can be grouped in three categories [255] listed in Tables A.1 A.2 A.3.

Table A.1: Common prosthesis gait deviations: Transtibial

Name	Description	Causes
Absent knee flexion	Knee fully extended at heel strike	Faulty suspension of the prosthesis - too soft heel cushion or plantar flexor bumpers Foot placement too far forward on stepping Lack of pre-flexion of the socket Discomfort/pain Quads weakness
Excessive knee flexion	Increased knee flexion at heel strike (or mid stance), patient feels as though walking downhill	Faulty suspension of prosthesis Prosthetic foot set in too much dorsiflexion Stiff heel cushion Flexion contracture of the knee Foot too posterior in relation to socket
External rotation of foot at heel strike	External rotation of the prosthesis/foot at heel strike	heel to hard loose socket
Knee instabiliy	Knee flexion 'jerky' in presentation during heel strike to foot flat	Weak Quadriceps
Valgus/Varus moment	Knee shifts medially or laterally during prosthetic stance phase	Foot placement (medial placement causes lateral thrust and vice versa) Foot alignment on the prosthesis Socket loose
Drop Off	Heel off occurs too early causing early knee flexion	Foot too posterior on the prosthesis in relation to the socket Excessive dorsiflexion of the foot on the prosthesis Soft heel bumper on the prosthesis
Knee Hyperextension	Delayed heel causing hyperextension of the knee, walking up hill sensation	Foot set too far forward on the prosthesis in relation to socket Too hard a heel cushion Too much plantar flexion on the foot
Whip	During swing phase foot 'whips' laterally or medially	Poor suspension Knee internally or externally rotated
Pistoning	Amputee drops into the socket as the foot moves into flat foot, tibia moves vertically during alternately weight bearing and non-weight bearing periods of gait	Lack of prosthetic socks Suspension loose or inadequate Too large or faulty socket

Table A.2: Common prosthesis gait deviations: Transfemoral

Name	Description	Causes
Prosthetic Instability	The prosthetic knee has a tendency to buckle on weight bearing	Knee set too far anterior Heel cushion too firm Weak hip extensors Heel of the shoe too high causing the pylon of the prosthesis to move anteriorly Severe hip flexion contracture
Foot Slap	Foot progresses too quickly from heel strike to foot flat, creating a slapping noise	Patient forcing foot contact to gain knee stability Heel cushion too soft Plantar flexion cushion too soft Excessive dorsiflexion
Abducted Gait	Increased base of support during mobility, prosthetic foot placement is lateral to the normal foot placement during the gait cycle	Prosthesis too long Socket too small Suspension belt may be insufficient-band may be too far from the ileum Pain in the groin or medial wall of the prosthesis Hip abductor contractures Lateral wall of the prosthesis not supporting the femur sufficiently Socket of prosthesis abducted in alignment Fear/lack of confidence transferring weight onto prosthesis Alignment of the lower half of the pylon of the prosthesis in relation to socket
Lateral Trunk Bending	Trunk flexes towards prosthesis during prosthetic stance phase	Prosthesis too short Short stump length Weak or contracted hip abductors Foot outset excessively in relation to socket Lack of prosthetic lateral wall support Pain on the lateral distal end of the stump Lack of balance Habit
Anterior Trunk Bending	Trunk flexes forwards during prosthetic stance phase	Short stump length Weak or contracted hip abductors Foot outset excessively in relation to socket Lack of prosthetic lateral wall support Pain on the lateral distal end of the stump Lack of balance Habit
Increased Lumbar Lordosis	Lumbar lordosis is exaggerated during prosthetic stance phase	Poor shaping of posterior wall of the prosthesis or pain on ischial weight bearing, resulting in anterior pelvic rotation Flexion contracture at the hip Weak hip extensor Habit Poor abdominal muscles Lack of support from the anterior wall of the socket Insufficient socket flexion
Whip (during swing phase)	At toe off heel moves laterally or medially	Prosthetic knee alignment Incorrect donning of the prosthesis i.e. applied internally rotated or externally rotated weakness around femur Prosthetic too tight

Pistoning	Socket dropping off when prosthesis lifted	Insufficient suspension Socket too loose or delayed knee flexion during toe off due to increased resistance of the prosthesis Alignment of prosthesis prosthetic knee lack of friction
Excessive Heel Rise	Prosthetic heel rises more than sound side	Amputee generating more force then needed to gain knee flexion Poor/lack of extension aid
Reduced Heel Rise	Prosthetic heel does not rise as much as sound side	Locked knee Lack of hip flexion Too much friction on free knee Extension aid to tight
Circum-duction	Lateral curvature of swing phase of prosthesis	Prosthesis too long Fixed knee and poor hip hitching Poor suspension causing prosthesis to slip Excessive foot plantar flexion Abduction contractures Weak hip flexors Socket too small Insufficient knee flexion
Vaulting	Amputee rises onto toe of the non prosthetic limb during prosthetic swing phase	Prosthesis too long Fear of catching toe on the floor Insufficient knee flexion due to decreased confidence a 'locked/fixed knee' Poor suspension prosthesis-slips off during swing phase Socket too small Excessive friction on knee flexion of the prosthesis
Terminal Impact	Forcible impact as knee goes into extension at end of terminal swing phase, just before heel strike	Lack of friction of knee flexion Extension aid too excessive Absent extension bumper Amputee deliberately snaps knee into extension by excessive force to ensure extension

Table A.3: Common prosthesis gait deviations: Both Transfermoral and Transtibial

Name	Description	Causes
Uneven Step Length	Steps are of uneven duration or length, usually a short stance phase on the prosthetic side	Fixed flexion deformity at knee Insufficient friction of prosthetic knee creating an increased step length on prosthetic side Hip flexion contracture Pain leading to decreased weight bearing on prosthetic side Fear Poor balance Painful poorly fitting socket
Uneven Arm Swing (secondary deviation)	Arm on the prosthetic side is held close to the body	Poor prosthetic fit Poor balance Fear Habit Always due to other gait deviations and lack of training

A.2/ GAIT ANALYSIS REPORT

A clinical report for the patient from Proteor is presented below. All the data about patient is removed from the report to guarantee the privacy of the person.

MODELE

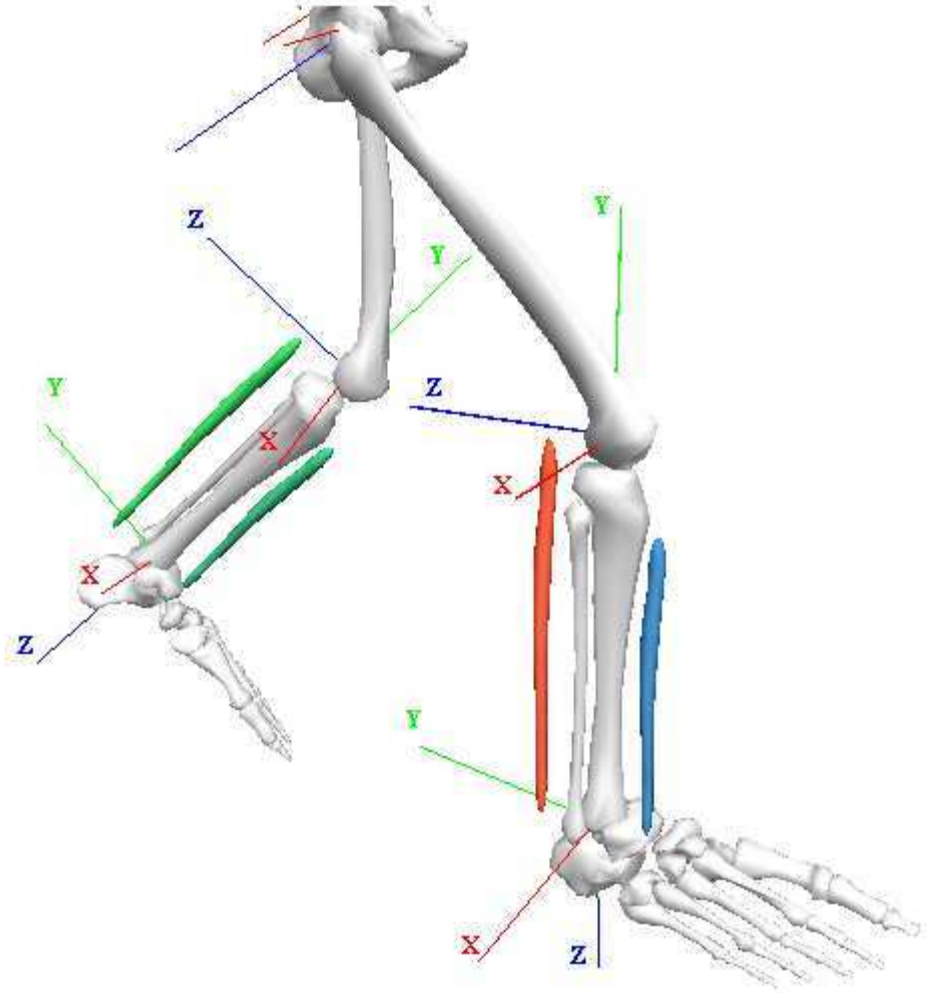
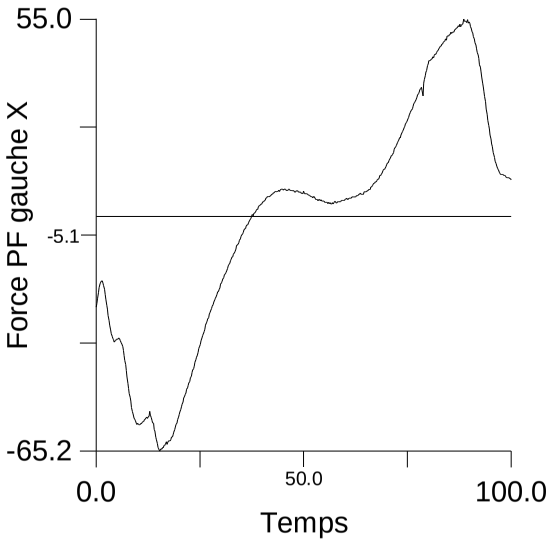
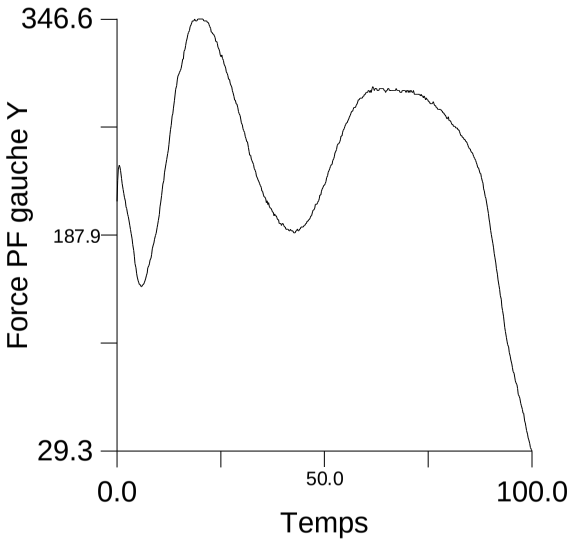
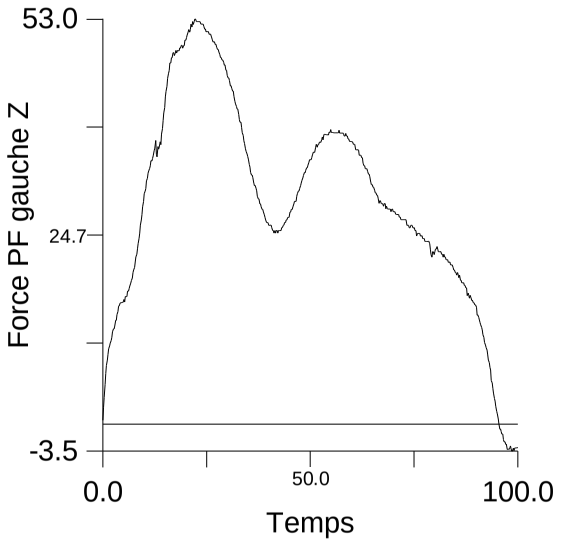
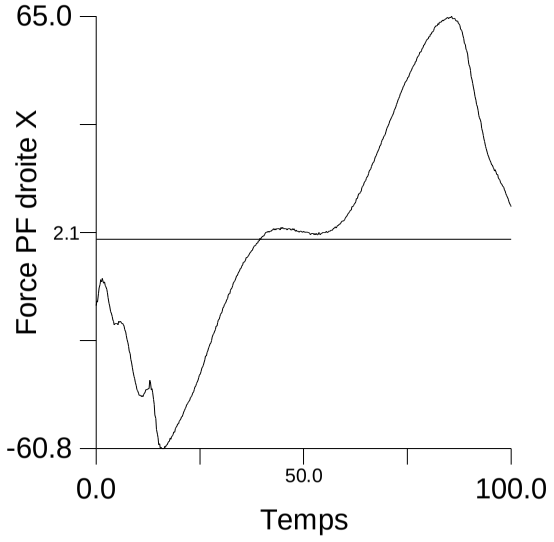


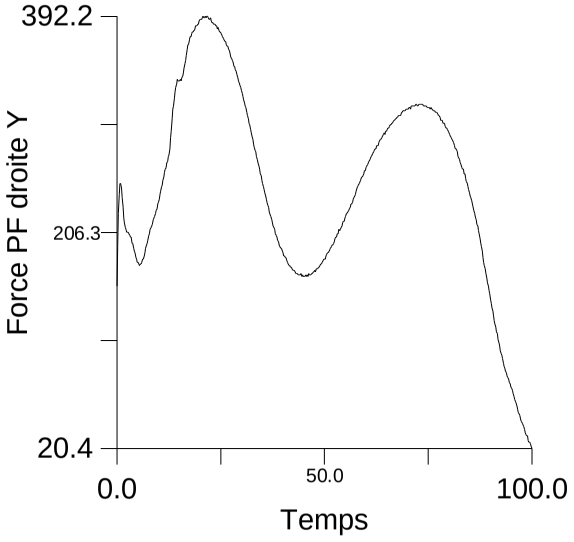
PLATE FORME DE FORCE

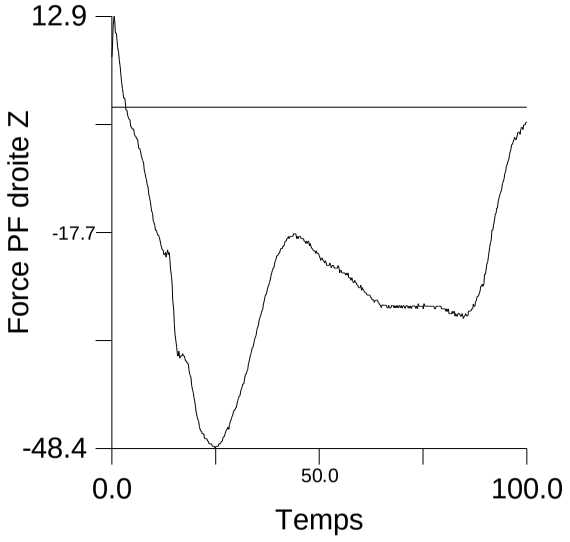




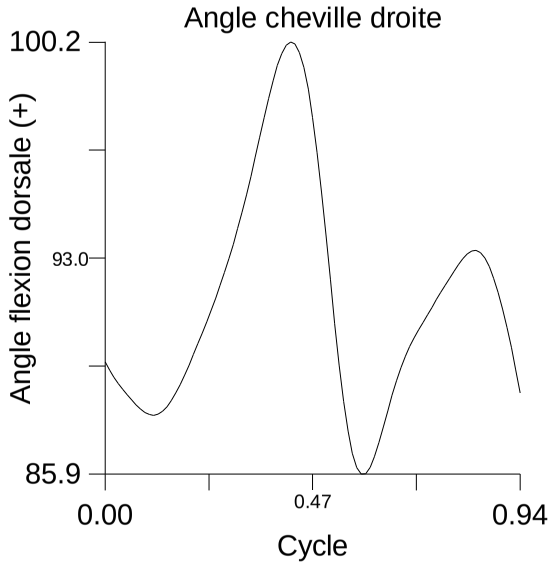


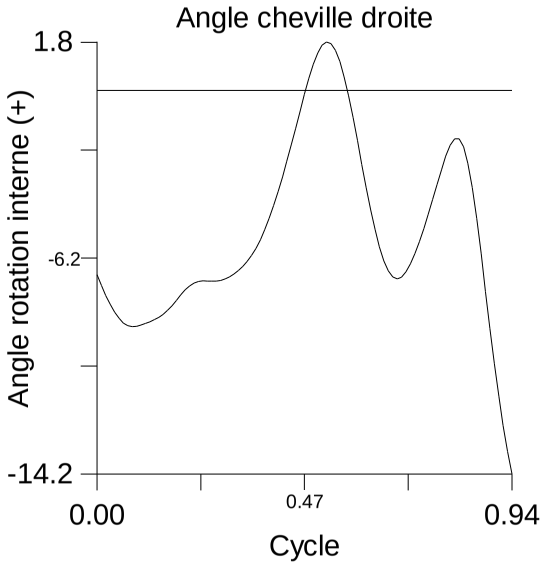


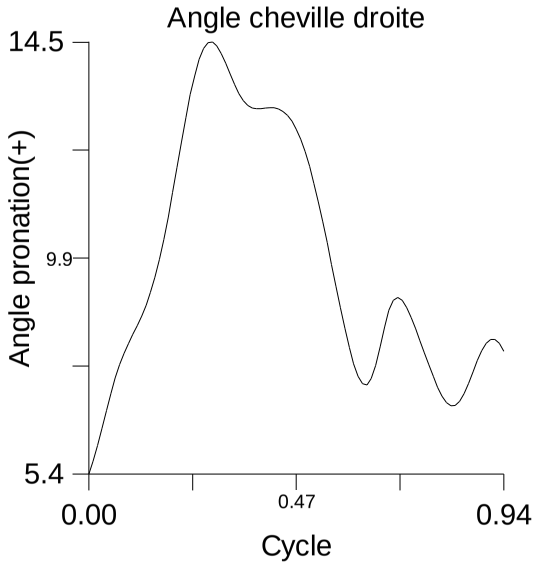


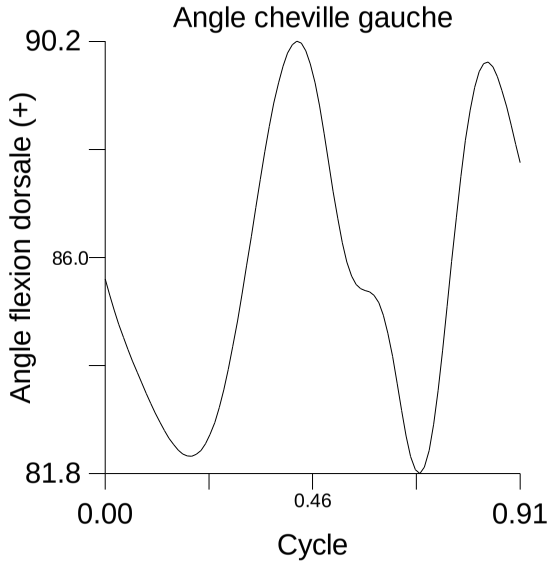


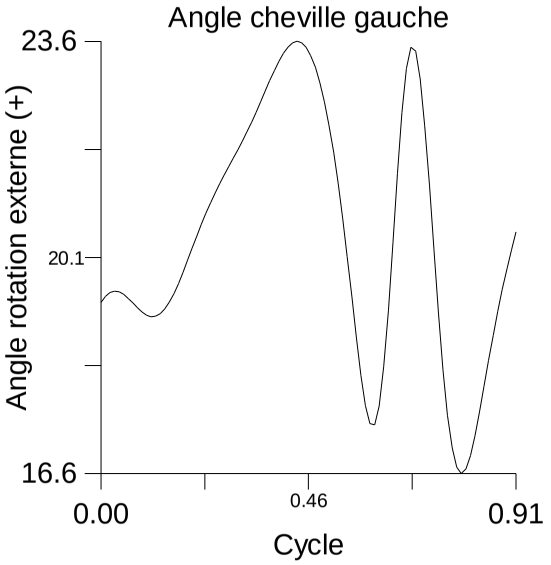
CINEMATIQUES CHEVILLES

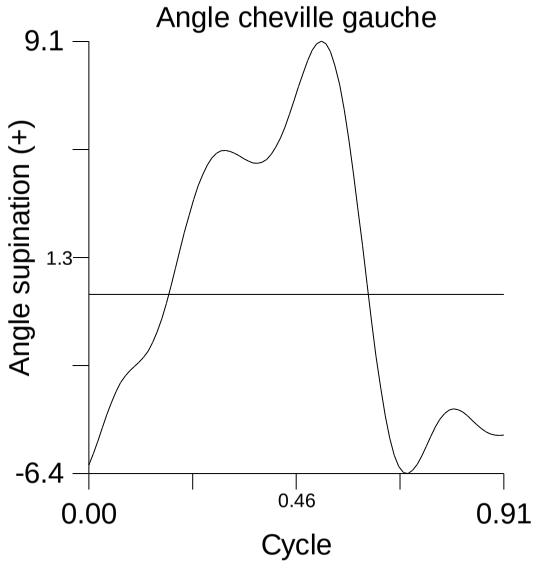




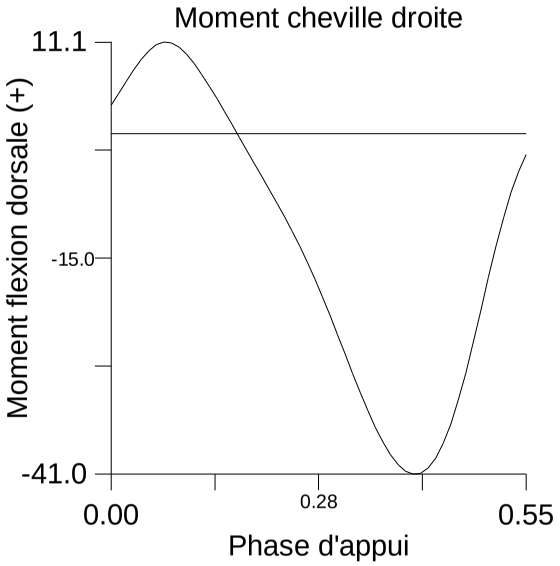


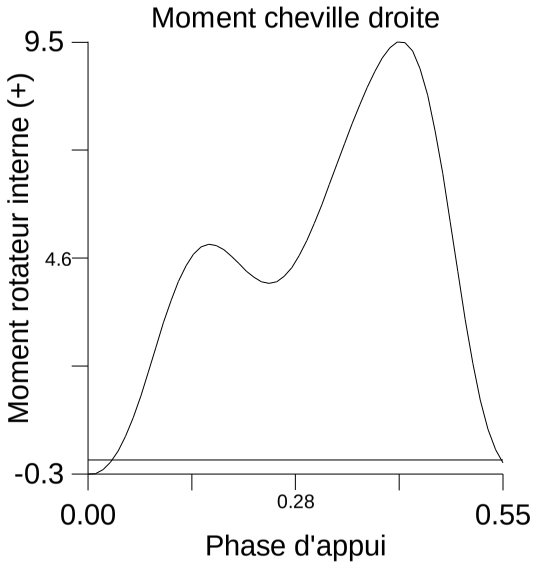


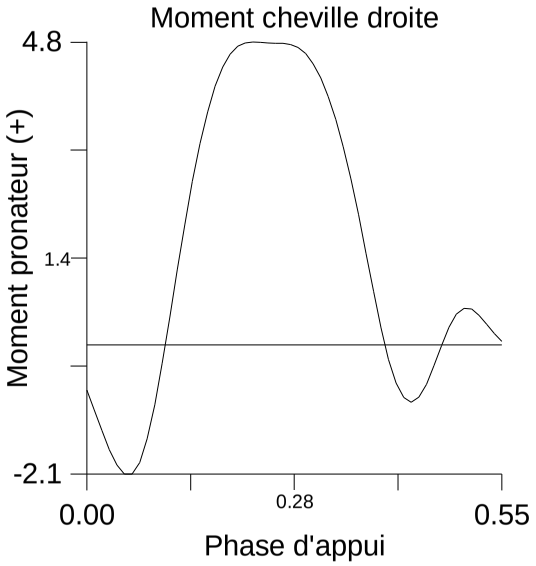


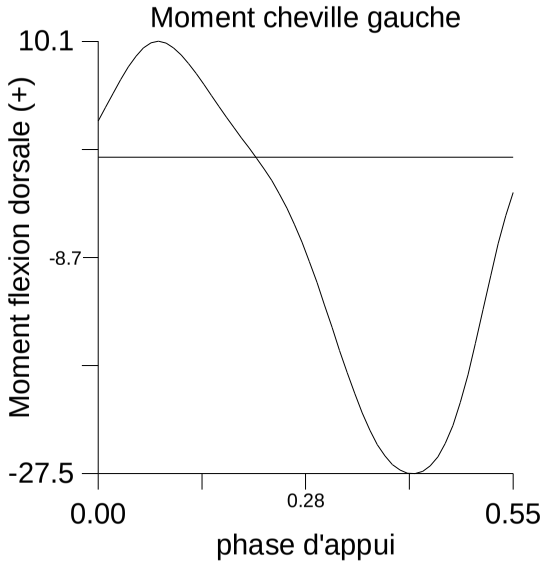


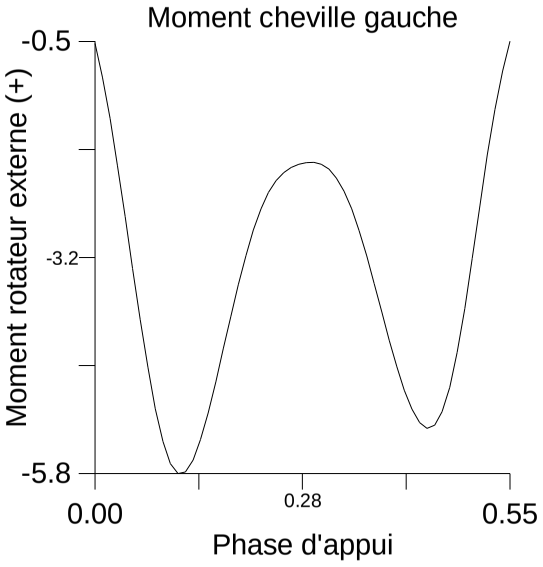
CINETIQUES CHEVILLES

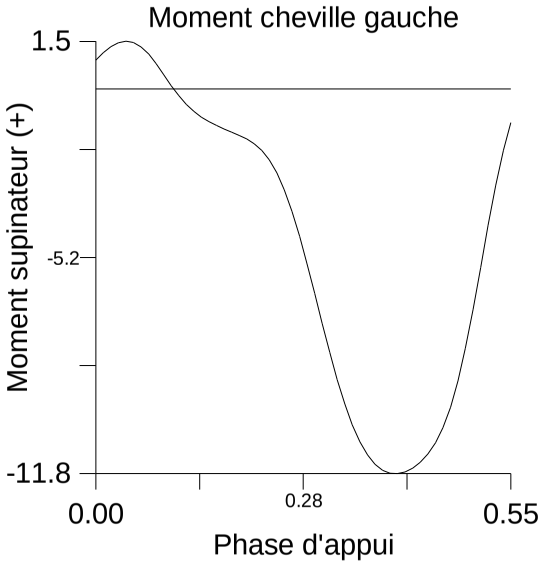




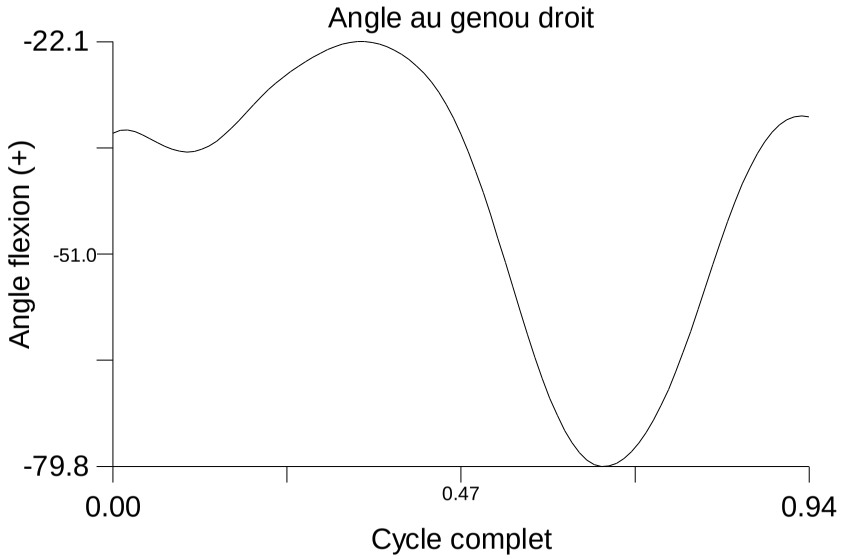


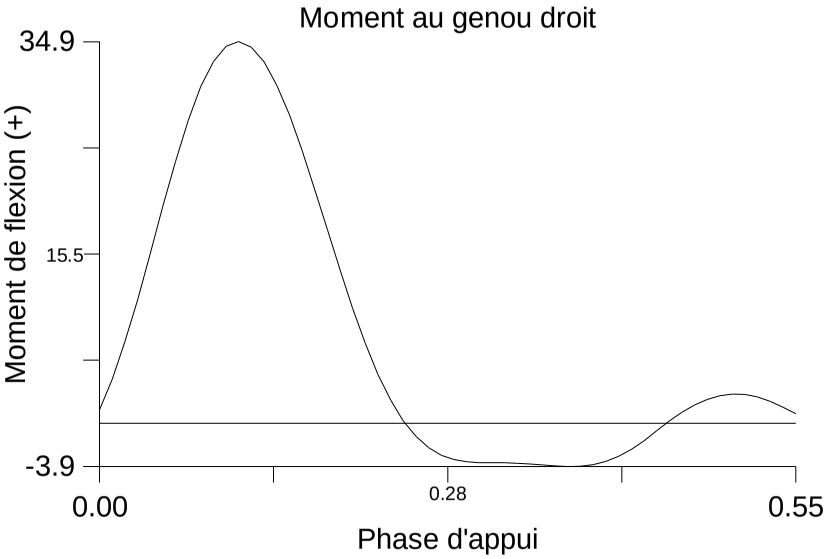


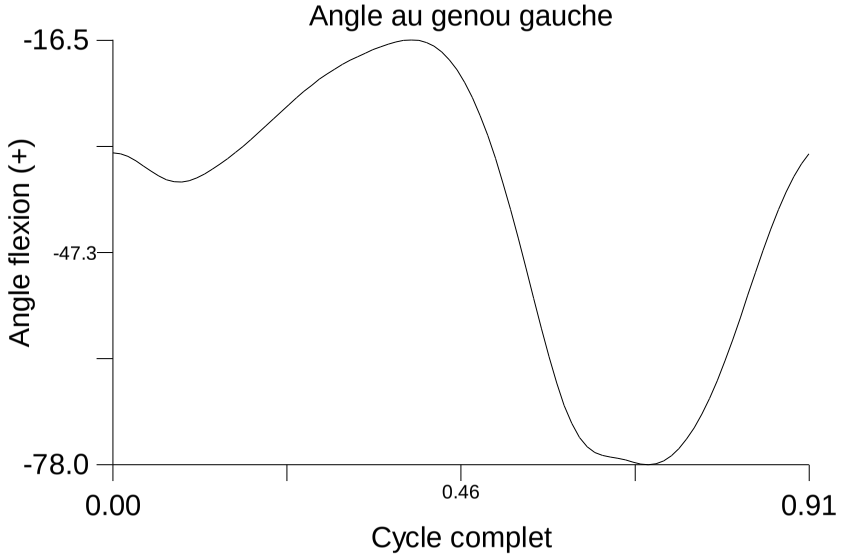


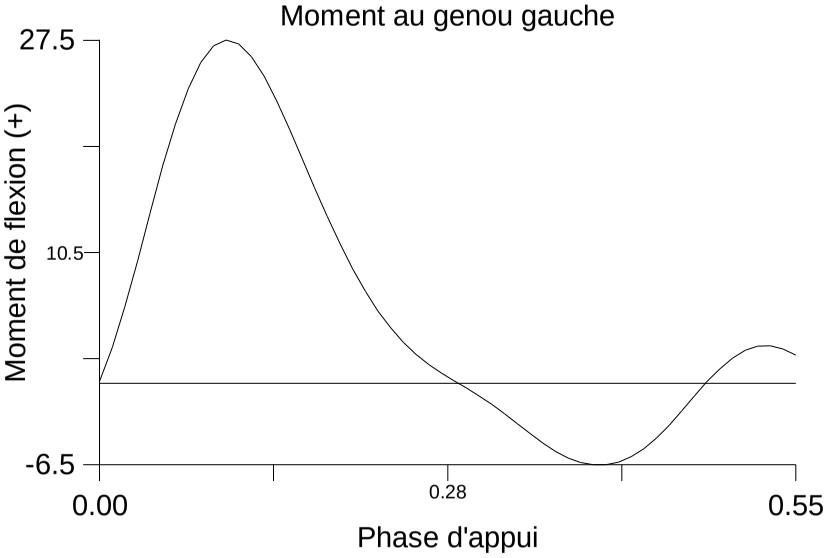


CINEMATIQUES / CINETIQUES GENOUX

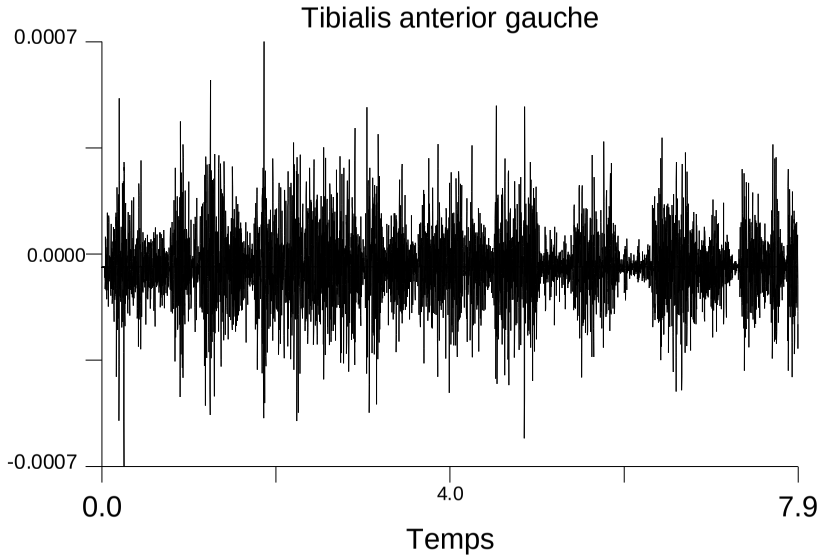


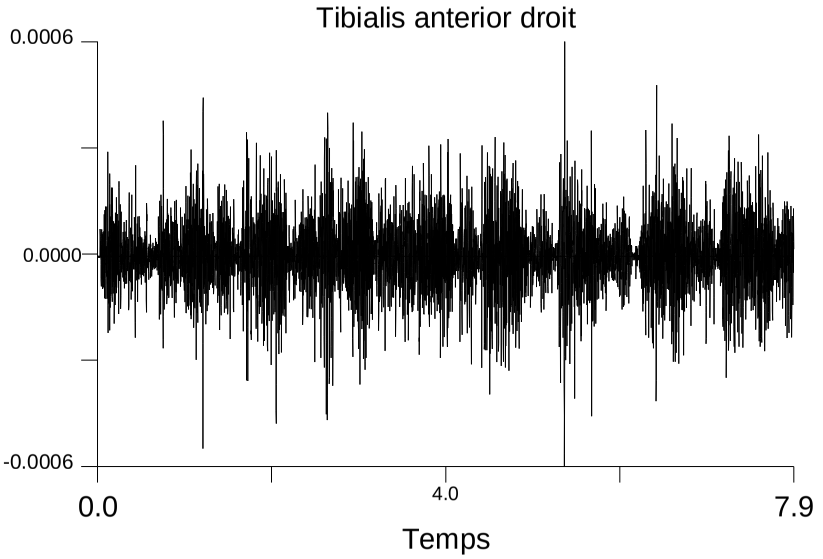


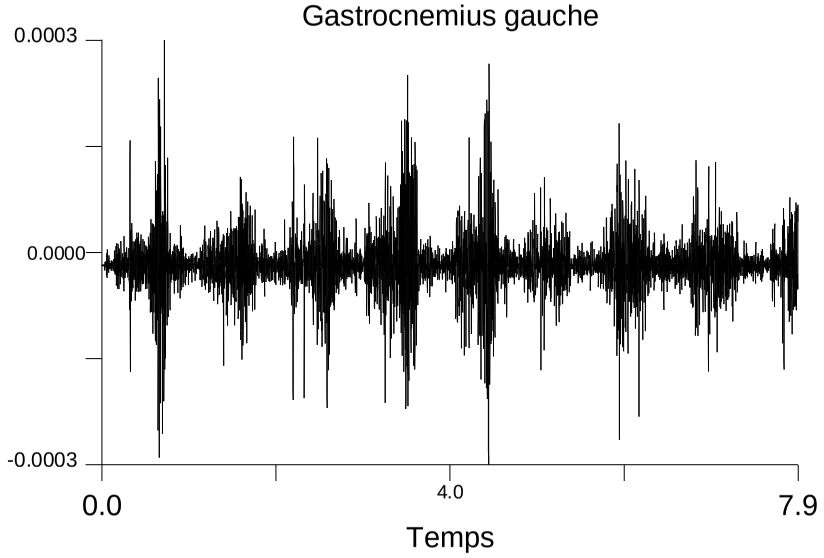


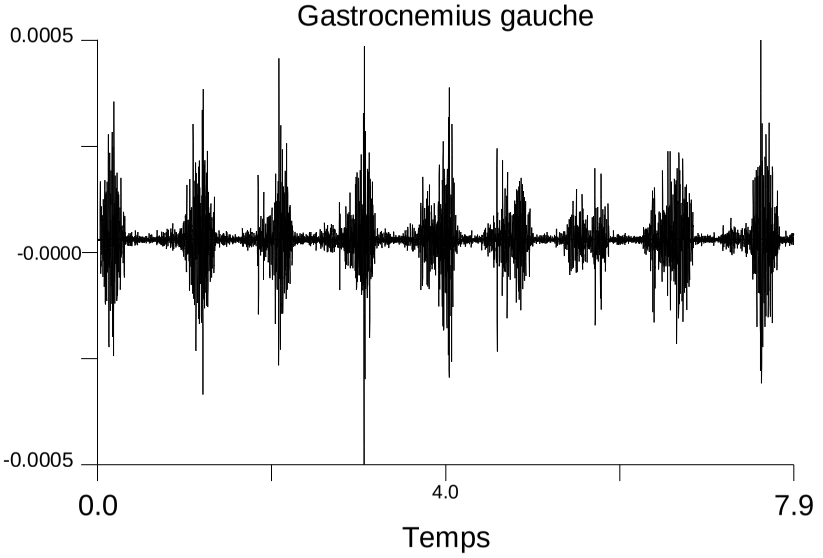


Signaux EMG

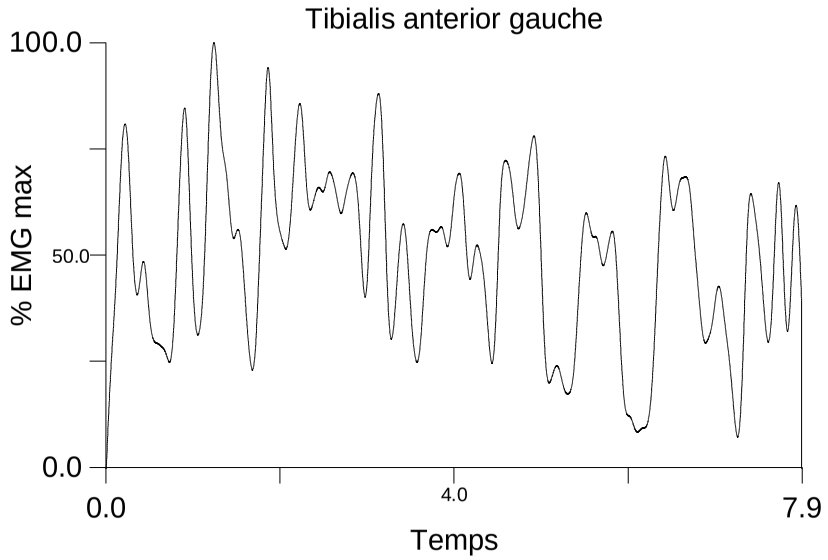

Tibialis anterior gauche
0.0007
0.0000
-0.0007
0.0
4.0
7.9
Temps

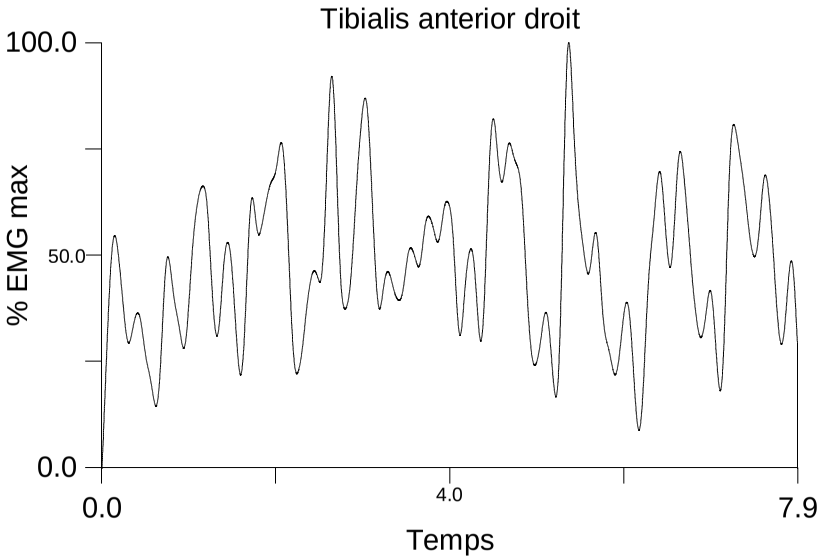

Tibialis anterior droit
0.0006
0.0000
-0.0006
0.0
4.0
7.9
Temps

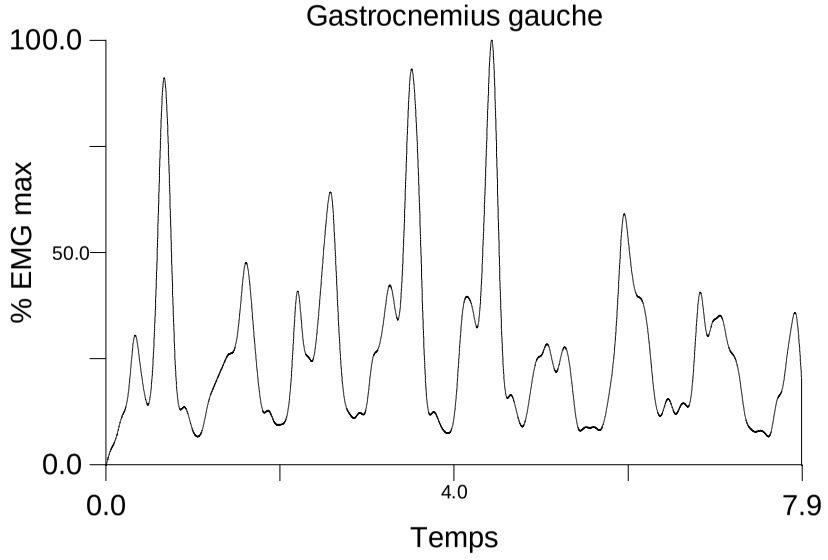

Gastrocnemius gauche
0.0003
0.0000
-0.0003
0.0
4.0
7.9
Temps

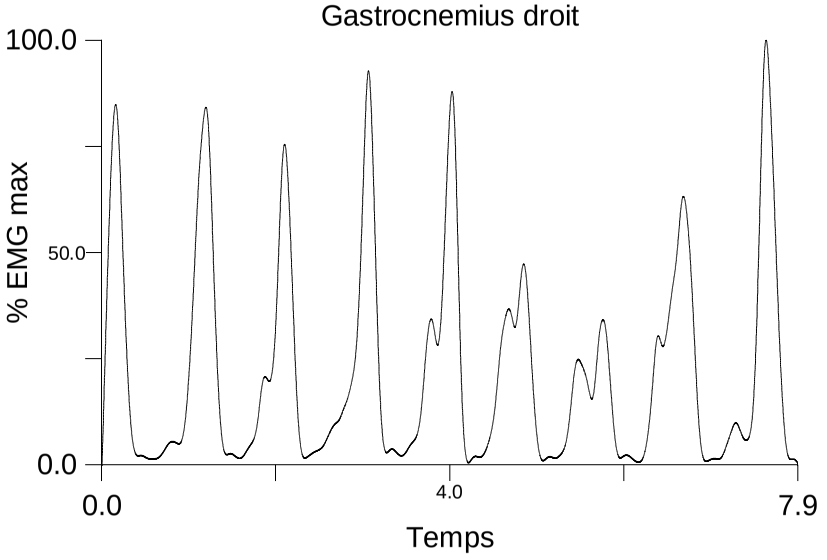

Gastrocnemius gauche
0.0005
-0.0000
-0.0005
0.0
4.0
7.9
Temps

EMG

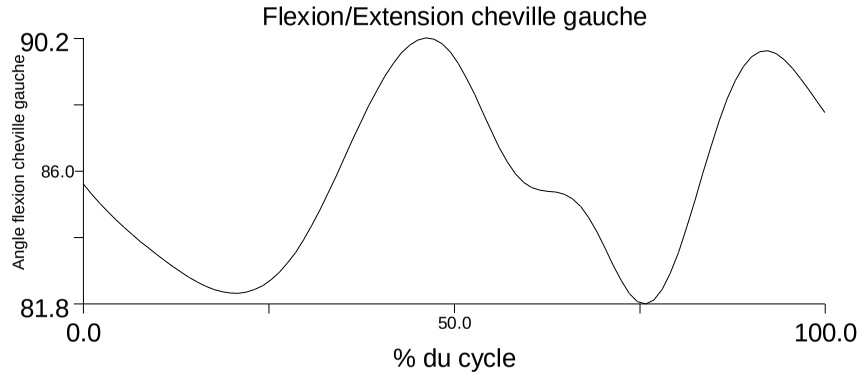

Tibialis anterior gauche
100.0
50.0
0.0
% EMG max
0.0
4.0
7.9
Temps

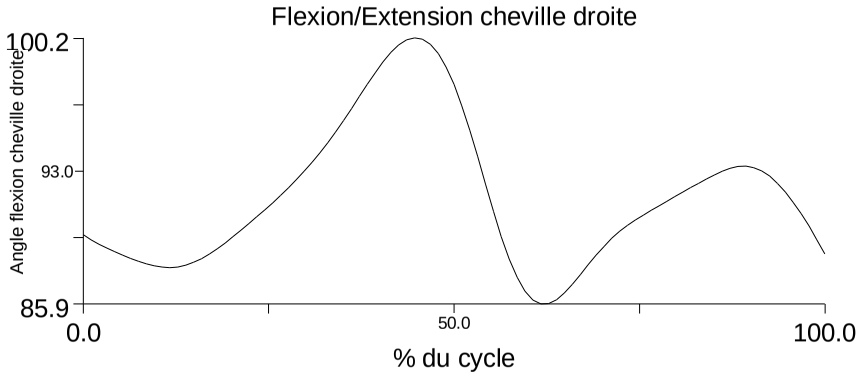

Tibialis anterior droit
100.0
50.0
0.0
% EMG max
0.0
4.0
7.9
Temps

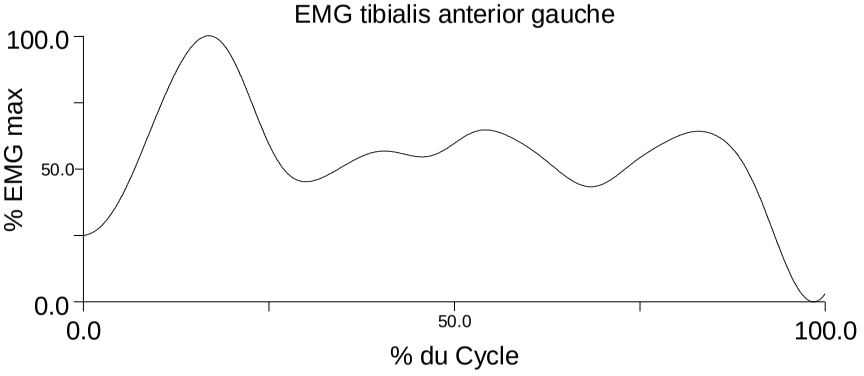

Gastrocnemius gauche
100.0
50.0
0.0
% EMG max
0.0
4.0
7.9
Temps

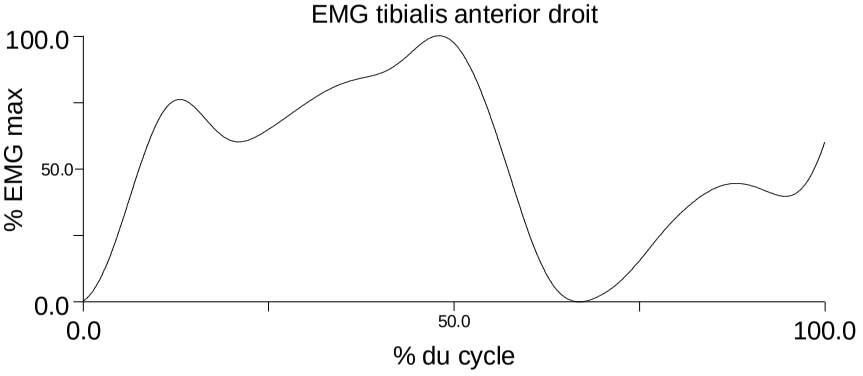

Gastrocnemius droit
100.0
50.0
0.0
% EMG max
0.0
4.0
7.9
Temps

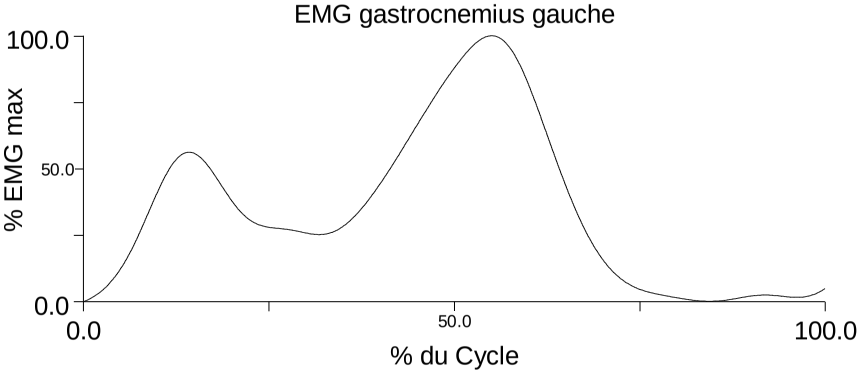
CHEVILLES : ANGLES / EMG

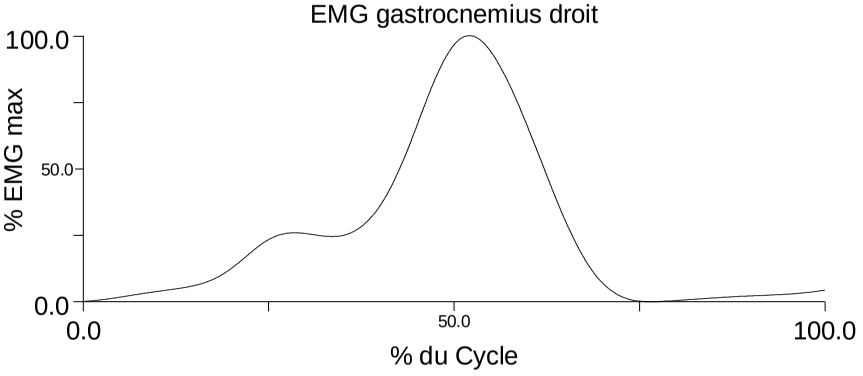




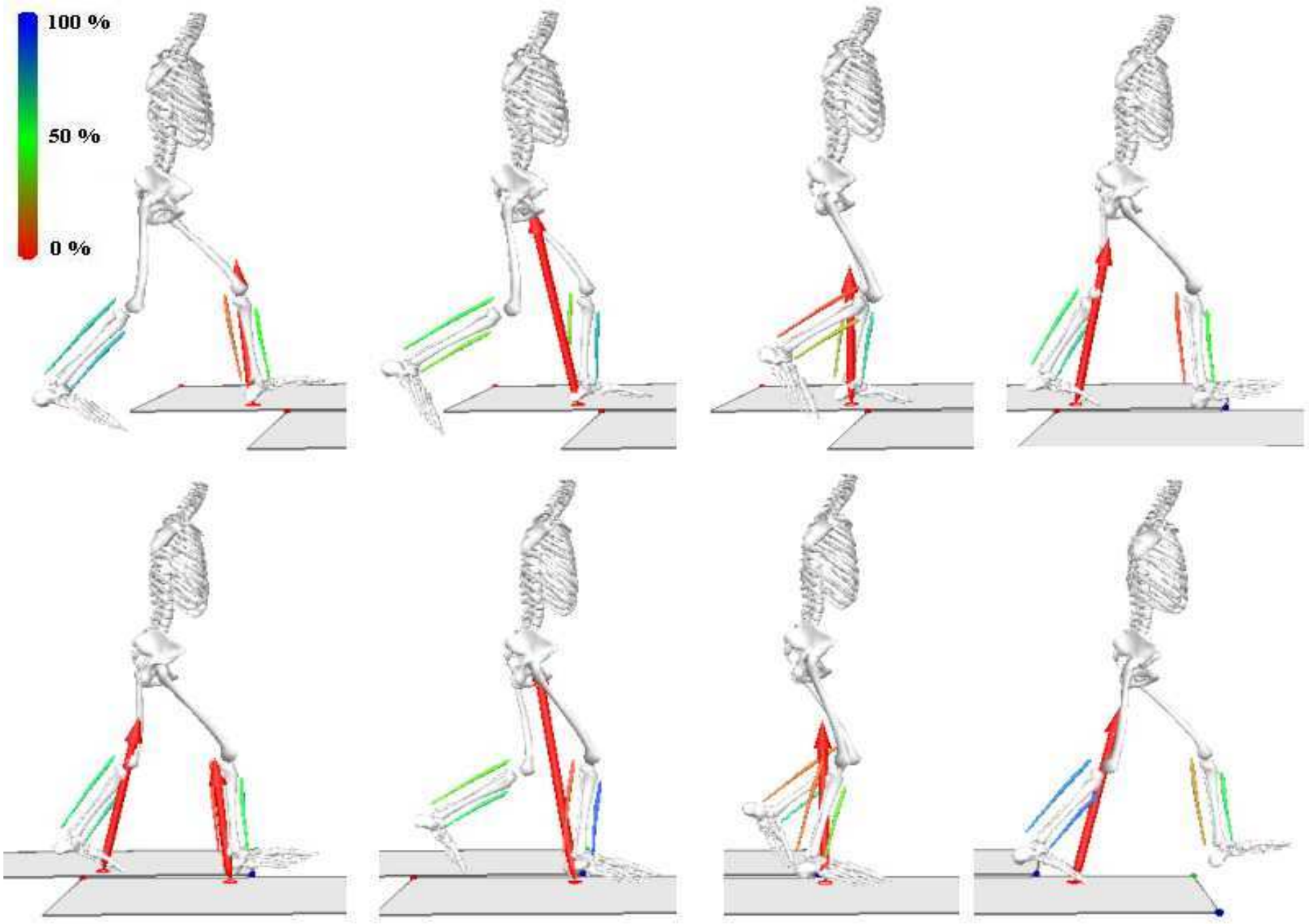








Cycle de marche et EMG



A.3/ Information about the Setup and Data

Data acquired in the experiment with Kinect and Vicon. Data about the Kinect camera placement is summarized in Table A.4, and data about test subjects can be found in A.5.

Table A.4: Aquistion setup for Kinect-MOCAP database

parameter	value
Height of camera placement	0.85 m.
Minimum distance between subject and the camera	1.6 m.
Maximum distance between subject and the camera	4.6 m.
Walking distance	3 m.
Frame rate	30 fps.

Table A.5: Custom dataset presentation

num	Gender	Age	Height, sm	Weight, kg
1	M	29	180	68
2	M	31	178	61

A.4/ Kinect Joints Visualization

Figure A.1 and Figure A.2 show examples of knees and feet joints positions estimated by a Kinect v.2 placed under and angle about 30 degrees to the moving direction.

A.5/ Normal Gait Model Visualization for DAI

Figure A.3 shows the mass distribution of the angle covariance features on the training part of DAI [169] dataset.

A.6/ LSTM Model Details

Figure A.4 shows the LSTM model architecture used in this work, with the input features visualized. Note that the final model is bi-directional, which is not shown on the Figure, however, we

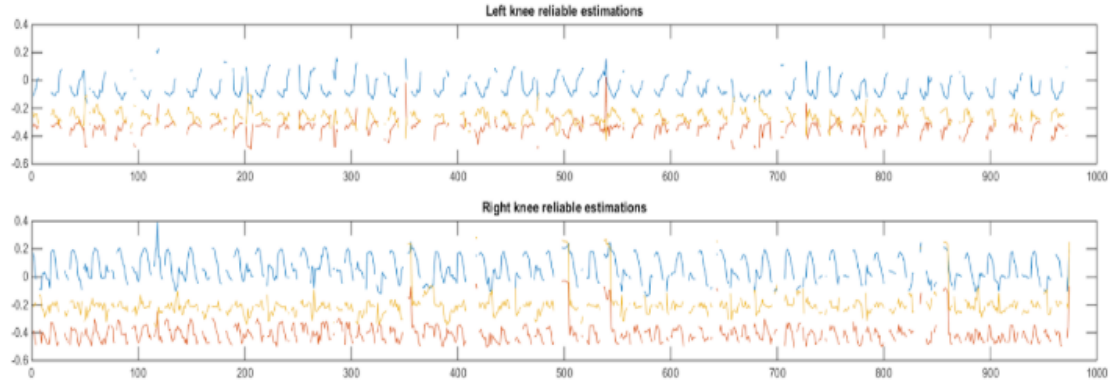


Figure A.1: Visualization of the X(blue), Y(red), Z(yellow) knee joint coordinates dynamics. Frames where status of joints was 'not tracked' are excluded.

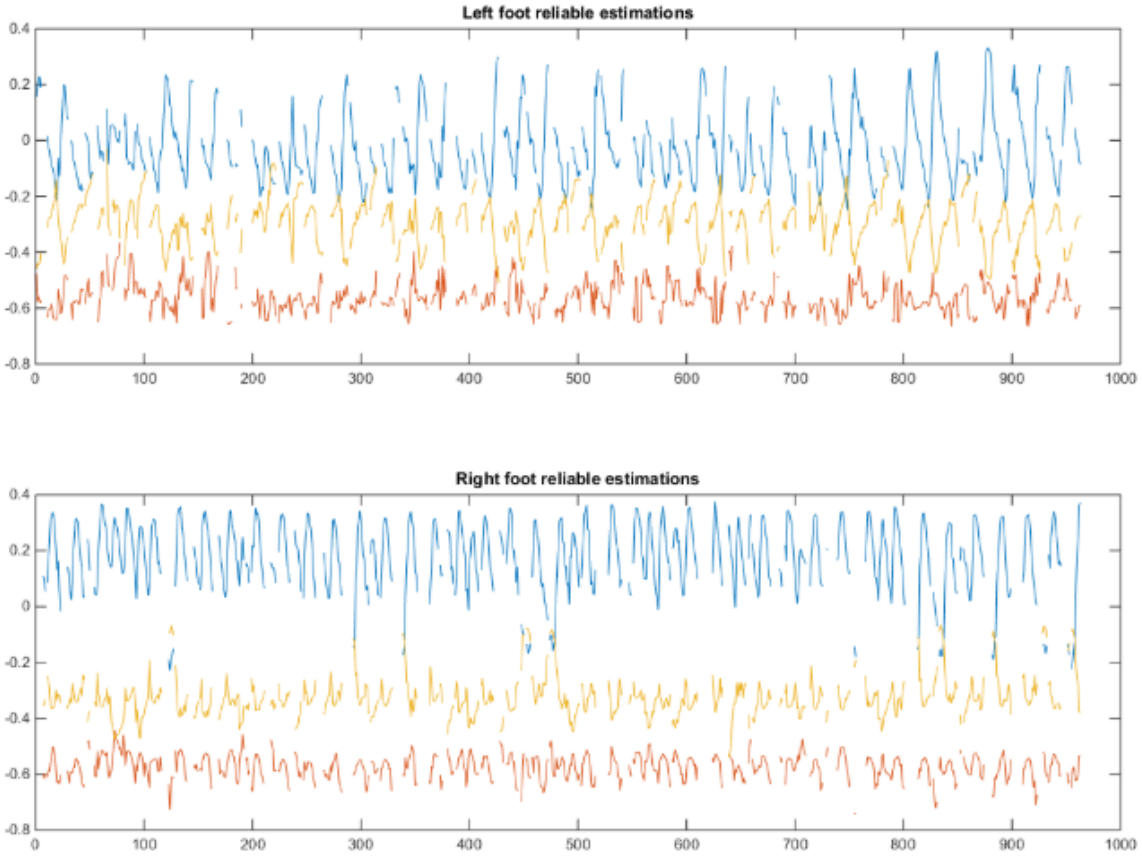


Figure A.2: Visualization of the X(blue), Y(red), Z(yellow) foot joint coordinates dynamics. Frames where status of joints was 'not tracked' are excluded.

experimented with both normal and bi-directional LSTMs in Chapter 5.

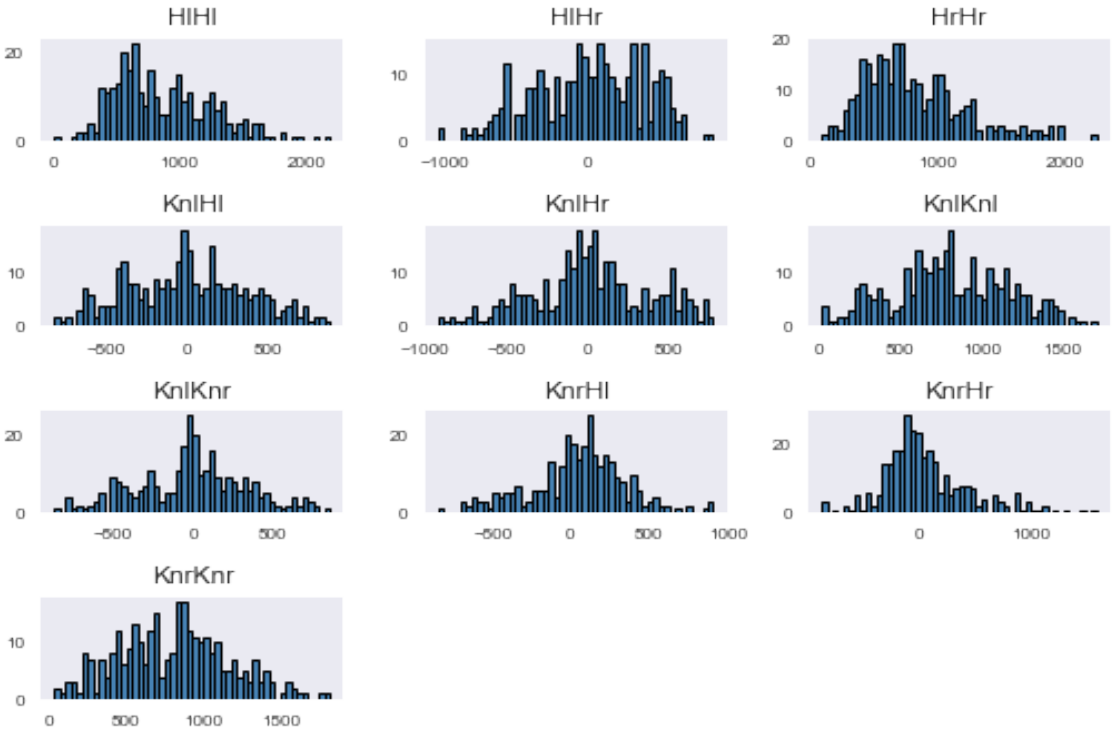


Figure A.3: The mass distribution of the angle covariance features on the training part of DAI follows the normal model.

A.7/ MMGS Database Details

Detailed information about the database participants is summarized here in Table A.6. This can be used in case a need to normalize the data, or perform gender recognition.

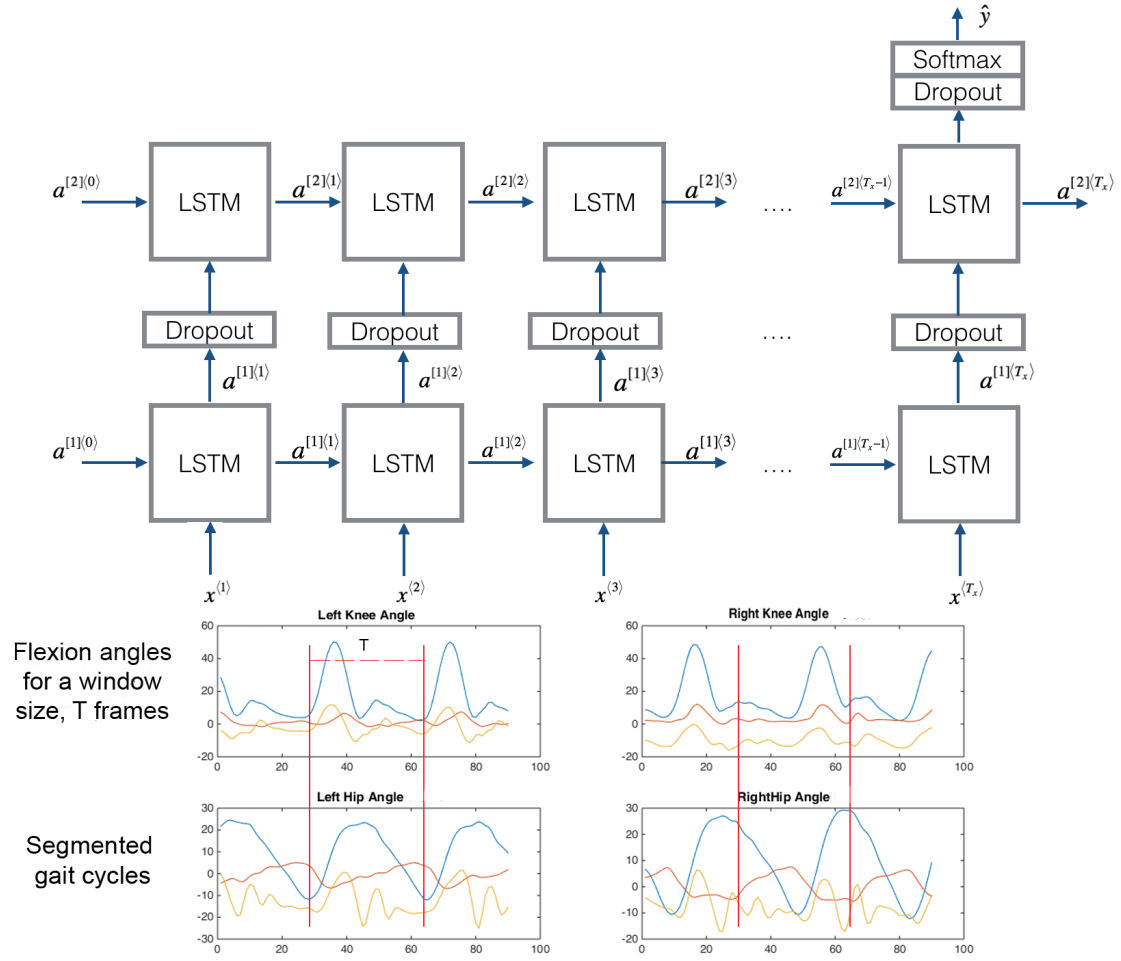


Figure A.4: The LSTM model architecture and input features.

Table A.6: Custom dataset presentation

num	Gender	Age	Height, sm	Weight, kg
1	M	23	182	89
2	M	26	167	67
3	M	34	188	106
4	M	30	180	65
5	M	28	170	73
6	M	31	176	65
7	M	23	170	71
8	F	55	156	58
9	M	42	175	85
10	M	27	170	75
11	M	28	175	70
12	F	28	176	58
13	F	25	171	53
14	M	47	178	70
15	M	25	175	84
16	M	24	181	86
17	M	35	176	65
18	M	29	183	70
19	M	28	180	80
20	M	28	172	65
21	M	34	175	63
22	M	24	180	65