



**Politecnico
di Torino**

POLITECNICO DI TORINO

MASTER'S DEGREE IN BIOMEDICAL ENGINEERING

Master's Degree Thesis

**A comparative evaluation of two
markerless methods for sagittal lower
limb joint kinematics estimation based
on a single RGB-D camera**

Author:

Alba Rago

Supervisors:

Prof. Ing. Andrea Cereatti

Ing. Diletta Balta

July 2022

Abstract

Gait analysis is a valuable and widespread tool for assessing the quality of human locomotion. To date, marker-based optoelectronic stereo-photogrammetry (MB) represents the gold standard for the evaluation of lower limb joint kinematics and the most accurate solution. However, MB systems require specialized operators, dedicated spaces, and long preparation and processing time. Additionally, the presence of markers attached to the subject's body and a cumbersome set-up of many cameras may affect the spontaneity of the movement.

Video-based markerless systems (ML) represent a low-cost, powerful, and promising alternative to MB systems. The use of ML techniques allows making the experimental sessions faster and easier since it does not require the application of markers on the skin of the patients. Most recently, several companies are producing inexpensive tracking consumer electronics systems constituted by an RGB camera integrated with an infrared depth sensor (RGB-D camera). These cameras (e.g. Microsoft Kinect, IntelRealSense D435) often come with software development kits (SDK) for real-time tracking of body position and orientation primarily focused on gaming purposes.

Very recently (2020), a new RGB-D camera (Azure Kinect) was released by Microsoft and compared to the previous versions of Kinect, this camera is targeted towards other markets such as logistics, robotics, health care, and retail. The improved performances suggest the possibility to apply these technologies for the development of clinical-based applications.

Within this general context, this thesis project aims: (i) to investigate whether motion tracking through the body tracking SDK integrated into the Azure Kinect DK could be employed to perform gait analysis for clinical purposes and (ii) to compare the performances of the above-mentioned SDK to an improved custom version of a 2D markerless method (MLM) based on a subject-specific kinematic model developed by Balta et al., 2020.

Before methods comparison, a characterization of the depth sensor performance was performed to evaluate its accuracy and precision (repeatability) in both static and dynamic conditions. A static acquisition was conducted by placing a known-sized object (a box) in front of the camera at four distances (1200, 1820, 2400 and 3000 mm). It was observed that the farther the object, the lower the accuracy of the distance estimation and that the repeatability was lower in correspondence of the object's edges. A dynamic acquisition was performed by using an oscillating pendulum to simulate the swing phase. As result, the faster the object is moving, the higher the number of invalidated pixels (black dots) in the depth images. As reported by the producers, in certain situations the depth sensor may not provide the depth information for all pixels resulting in the decrease of the object's area in the image.

After the above-described sensor characterization, a comparison between the two ML methods was conducted on the gait cycles recorded with the Azure Kinect. The sagittal lower limb joint kinematics was computed and validated against a standard MB gait analysis protocol. Five healthy subjects were recorded in a gait analysis laboratory equipped with the Azure Kinect placed laterally to the walkway and a 3D MB system (Vicon Vero) that was used as the gold standard. The acquisitions could not be synchronized due to the mutual interference of the two systems caused by the same working IR wavelength. This caused a large deterioration in the depth images. Ten gait trials per subject were recorded (5 for left and 5 for right). One single gait cycle for each trial was manually identified from each recording by observing the initial and final contacts of the foreground foot. The lower limb joint centres were estimated with the two methods to compute the hip, knee, and ankle angles in the sagittal planes.

Seven significant gait variables were extracted for each trial: the hip max extension in stance phase (H3), the ankle max dorsiflexion in stance phase (A3) and swing phase (A5), the knee flexion at initial contact (K1), the knee max flexion at loading response (K2), the knee max extension in stance phase (K3), and the knee max flexion in swing phase (K5). Results were averaged over the total number of trials and compared by calculating the mean differences (d) obtained from the SDK and the MLM with respect to the marker-based system. The reliability of the methods was evaluated with the intraclass correlation coefficient (ICC) computed for each variable.

Both the SDK and MLM methods provided good estimates for knee and hip kinematics while higher differences with respect to the marker-based system are provided for ankle kinematics, mainly for the SDK. The resulting waveforms of the ankle angles from the SDK joint centres estimations were substantially different from the marker-based ones, particularly during the stance and swing phase where the lateral malleolus position was visibly misidentified. ML performances were higher in the stance and swing phase of knee kinematics ($d < 0.8^\circ$) while SDK differences were lower during the first 40% of the gait cycle ($d < 1.4^\circ$). Regarding hip kinematics, ML has shown a lower mean difference than the SDK during the stance phase. The overall performances of the SDK are also affected by the fact that in 2 out of 5 subjects the contralateral limb was misidentified instead of the foreground one. Based on the ICC values, ML revealed moderate (0.5 - 0.75) to good (0.75 - 0.9) reliability except for the knee parameters in stance and swing phase and the hip parameter. The SDK showed lower values of ICC with respect to the MLM except for the hip and knee variables in the stance phase.

This preliminary investigation demonstrated that although the MLM method is more time-consuming than the SDK and requires a manual identification on the image of the joint cen-

tre's positions for the creation of the models, its accuracy overcomes the SDK limitations. Nonetheless, it is the author's opinion that the accuracy of the joint centre's estimation could be further improved by implementing a 3D model which could embed the outputs of both the SDK and the MLM method. In future work, these investigations could be conducted also in patients with altered gait functions, e.g., children with cerebral palsy, to study the performances of the methods on unusual gait cycles.

A summary of the work conducted in this thesis project named "*A comparative accuracy study of two markerless methods for estimating the sagittal lower limb joint kinematics with a single RGB-D camera*" has been submitted for the annual conference of the Italian Society of Clinical Movement Analysis (SIAMOC) that will be held in Bari in October 2022.

Acknowledgments

I would like to acknowledge my supervisor professor Cereatti who made this work possible by giving me the opportunity to take part in this research project and his precious guidance. A warm thank you to Ing. Balta whose contribution and advice helped me through all the stages of this work.

Contents

1	OVERVIEW	1
2	BACKGROUND	3
2.1	Clinical relevance and aim of the work	3
3	FUNDAMENTALS OF GAIT ANALYSIS	6
3.1	The gait cycle	6
3.2	Joint kinematics	8
4	MOTION CAPTURE	12
4.1	Marker-based system	12
4.2	Depth camera and markerless systems	14
5	SENSOR CHARACTERIZATION	18
5.1	Azure Kinect DK	18
5.2	Preliminary characterization of depth camera performance	21
5.2.1	Static Recordings	21
5.2.2	Dynamic Recordings	24
6	METHODS FOR JOINT KINEMATIC ESTIMATION	28
6.1	Model-based Markerless Method (MLM)	28
6.2	Body Tracking SDK	40
7	MATERIALS AND METHODS	44
7.1	Experimental Protocol and Setup	44
7.2	Infrared interference Issues	47
7.3	Data Processing	48
8	RESULTS AND DISCUSSIONS	57

9 CONCLUSIONS

66

Acronyms

CCS Cartesian Coordinate System

DoF Degrees of freedom

GT Greater Trochanter

ICC Intraclass correlation coefficient

JCS Joint Coordinate System

LE Lateral femoral Epicondyle

LM Lateral Malleolus

MB Marker-based

ML Marker-less

MLM Markerless Method

NFOV Narrow field-of-view

ToF Time-of-flight

WFOV Wide field-of-view

List of Figures

3.1	Phases of a typical gait cycle. Figure from Pirker et al. 2016 [1]	7
3.2	The Joint Coordinate System (JCS) of the right ankle complex. Figure from Wu et al. 2002 [2]	9
3.3	Illustration of the Joint Coordinate System (left) and the three rotational movements (right) associated to the knee joint. Figure adapted from Grood et Suntay. 1983 [3]	10
3.4	The Joint Coordinate System (JCS) of the hip joint. Figure from Wu et al. 2002 [2]	11
3.5	Mean (solid line) and standard deviation (dotted line) of hip, knee and ankle angles in sagittal plane. All angles are in degrees. Figure adapted from Kadaba et al. 1990 [4]	11
4.1	Illustration of a movement analysis laboratory with stereophotogrammetry system, force plates and the associated reference systems. Figure from Capozzo et al. 2005 [5]	13
4.2	Illustration of the camera set up and the movement detection performed through Theia 3D Markerless. Figure from Theia Markerless	14
4.3	Illustration of the Time-of-Flight principle, implemented by the Azure Kinect and the Kinect 2	15
4.4	Example of pose estimation based on a Convolutional Neural Network (CNN). First, the CNN was trained through supervised learning with a set of manually labelled images (A) then the trained network is capable of processing and labelling new images (B). Figure from Cronin et al. 2021 [6]	16
4.5	(A) Comparison of 3D pose estimation outputs of the marker-based (Mocap) and the OpenPose markerless method (OpenPose) during walking; (B) Examples of 2D estimations success and failure. Figure adapted from [7].	17
5.1	Illustration of the Azure Kinect and the output images from the Azure Kinect Viewer. Figures adapted from Azure Kinect Viewe	18

5.2	Examples of depth images taken with narrow field-of-view (left) and wide field-of-view (right) mode.	19
5.3	Illustration showing the effect on depth pixels of a saturated IR signal. The arrows point to the invalidated pixels in both the depth (left) and IR (right) images. Figure from [8].	20
5.4	Illustration showing the effect on depth pixels of an IR signal which is too low. The arrows point to the invalidated pixels in both the depth (left) and IR (right) images. Figure from [8].	20
5.5	a) Average depth pixels values over 126 frames of static acquisitions at a fixed distance. The mean distance is represented by the dotted line. b) Errors with respect to the object's real distance from the camera, placed at different positions ($D1 = 1200$ mm , $D2 = 1820$ mm, $D3 = 2400$ mm, $D4 = 3000$ mm)	22
5.6	Illustration of the standard deviation map calculation. The standard deviation was computed over each pixel for 126 frames of static acquisitions at fixed distance	22
5.7	Color image (top) and standard deviation map (bottom) of the known-sized object placed at a fixed distance of 2.6 m.	23
5.8	From left to right: RGB, Depth image and segmentation mask of the oscillating pendulum.	24
5.9	a) Identification of the lower extremity of the pendulum through an ellipse which was fitted on the bar. The image reference system is on the top left of the image. b) The reference system of the image is aligned to the inertial principal axes of the pendulum for the identification of the coordinates of the extreme point of the bar (E).	25
5.10	Speed (blue) and pendulum area (orange) for each frame of the swinging motion. The figure on the right is a zoom of the figure on the left.	25
5.11	Column and rows coordinates of the pendulum extremity. The dotted lines represent the 17 selected frames in which the pendulum was in a vertical position	26
5.12	Standard deviation computed over each pixel for 17 frames of dynamic acquisition of the pendulum in a vertical position	26
5.13	Illustration of invalidated pixels of the shank (left) and thigh (right) during the preliminary recordings of a gait cycle in the sagittal plane. The invalidated pixels mainly occur along the edges of the subject body and in the direction of progression	27

6.1	Block diagram of the subject segmentation step of the markerless method for the estimation of the lower limb joint kinematics.	31
6.2	Illustration of the three subject-specific lower limb models in a static up-standing position, during the loading phase and the swing phases.	32
6.3	Illustration of the lower limb model structure with the foot, shank and thigh models and their coordinate systems.	33
6.4	Overview of the bottom-up approach implemented for the lower limb body segment pose estimation.	34
6.5	Shank region and lateral malleolus (LM) position identification in a frame in which the foreground and the background shank are separated (right) and overlapped (left).	35
6.6	Envelope of the depth values histogram computed for the semicircular ROI centred in the LM of the foreground ankle with a radius equal to the LE-LM segment. The foreground shank is identified through the range of depth values between the two black dashed lines.	35
6.7	Thigh identification in an image where the hand, foreground thigh and background thigh are overlapped. RGB image (top left), thigh region identification and LE position (top right). Envelope of the depth values histogram computed for the semicircular ROI centred in the LE of the foreground knee with a radius equal to the LE-GT segment (bottom). The foreground thigh range is identified between the light blue and the black dashed lines.	37
6.8	Illustration of the hip, knee and ankle angles of the foreground limb calculated starting from the GT, LE and LM 2D coordinates and the foot inclination (T) extracted through the MLM method	39
6.9	Illustration of real-time body tracking performed with the Azure Kinect SDK. Figure from: Microsoft Azure Kinect DK documentation	40
6.10	Illustration of the Body Tracking SDK system architecture. Figure from: 3D Skeletal Tracking on Azure Kinect	41
6.11	Illustration of the joint positions and connections provided by the Azure Kinect SDK. The rectangular box includes the joints under investigation in this study. Figure adapted from: Microsoft Azure Kinect DK documentation	42
6.12	Illustration of the information contained in a body structure. Figure adapted from the Microsoft Azure Kinect DK documentation	43
7.1	Illustration of the experimental set up	45
7.2	Active Wand for the optical stereo-photogrammetric system calibration	45

7.3	Illustration of the marker placement for the acquisitions with the Vicon system. Figures adapted from Vicon - lower boy plug-in-gait	46
7.4	Illustration of the RGB (left) and the depth (right) images captured with the Azure Kinect during synchronous recording with the marker-based system. Black invalidated pixels can be seen particularly in correspondence of the reflective markers (red arrows).	47
7.5	Overview of the methods.	48
7.6	Illustration of the recording and processing procedure adopted for the extraction of the joint angles from the marker-based recordings.	50
7.7	Example of a left trial recording with the marker-based system. Each marker relative to the anatomical landmarks is labelled according to the plug-in gait lower limb model.	51
7.8	Example of the plug-in-gait modelling outcome with the Vicon-Nexus system.	51
7.9	Examples of foot-floor contact in the intial contact (IC) and final contact (FC) frame selected for the identification of a single gait cycle.	52
7.10	Examples of frames selected for the load model, flex (or swing) and static model definition for one of the right limb trials.	52
7.11	Illustration of the hip, knee and ankle angles of the foreground limb calculated starting from the HIP, KNEE, ANKLE and FOOT 2D coordinates extracted thourgh the SDK method	55
7.12	Illustration of the seven gait variable extracted from the hip, knee and ankle angles in the sagittal plane for the comparison of the methods.	56
8.1	Illustration of the joint positions resulting estimations trough the MLM method for some frames of a right trial.	57
8.2	Illustration of the joint positions resulting estimations extracted through the SDK method for some frames of a right trial.	58
8.3	Example of wrong leg identification with the body tracking SDK method. In this frame the supposed left hip, knee, ankle and foot coordinates were misidentified as the right limb ones.	58
8.4	Illustration of hip, knee and ankle angles in the sagittal planes of ten gait trials of a single subject estimated with the marker-based (MB), model-based (MLM) and body tracking SDK methods. While the markerless angles result from synchronous recordings the MB angles were extracted from different recordings of the same subject with markers attached to the skin. . .	62

LIST OF FIGURES

8.5 Five summary statistics for the seven gait variables extracted from the joint angles computed with the marker-based (MB), model-based markerless method (MLM) and body tracking SDK computed for each gait variable. 63

8.6 Examples of ankle angle evaluations from the coordinates extracted with the SDK method in two different phases of the gait cycle, during the stance (left) and swing (right) phase. The coordinates of the ankle, foot and knee are visibly misidentified resulting in incorrect ankle angle estimations. . . . 64

8.7 Illustration of ankle angle parameters estimation from the dorsiflexion angles extracted with the marker-based (MB), model-based (MLM) and body tracking SDK methods. The SDK ankle curve results in incorrect estimations of the ankle max flexion in the stance phase (A3) and swing phase (A5). 64

List of Tables

5.1	Real distance of the known-sized object, mean distances computed from the Azure Kinect depth images and their differences.	21
8.1	Reliability of the model-based markerless (MLM) and body tracking SDK methods computed for each gait variable	59
8.2	Grand mean and standard deviation of the gait parameters over ten trials per five subjects computed for the marker-based (MB), model-based markerless (MLM) and body tracking SDK methods. Differences of the two markerless methods estimations with respect to the MB method.	61

CHAPTER 1

OVERVIEW

Chapter 1 is the current section which includes an overview of the work.

Chapter 2 describes the clinical relevance of gait analysis and the aim of this thesis project.

Chapter 3 is an introductory section about the methods of gait analysis. The gait cycle and its phases are presented going into the details of the lower limb joint kinematics and their evaluation.

Chapter 4 describes the methods for motion capture focusing on the optical methods which can use either a marker-based or markerless approach. An example of a 3D marker-based system, the optoelectronic stereophotogrammetry and its limitations is provided and then an overview of the potentiality of 3D and 2D markerless systems and some of their applications in gait analysis are presented focusing on the 2D markerless gait analysis protocol that will be under investigation.

Chapter 5 is dedicated to the description of the hardware (Azure Kinect DK) and to the preliminary investigations that were conducted to evaluate the depth sensor accuracy and repeatability both in static and dynamic acquisitions.

Chapter 6 In this chapter the two methods for the estimation of lower limb joint kinematics are described in detail. The first one includes the body tracking SDK developed by Microsoft, which is based on a deep-learning algorithm for pose estimation. The second one was developed by Balta et al. (2020) to estimate the lower limb kinematics by defining three subject-specific kinematic models.

Chapter 7 describes the core of this study. The experimental setup and the acquisitions protocol are described in detail. The issue related to the problem of interference between the marker-based system and the Azure Kinect camera is shown and then the

estimation of the lower limb joint kinematics is described in detail, focusing on the gait variables that were selected for the comparison study.

Chapter 8 This chapter presents and discusses the results of this study, focusing on the potentialities and limitations of the methods under investigation.

Chapter 9 This chapter contains the conclusions of this thesis project, the major highlights of the work, its limitations, and a description of future works.

BACKGROUND

2.1 Clinical relevance and aim of the work

Walking is the most frequent activity in everyday life and is strictly related to a person's health. The assessment of gait through its instrumented measurement and the estimation of gait parameters has well-established importance in clinical practice and biomechanical research. Gait analysis allows you to easily analyze lower limb functionality, the efficacy of a clinical therapy, or training sessions. It is especially useful in assessing individuals with cerebral palsy [9], elderly people who are more prone to falling, and patients who have had lower limb orthopaedic surgery or other related injuries that compromise normal locomotion.

In gait analysis, the evaluation of lower limb joint kinematics is a useful and widespread tool for assessing the conditions of human locomotion. To date, marker-based (MB) systems represent the gold standard for the evaluation of lower limb joint kinematics thanks to their accuracy [10]. MB systems use multiple cameras and image processing techniques to detect the instantaneous 3D position of markers placed on anatomical landmarks of interest. Although MB systems have some disadvantages, they require specialized operators, dedicated spaces, and long preparation and processing time, the marker placement may be subjected to human errors and additionally the presence of markers attached to the subject's body and a cumbersome set-up of many cameras may affect the spontaneity of the movement [11]. The need for an accurate, affordable, and time-saving technology is fulfilled by video-based markerless (ML) methods which are recently gaining momentum in clinical gait analysis. ML systems represent a low-cost, powerful, and promising alternative to MB systems. The use of ML techniques allows making the experimental sessions faster and easier since it does not require the application of markers on the skin of the patients.

After a literature review, an insight was gained into video ML systems for gait analysis

which includes both single and multi-camera approaches and focuses either on 3D or 2D analysis. The use of a single camera is preferable as it is both space and cost-saving. Since the movement involved during walking could be mainly evaluated in the sagittal plane, a 2-dimensional approach could be considered sufficient [12]. By using only a single RGB camera the problems of occlusion during movements are increased leading to increased errors in the joint centre's identifications. One more promising technology is the use of RGB-Depth cameras which include an infrared sensor for distance estimation.

Most recently, several companies are producing inexpensive tracking systems constituted by RGB-D cameras (e.g. Microsoft Kinect, IntelRealSense D435) that come with software development kits (SDK) for the real-time tracking of body position and orientation. These technologies were primarily focused on gaming purposes. Recently (2020), Microsoft released a new RGB-D camera (Azure Kinect [13]) which compared to the previous versions of Kinect, is targeted toward other markets such as logistics, robotics, health care, and retail. The Azure Kinect includes an infra-red (IR) sensor for distance estimation that has greater accuracy than its predecessors and a novel motion tracking algorithm (body tracking SDK) for the estimation of the body joints' 3D positions and orientations which is based on deep neural networks.

In this context, this thesis study has two main purposes:

- To investigate whether motion tracking through the body tracking SDK integrated into the Azure Kinect DK could be employed to perform gait analysis for clinical purposes
- To compare the performances of the above-mentioned SDK to an improved custom version of a 2D markerless method (MLM) based on a subject-specific kinematic model developed by Balta et al., 2020 [14].

The first part of this thesis is a general introduction to the gait analysis focused on the lower limb joint kinematics and the optical methods for the assessment of gait. The SP system, which has been used in the experimental investigations of this work is then described. In addition, a literature review on the markerless approach in gait analysis is conducted, paying attention to the RGB-Depth sensor technology and then describing the main hardware component of this work which is the Microsoft Azure Kinect. The experimental work is organized into three sections.

1. The first section is dedicated to the Azure Kinect's depth sensor characterization. Preliminary investigations were conducted both in static and dynamic acquisition to assess the depth sensor accuracy and repeatability and to verify the performances reported in the hardware specifications.

2. In the second section the methods for the hip, knee and ankle angles estimation in the sagittal planes are presented. The first method is based on the Azure Kinect Body tracking SDK joint position estimation while the second method is an improved version of the algorithm implemented by Balta et al. (2020).
3. In the third section the experimental acquisitions and the extraction of the gait variables of interest are described.

The sagittal lower limb joint kinematics was computed and validated against a standard MB gait analysis protocol. Ten gait cycles of five healthy subjects were recorded in a gait analysis laboratory equipped with the Azure Kinect placed laterally to the walkway and a 3D MB system (Vicon Vero) that was used as the gold standard. The acquisitions could not be synchronized due to the deterioration in the depth images caused by the mutual interference of the two systems which had the same working IR wavelength (850 nm).

To evaluate the performances of the methods a single gait cycle was manually identified from each recording by observing the initial and final contacts of the foreground foot. The lower limb joint centres were estimated with the three methods to compute the hip, knee, and ankle angles in the sagittal planes. Seven significant gait variables were extracted for each trial: the hip max extension in stance phase (H3), the ankle max dorsiflexion in stance phase (A3) and swing phase (A5), the knee flexion at initial contact (K1), the knee max flexion at loading response (K2), the knee max extension in stance phase (K3), and the knee max flexion in swing phase (K5). Results were averaged over the total number of trials and compared by calculating the mean differences obtained from the SDK and the MLM with respect to the MB system. The reliability of the methods was evaluated with the intraclass correlation coefficient (ICC) computed for each variable.

In the last part of this work, the performances of the MLM and SDK methods are outlined and compared. The limitations of both methods are analyzed and a possible future development that paves the way to a further improvement in a ML gait analysis protocol for the sagittal lower limb kinematic estimation is formulated.

FUNDAMENTALS OF GAIT ANALYSIS

3.1 The gait cycle

The gait cycle is the functional unit of gait analysis, it is defined as the time interval between two subsequent contacts of the same foot and can be divided into the stance and the swing period. The stance period begins with the initial contact of the foot, also called heel-strike, and is followed by the swing period when the foot is raised above the ground for the advancement of the limb. During the stance phase, which holds approximately 60% of the entire gait cycle, the floor is in contact with the ground. The swing period covers the remaining 40% of the gait cycle, it begins with the toe-off event and ends with the subsequent contact of the foot. The stance and the swing period of the gait cycle can be divided into different phases, each with a precise functional purpose and a specific movement scheme to achieve it. One gait cycle is composed of 8 different phases (Figure 3.1) that follow one another to perform three main tasks: weight acceptance, the single support and, during the swing period, the advancement of the limb [15].

Weight Acceptance: The first task of the stance phase is weight acceptance, during which the shock of the foot that strikes the ground is absorbed and the limb is stabilized under the body weight to preserve the progression. This task includes two phases of the gait cycle:

- Initial contact (0-2% of the gait cycle): The initial contact begins with the heel's strike in a normal gait cycle. The hip is flexed, the knee is extended and the ankle is near to its neutral position.
- Loading response (0-10%): The loading response represents the beginning of the double-support, during this phase the weight of the body is shifted on the advancing limb. The

heel rolls and the knee is flexed to absorb the impact of the foot.

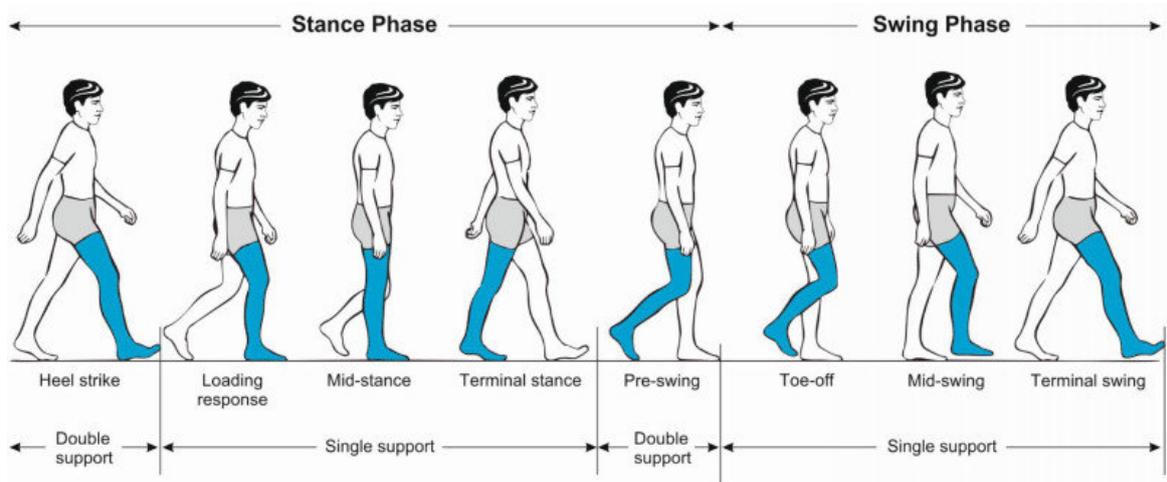


Figure 3.1: Phases of a typical gait cycle. Figure from Pirker et al. 2016 [1]

Single-limb support: The single-limb support begins when the contralateral limb is raised above the ground to start the swing phase and ends when the contralateral foot touches the ground. During this period, the supporting limb holds the weight of the whole body to allow the progression of gait. The single support consists of two phases:

- **Mid stance (10-30%):** It begins when the contralateral foot is lifted off the ground and continues until the weight of the body is aligned on the forefoot. In this phase, the ankle is dorsiflexed, and the hip and knee are extended to move the limb forward the supporting foot.
- **Terminal Stance (30-50%):** It begins with the heel lift and ends when the contralateral foot touches the ground. During this phase, the knee and hip extend further and then the knee is slightly flexed. The progression of the body continues beyond the supporting foot.

Advancement of the limb: For the advancement of the limb, which occurs primarily during the swing phase, the limb must be properly positioned at the end of the support phase. The limb advancement task covers the last four phases of the gait cycle:

- **Pre-swing (50-60%):** It is the final phase of the stance period. The weight is transferred to the contralateral limb. The ankle and knee increase the flexion and the hip extension is reduced in preparation for the swing.
- **Initial Swing (60-73%):** It begins with the toe-off event when the foot is lifted off the ground. The hip is flexed, and the knee reaches the maximum flexion to allow the advancement of the limb, the ankle is only partially flexed.

- Mid-Swing (73-87%): Through an additional flexion of the hip, the limb advancement crosses the gravity line. The knee begins its extension, and the ankle continues the dorsiflexion.
- Terminal Swing (87-100%): During the stride's last phase, the limb advancement is completed. The knee is extended, the hip is flexed, and the ankle is dorsiflexed. The limb is now decelerated and prepared for weight acceptance.

3.2 Joint kinematics

The term *kinematics* refers to the study of the geometric and time-dependent aspects of motion without analyzing the forces causing the motion [16]. In a subject's movement, this refers to the evaluation of the position, velocity, and acceleration of the segments of the body and its joints. In gait analysis, the evaluation of joint angles, meaning the angles between two segments of the body in a certain plane, is a useful and widespread tool for assessing the conditions of human locomotion. Three joint angles that are commonly studied to describe the body motion during gait are the hip, knee and ankle angles in the sagittal plane. When reporting joints motion it is recommended by the Standardization and Terminology Committee (STC) of the International Society of Biomechanics (ISB) to adopt the Joint Coordinate System (JCS) provided by Grood and Suntay in 1983 for the knee joint [3] and the resulting approach reported by the same committee for the hip and ankle joints [2]. The three-dimensional motion of joints can be described by six components or degrees of freedom (DoF). The DoF description is based on the JCS and includes three translations and three rotations. According to Grood and Suntay (1983), the JCS is defined starting from two Cartesian Coordinate Systems (CCS) associated to the two adjacent body segments of the joint. The origins of these two CCS coincide with the linear translation reference point of the joint. The JCS contains two body-fixed axes e_1 and e_2 and one 'floating' axes e_3 which is mutual perpendicular to them.

The Ankle joint The ankle joint complex is composed of the talocrural and the subtalar joints. The talocrural is the articulation connecting the talus and the tibia/fibula while the subtalar connects the talus and the calcaneus. The JCS for the ankle joints complex is illustrated in figure 3.2, it is composed of three main rotational axes:

- e_1 is the axis fixed to the tibia/fibula, it is the Z-axis of the tibia/fibula CCS. The rotational movements around this axis are the dorsiflexion (positive) and the plantarflexion (negative).
- e_2 is the floating axes, perpendicular to e_1 and e_3 . The rotational movements associated to

these axes are the inversion (positive) and the eversion (negative).

- e_3 is the axis fixed to the calcaneus, it is the y-axis of the calcaneus CCS. The rotational movements around this axis are the internal rotation (positive) and the external rotation (negative)

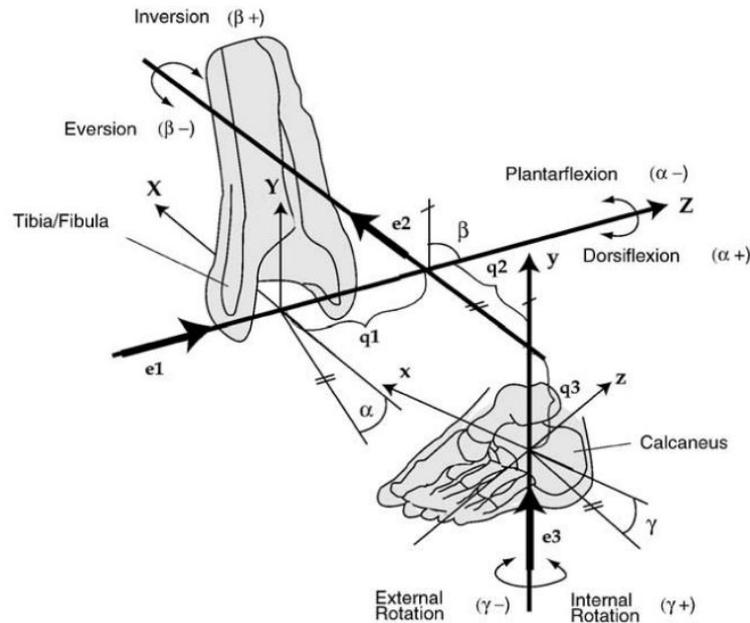


Figure 3.2: The Joint Coordinate System (JCS) of the right ankle complex. Figure from Wu et al. 2002 [2]

The most significant movement of the ankle joint is the plantar/dorsiflexion that occurs in the sagittal plane. The ankle plantar and dorsiflexion during one gait cycle are reported in figure 3.5. At the very beginning of the gait cycle, the ankle joint is close to its neutral position, then after the heel strike, it undergoes a rapid plantar flexion down to -10 degrees reaching the foot flat position. Subsequently, the ankle is slowly dorsiflexed until 10 degrees. The ankle is then plantarflexed reaching the swing phase when it is dorsiflexed.

The Knee joint The knee joint connects the femur to the tibia and includes the tibiofemoral and the patellofemoral articulations.

The JCS for the ankle joints complex is illustrated in figure 3.3, it is composed of three main rotational axes:

- e_1 is the axis fixed to the femur, it is the X-axis of the femur CCS. The rotational movements around this axis are the flexion (positive) and the extension (negative).
- e_2 is the floating axes, perpendicular to e_1 and e_2 . The rotational movements associated to this axes are the abduction (positive) and abduction (negative).

- e_3 is the axis fixed to the tibia, it is the z-axis of the tibial CCS. The rotational movements around this axis are the internal (negative) and the external (positive) rotation.

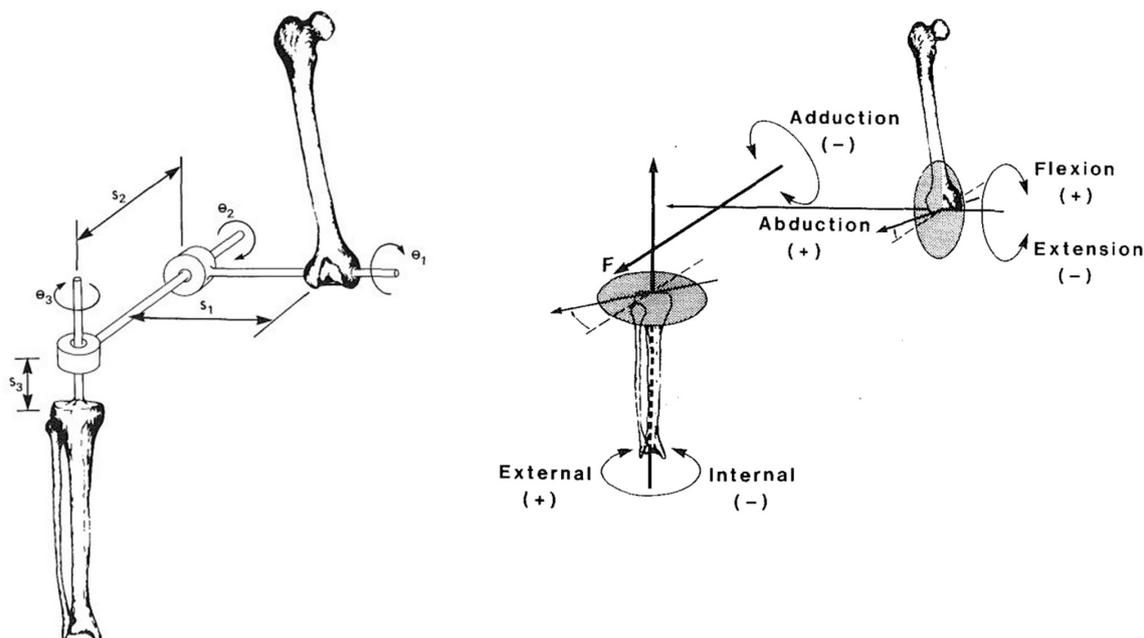


Figure 3.3: Illustration of the Joint Coordinate System (left) and the three rotational movements (right) associated to the knee joint. Figure adapted from Grood et Suntay. 1983 [3]

The most significant movement of the knee joint is the flexion/extension in the sagittal plane. As shown in figure 3.5 during a normal gait cycle the knee flexion at initial contact is about 5 degrees and increases during the loading response. During the stance phase, the knee flexion is minimum while it reaches its maximum flexion of about 60 degrees during the swing phase.

The Hip joint The hip joint connects the head of the femur and the acetabulum of the pelvis and can be modelled as a spherical hinge [17]. The JCS for the ankle joints complex is illustrated in figure 3.4, it is composed by three main rotational axis:

- e_1 is the axis fixed to the pelvis, it is the Z-axis of the pelvic CCS. The rotational movements around this axis are the flexion and the extension.
- e_2 is the floating axes, perpendicular to e_1 and e_3 . The rotational movements associated to this axis are the adduction and abduction.
- e_3 is the axis fixed to the femur, it is the y-axis of the femur CCS. The rotational movements around this axis are the internal and the external rotation.

The most significant movements of the hip during the gait cycle occur in the sagittal plane, as shown in figure 3.5. The hip is flexed during the loading response and the swing phase while reaching its maximum extension during the stance phase.

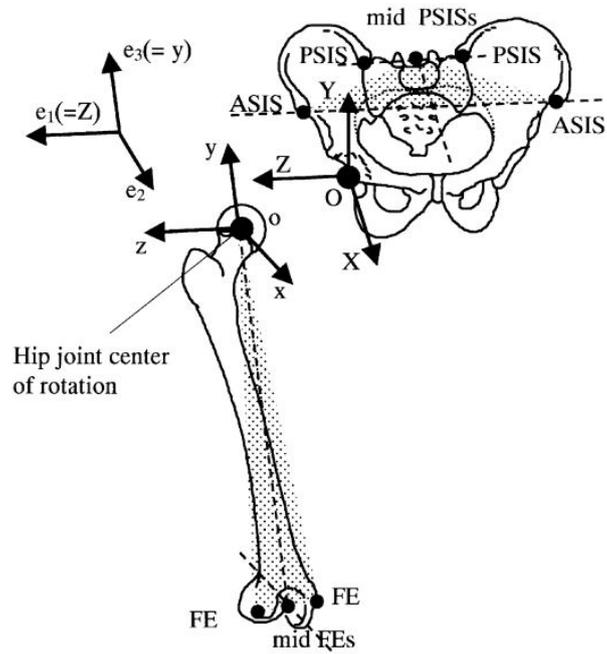


Figure 3.4: The Joint Coordinate System (JCS) of the hip joint. Figure from Wu et al. 2002 [2]

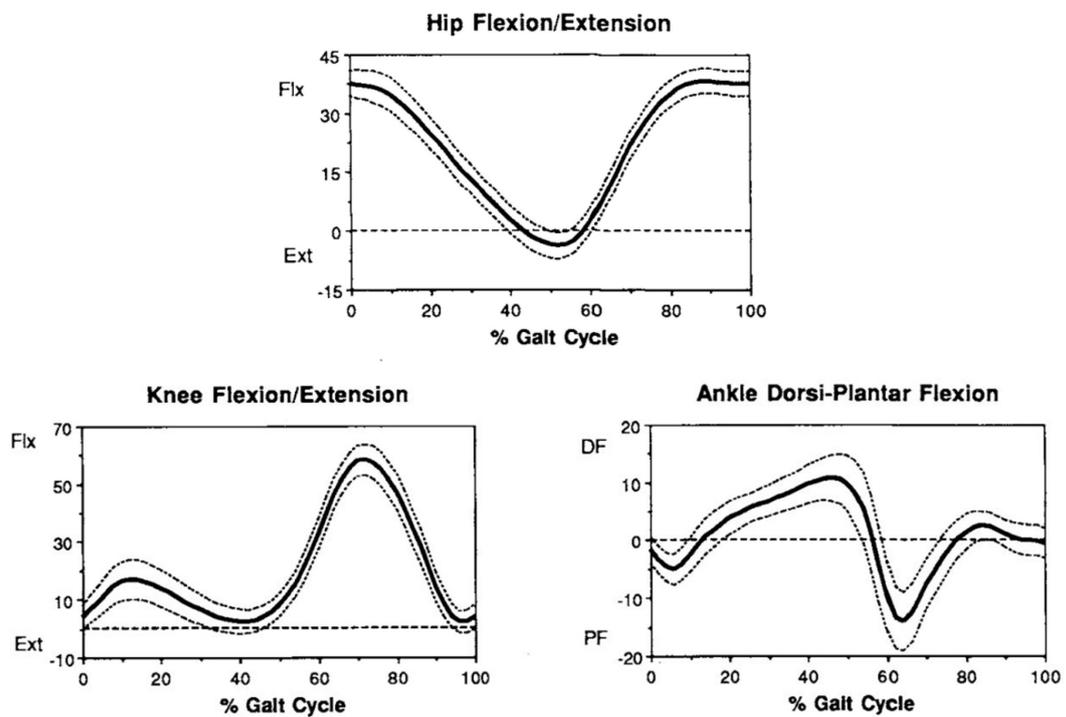


Figure 3.5: Mean (solid line) and standard deviation (dotted line) of hip, knee and ankle angles in sagittal plane. All angles are in degrees. Figure adapted from Kadaba et al. 1990 [4]

MOTION CAPTURE

Human motion capture refers to the recording of movement, which is useful in sport, research, and rehabilitation purposes. There are different approaches for motion capture either based on electro-mechanical, electro-magnetical, inertial or optical principles. Optical motion capture systems use video technology to record movement. There are two types of video-based motion capture: the marker-based (MB) systems and the markerless (ML) systems. 3D Video-based MB systems are widespread in clinical gait analysis for the assessment, treatment planning, and evaluation of gait conditions but have some limitations, markerless systems instead are gaining momentum recently due to their affordability and reduced processing time [10]. This chapter comprises a description of a three-dimensional MB system, the optoelectronic stereophotogrammetry (SP), its limitations, and an overview of the potential of the markerless approaches and some of their applications in gait analysis focusing on the use of a single RGB-Depth camera for clinical gait analysis.

4.1 Marker-based system

Three-dimensional marker-based systems are a widespread and reliable method for evaluating the joint kinematics [10]. Optoelectronic stereophotogrammetry (SP) is an accurate, non-invasive system that uses multiple cameras and image processing techniques to detect the instantaneous 3D position of markers placed on anatomical landmarks of interest. The marker placement is crucial for the accuracy of the estimations, it requires trained operators and long preparation time and might be subjected to human errors. The investigations with SP systems must be conducted in gait laboratories and the overall cost of those systems is still very high [11]. Optical SP systems can be divided into active-marker systems (e.g. Optotrack) that use infrared (IR) light-emitting markers and passive-marker systems (e.g. Vicon) that use retro-reflective markers whose position is tracked by a set of IR cameras. Active LED markers are activated in sequence so that only one marker at a time can be

detected by the cameras. Passive markers receive the IR light from all cameras at the same time, they cannot be distinguished from one another, so they need to be labelled in post-processing. While the active markers systems give more accurate measurements, they need cables and power supplies that could limit the ease of movement.

To reconstruct the 3D position of the markers inside the volume capture, the data from the 2D images collected from each camera are combined using triangulation. To achieve this aim at least two cameras are needed but markers can be occluded by the subject's movements so usually at least five or six cameras placed in different positions above the volume capture are used [5]. To perform the triangulation the first step is the calibration of the system to obtain the intrinsic parameters of each camera (e.g. the focal length, the optical distortion introduced by the cameras lens) and the position and orientation of each camera reference frame with respect to a global reference frame (extrinsic parameters).

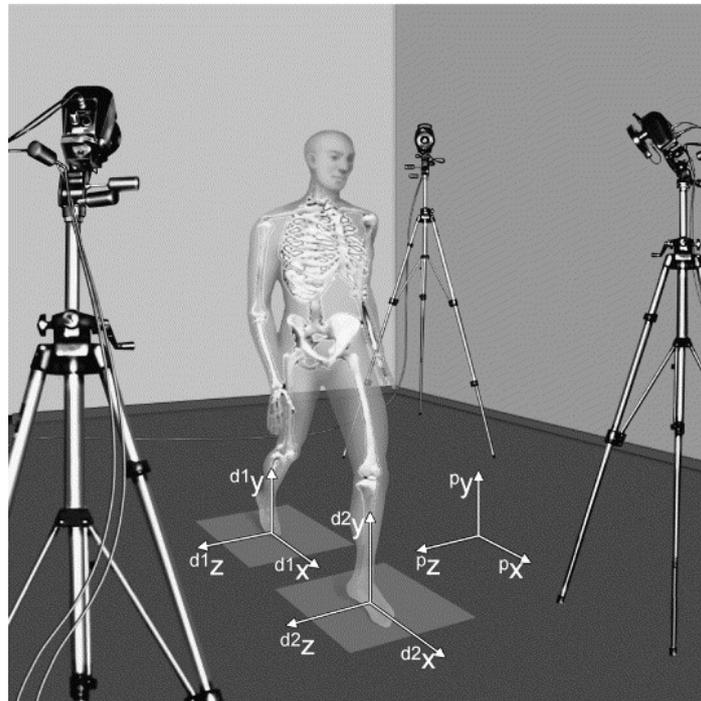


Figure 4.1: Illustration of a movement analysis laboratory with stereophotogrammetry system, force plates and the associated reference systems. Figure from Cappozzo et al. 2005 [5]

To reconstruct the joint kinematics a biomechanical model of the body is needed. The body is modelled with different segments, corresponding to the bones of interest and connected by rotational joints. Under the assumption that they represent rigid bodies, their position and orientation can be tracked over time and the joint angles can be estimated. As reported in literature [5], the main source of errors in marker-based SP may arise because of instrumental errors [18], soft tissue artefacts [19] or anatomical landmark position uncertainties [20].

4.2 Depth camera and markerless systems

ML systems are made of one or more cameras to record movement and software algorithms to extract the subjects' body segment position and orientation. ML systems currently embody a promising alternative to marker-based systems because of their affordability, their reduced recording and processing time and the possibility to perform out-of-lab recordings. These techniques do not require highly trained operators as they can be fully automatic and also provide the possibility to re-process old datasets with improved versions of more accurate algorithms [11].

Thanks to the major developments in computer vision, this technology and its applications have substantially increased over the past few years. Also, the accuracy of these methods is currently under many investigations to demonstrate their potential for clinical and rehabilitation purposes [11].

Markerless methods include both multi or single-camera approaches for data collection, using standard RGB cameras or RGB-Depth cameras. One example of a multi-camera application in biomechanics is Theia3D [21], a markerless motion capture system that utilizes at least six synchronized cameras. Theia3D software enables the detection of numerous movements (Figure 4.2) and computes the 3D position and orientation of the body segments of one or more subjects also in different environments [21].

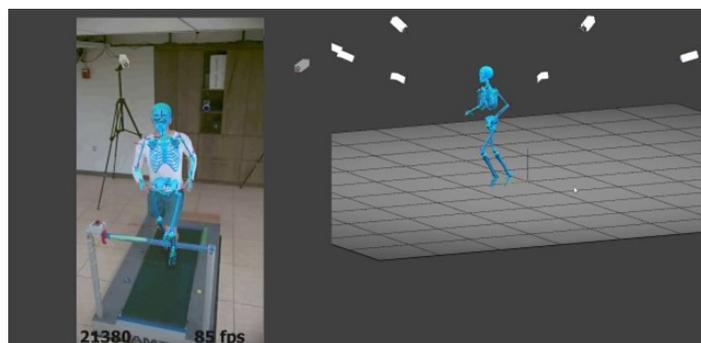


Figure 4.2: Illustration of the camera set up and the movement detection performed through Theia 3D Markerless. Figure from Theia Markerless

Markerless analyses are performed in two or three dimensions but generally for gait analysis, as the movement involved during walking could be mainly evaluated in the sagittal plane, a 2-dimensional approach could be considered sufficient [12]. Moreover, the use of a single RGB camera is a very cost and space-effective alternative with respect to multi-camera systems. An interesting 2D markerless method for gait analysis is the one reported by Castelli et al. [10]. This method enables the extraction of the lower limb sagittal plane kinematics during gait using only a single RGB camera (Vicon Bonita Video) through the

definition of a subject-specific kinematic model.

By using only a single RGB camera the problems of occlusion during movements are increased leading to increased errors in the joint centres' identifications. One more promising technology is the use of RGB-Depth cameras. RGB-D cameras, such as the Orbecc, the Intel Realsense and the Microsoft Kinect, are equipped with an integrated depth sensor that provides an image containing information on the distances from the device. Three main approaches are used in depth sensors to reconstruct the depth images:

- Fixed structured light, in which the infra-red (IR) emitter projects a known pattern which is often made of grids or horizontal bars onto the scene. This pattern strikes the surfaces and is distorted by the presence of objects. The reflected pattern is captured by an IR Depth Sensor and then by comparing the emitted pattern with the distorted one through a triangulation process, the distance between the object and the camera can be calculated.
- Stereoscopic vision, in which at least two cameras are needed to reconstruct the 3D scene. The depth information is calculated known the cameras' intrinsic and extrinsic parameters and the distance between them (also called baseline).
- Time-of-flight (ToF), in which the sensor emits IR light onto the scene and captures the reflected signal. The distance is then estimated from the phase shift calculated between the emitted and the reflected signal.

The distance information is then gathered to reconstruct a depth map which is a set of depth values for all the image pixels expressed in millimetres.

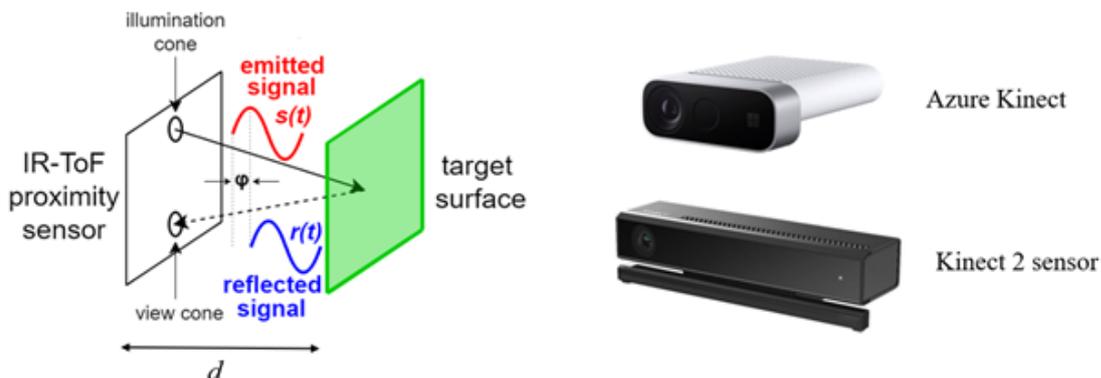


Figure 4.3: Illustration of the Time-of-Flight principle, implemented by the Azure Kinect and the Kinect 2

RGB-Depth cameras like the Microsoft Kinect often come with software development kits (SDK) for the joint position estimation to track one or more subjects' body positions and orientation. These algorithms are often based on deep learning techniques, which showed good estimation capabilities in human pose estimation. They use Convolutional Neural Networks (CNN) which are deep learning algorithms made of many hidden layers. CNNs are

trained on large datasets called training sets to recognize patterns in the images related to anatomical points of interest. The training is done through a process called supervised learning in which the network is trained to recognize patterns of pixels' RGB values, gradients, or texture features [11] on many images that contain labels of the body parts of interest. In this way, after analysing many examples, the neural network is capable to recognize the same features related to the anatomical landmarks on a completely new image which is not labelled.

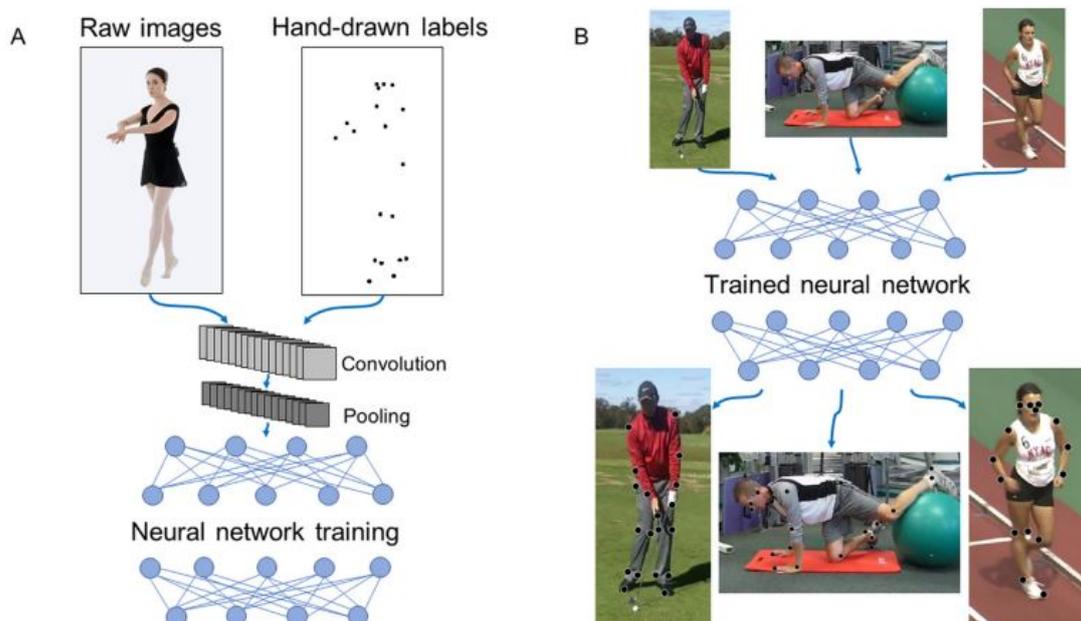


Figure 4.4: Example of pose estimation based on a Convolutional Neural Network (CNN). First, the CNN was trained through supervised learning with a set of manually labelled images (A) then the trained network is capable of processing and labelling new images (B). Figure from Cronin et al. 2021 [6]

A successful example is Openpose [22], an open-source and real-time system that enables 2D pose estimation of one or more subjects (Figure 4.5). OpenPose has been employed in gait analysis in a 3D markerless system of multiple synchronized cameras, that computed joint positions and whose results were compared to a marker-based method showing good accuracy [7].

The performances of deep-learning-based algorithms largely depend on the quality and variety of the training dataset and on the precision of the manual labelling performed on the training set [6]. The accuracy of detection might decrease in images that are much different from the ones they have been trained on or in the recognition of unusual movements [11].

For what concerns the clinical applications of gait analysis, a deep learning approach is not suitable because of the unavailability of large training sets of pathological data and

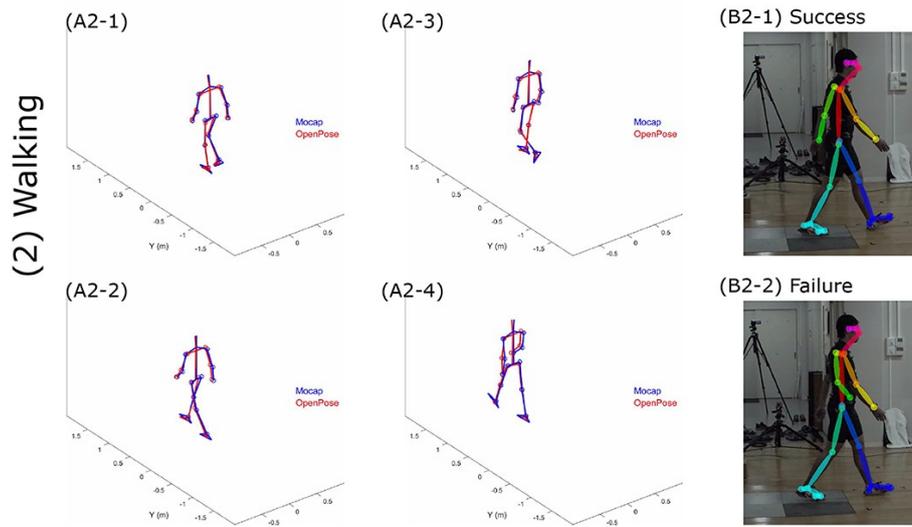


Figure 4.5: (A) Comparison of 3D pose estimation outputs of the marker-based (Mocap) and the OpenPose markerless method (OpenPose) during walking; (B) Examples of 2D estimations success and failure. Figure adapted from [7].

because it does not take into account the subject specificity which might lead to estimation inaccuracies. A different approach based on subject-specific kinematic models [10], [9], [14], [23] has been implemented to address this needs in clinical gait analysis.

The purpose of this thesis project is to compare the method developed by Balta et al. [23] with the outputs of the deep-learning-based algorithm of the Microsoft Azure Kinect body tracking SDK [24] in a 2D markerless gait analysis protocol.

SENSOR CHARACTERIZATION

5.1 Azure Kinect DK

The Azure Kinect development kit (DK) was released by Microsoft in 2020. It is a developer kit that contains one RGB camera integrated with a depth sensor, an array of 7 microphones and two orientation sensors (an accelerometer and a gyroscope) in a very compact and light-weight hardware [8]. The color and depth sensor max framerate is 30 frames per second. The Azure Kinect DK contains three software development kits (SDKs): a sensor SDK for sensor and device access, the Body Tracking SDK for the 3-dimensional tracking of bodies and the Speech Cognitive Services SDK [13]. The external pins on the camera also enable

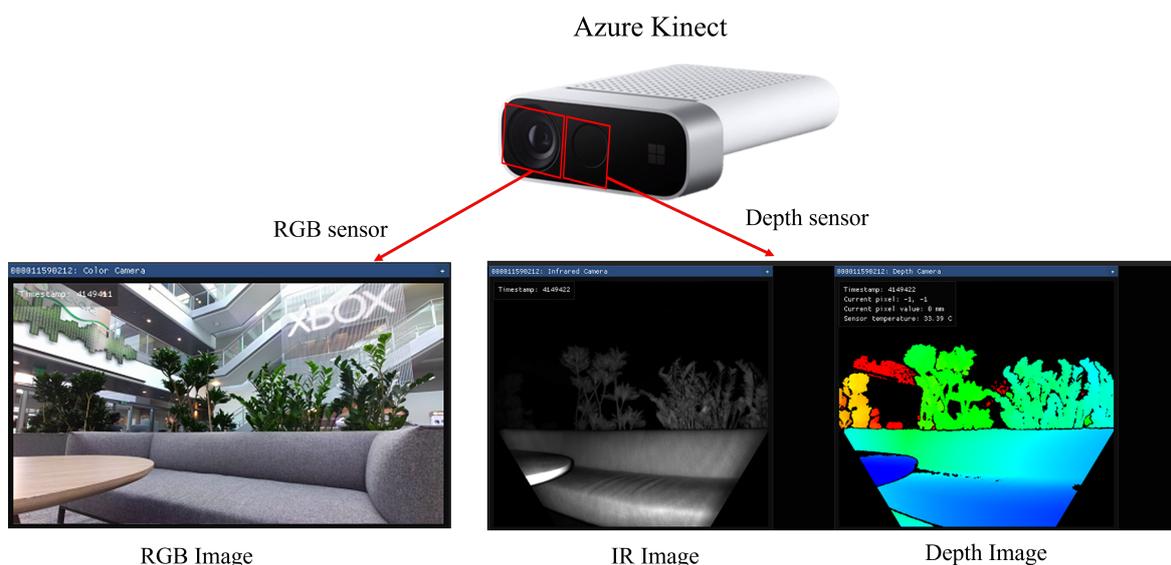


Figure 5.1: Illustration of the Azure Kinect and the output images from the Azure Kinect Viewer. Figures adapted from Azure Kinect Viewer

the possibility to synchronize two or more devices.

The Azure Kinect produces an RGB color image, a depth map, and a clear IR image. The pixel values in the IR image are proportional to the amount of light returned from the scene. Figure 5.1 shows an example of an RGB image (left), an IR image (centre) and a depth image (right). The Azure Kinect RGB video camera has 12 MegaPixels with an OV12A10 CMOS rolling shutter sensor. The RGB camera exposure, white balance and other image settings can be controlled by the user. The Depth camera [25] emits amplitude modulated continuous waves and is based on the Time-of-Flight (ToF) principle. The depth camera supports two operating modes: the wide field-of-view (WFOV) which is indicated for scenes with a large extension in the X and Y dimensions and small values of depth, and the narrow field-of-view (NFOV) which is indicated for scenes with larger extensions in the Z dimension. Two depth images captured with the NFOV and the WFOV modes are shown in figure 5.2.



Figure 5.2: Examples of depth images taken with narrow field-of-view (left) and wide field-of-view (right) mode.

As reported by the producers, in certain situations the depth map reconstruction may be subjected to errors and provide invalidated pixels [26]. Invalidated pixels have 0 values that show up as black spots in the depth images. This might occur because:

- The IR signal is too low for some pixels. An example is shown in figure 5.3.
- Pixels contain saturated IR signal. An example is shown in figure 5.3.
- Pixels are outside of the active IR illumination mask. This results in invalidated black pixels that can be seen around the corners of the depth images.
- Multi-path interference, this may occur for pixels that received IR signals from more than one object in the scene. This mainly occurs in the corners. Another example that can be seen around the object's edges are pixels which received mixed signals from the foreground and the background, this may cause ambiguities in the depth reconstruction.

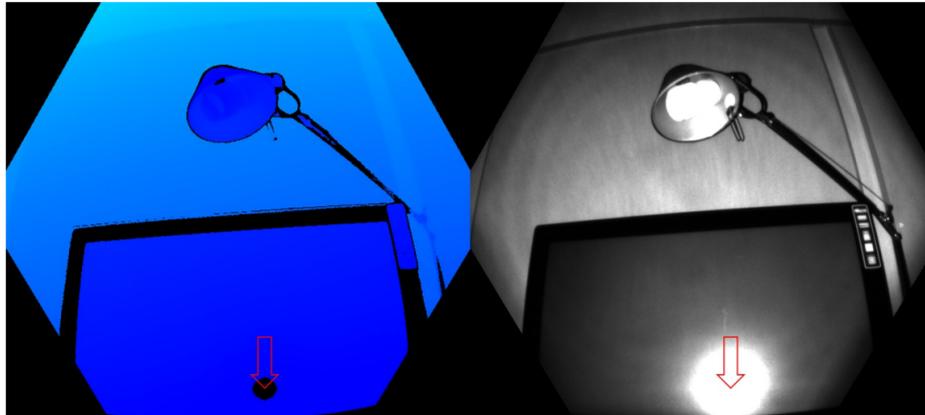


Figure 5.3: Illustration showing the effect on depth pixels of a saturated IR signal. The arrows point to the invalidated pixels in both the depth (left) and IR (right) images. Figure from [8].

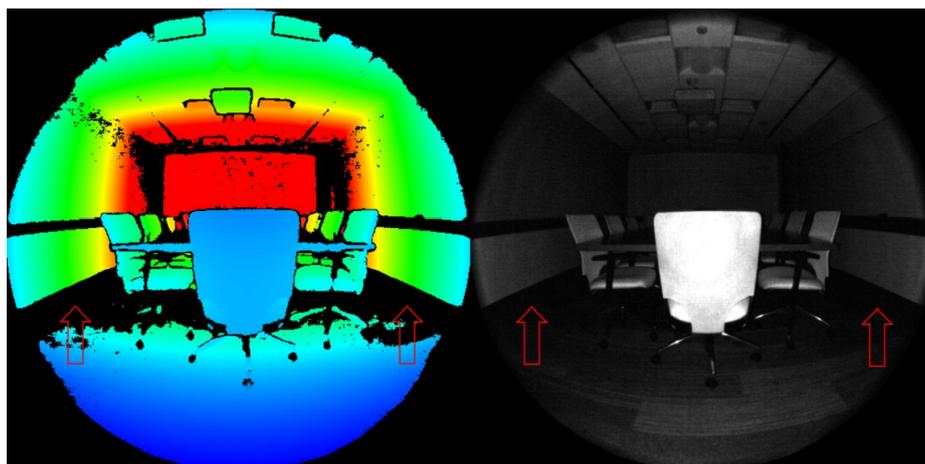


Figure 5.4: Illustration showing the effect on depth pixels of an IR signal which is too low. The arrows point to the invalidated pixels in both the depth (left) and IR (right) images. Figure from [8].

5.2 Preliminary characterization of depth camera performance

The accuracy of the depth distances estimations were investigated to verify the performances of the depth sensor reported by the producer whose distance error is less than $11 \text{ mm} + 0.1\%$ of the distance in the camera operating range (0.25 to 3.86 m) in NFOV unbinned mode.

5.2.1 Static Recordings

To evaluate the accuracy of the depth sensor a rectangular box was placed in front of the sensor at 2.6 m of distance and recorded for about 3 seconds at 30 fps in NFOV unbinned mode. A row of pixels belonging to the central part of the box was selected in each frame and the average distance over 126 frames was computed (Figure 5.5.a). It is noticeable that the distances from the mean value increase at the edges of the object which represent the lowest and highest column values.

The known-sized object was then placed in front of the camera at four different distances (120, 182, 240 and 300 cm). In each acquisition, the mean distance over 46 frames of the object from the depth image was computed with the same method described above. The errors with respect to the real distance are calculated and shown in figure 5.5.b.

From Table 5.1, it is possible to see that for all distances the error is under $11 \text{ mm} + 0.1\%$ of the distance in compliance with the device specifications. At the maximum distance of 300 cm the mean error with respect to the real distance is the highest (12.4 mm).

Distances (mm)	Mean Distance (mm)	Mean error (mm)
1200	1210.9	10.9
1820	1831.9	11.9
2400	2411.6	11.6
3000	3012.4	12.4

Table 5.1: Real distance of the known-sized object, mean distances computed from the Azure Kinect depth images and their differences.

To evaluate the repeatability of the acquisitions, the box was positioned at a fixed distance of 2.6 m. The standard deviation of each pixel over 84 Depth frames was computed. Results are shown in Figure 5.7, where it can be seen that the standard deviation is higher at the edges of the objects resulting in a lower accuracy with respect to the centre. The highest standard deviation values occur in the corners between the floor and the wall of the image.

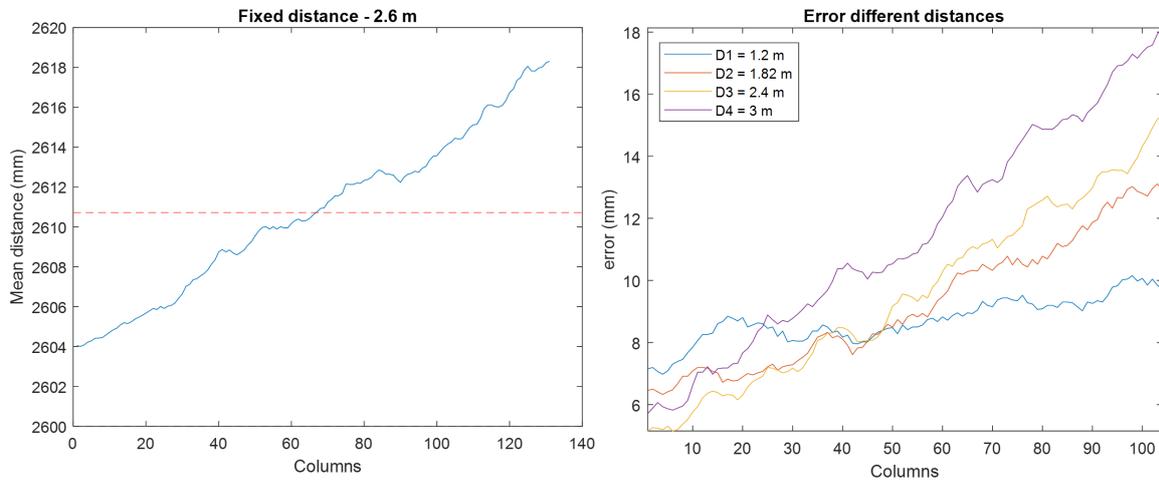


Figure 5.5: a) Average depth pixels values over 126 frames of static acquisitions at a fixed distance. The mean distance is represented by the dotted line. b) Errors with respect to the object's real distance from the camera, placed at different positions (D1 = 1200 mm, D2 = 1820 mm, D3 = 2400 mm, D4 = 3000 mm)

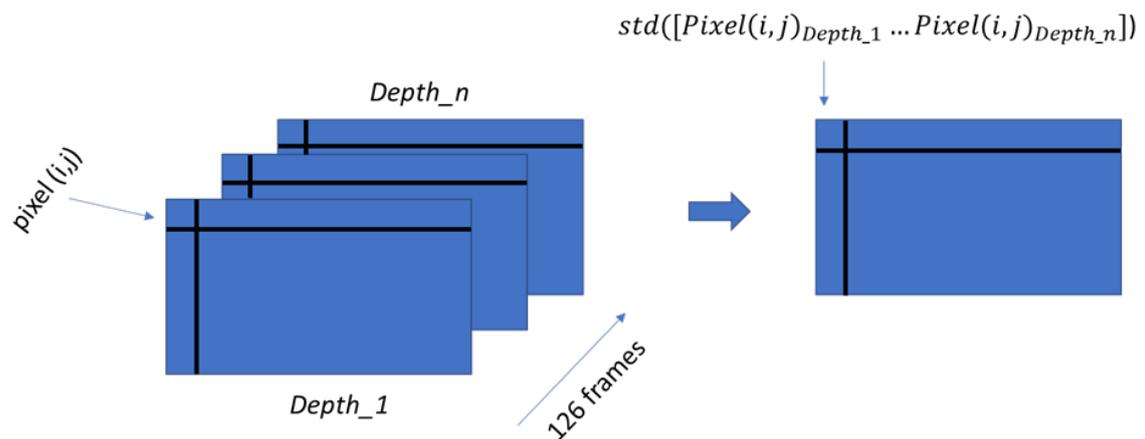


Figure 5.6: Illustration of the standard deviation map calculation. The standard deviation was computed over each pixel for 126 frames of static acquisitions at fixed distance

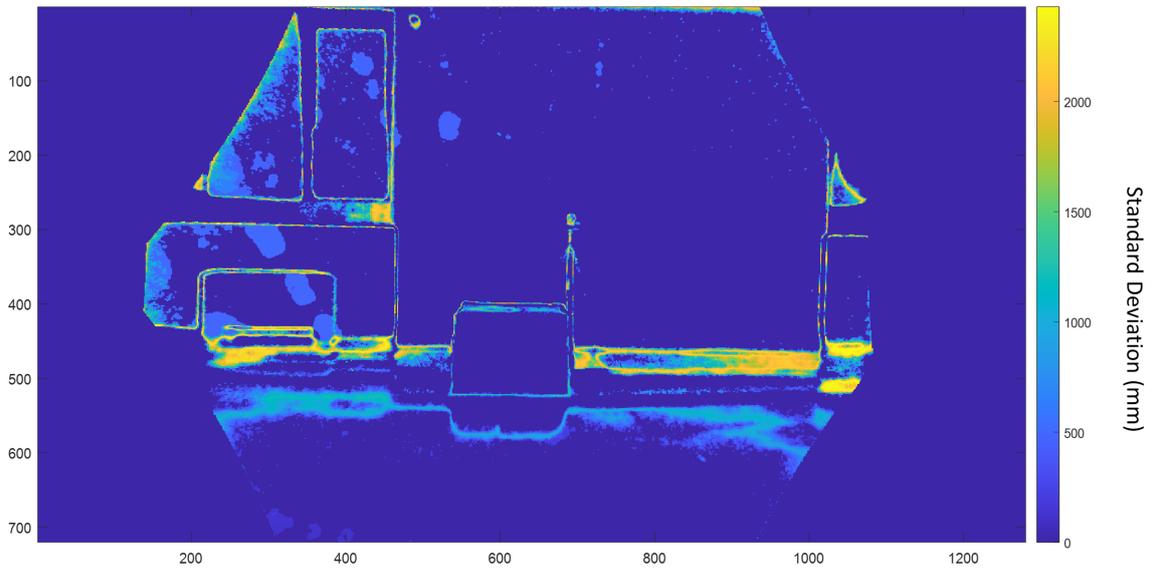


Figure 5.7: Color image (top) and standard deviation map (bottom) of the known-sized object placed at a fixed distance of 2.6 m.

5.2.2 Dynamic Recordings

To evaluate the influence of the speed on the Azure Kinect depth sensor, we considered the oscillations of a pendulum since the swinging phase of a typical gait cycle is quite comparable to the swinging motion of the pendulum. The pendulum was free to oscillate starting from a static horizontal position. About 30 seconds of oscillations at 30 fps were recorded.

The pendulum is a 61 x 5 cm wooden bar which is positioned at a 2.35 m distance from the camera. The pendulum was identified in each depth frame through a thresholding method. To compute the pendulum speed, the extremity of the bar was identified in each segmented image. To achieve this aim, the bar was fitted through the ellipse that best approximates it as

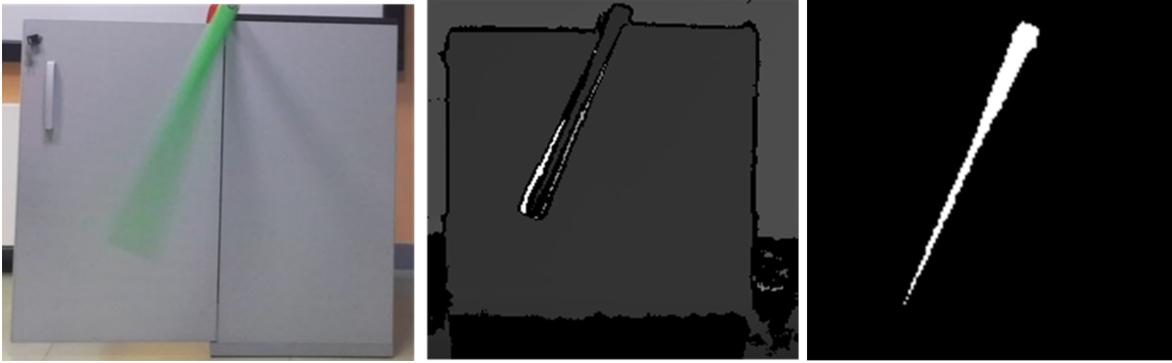


Figure 5.8: From left to right: RGB, Depth image and segmentation mask of the oscillating pendulum.

shown in figure 5.9. The principal axes of inertia and the centroid of this ellipse have been identified to calculate his orientation with respect to the x-axis. Then, the image reference system of the image was aligned to the principal axes of inertia and the extremity of the bar is the point (E, in Figure 3) with the higher x-coordinate. The pendulum speed was calculated as:

$$PendulumSpeed = \frac{E(t_i) - E(t_{i+1})}{t_{i+1} - t_i}$$

Where $E(t_i) - E(t_{i+1})$ are the extremities of the bar in two consecutive frames. The time difference is considered constant by assuming frame rate = 30 fps. From Figure 5.8, it is easy to notice that when the pendulum speed increases its area linearly decreases. Finally, it was possible to conclude that the faster the movement, the lower the depth accuracy.

To evaluate the repeatability of the depth sensor in dynamic acquisitions the standard deviation of each pixel was computed for the frames of the pendulum swinging motion where the bar was in a similar position. The frames in which the pendulum was in a vertical position, where the maximum speed is reached, were selected. These frames were found through the identification of the positive peaks of the extreme row of the pendulum (figure

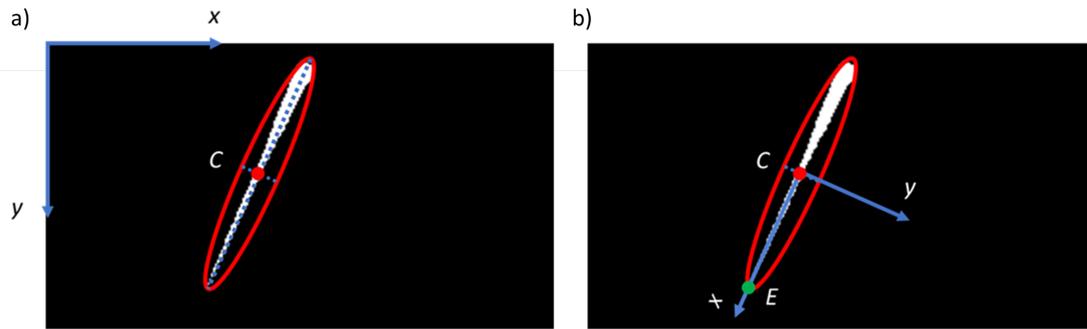


Figure 5.9: a) Identification of the lower extremity of the pendulum through an ellipse which was fitted on the bar. The image reference system is on the top left of the image. b) The reference system of the image is aligned to the inertial principal axes of the pendulum for the identification of the coordinates of the extreme point of the bar (E).

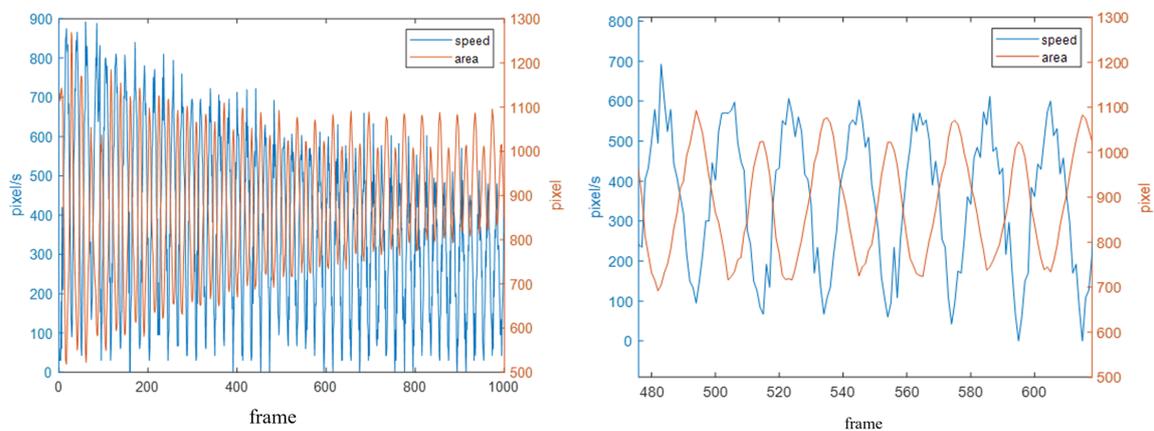


Figure 5.10: Speed (blue) and pendulum area (orange) for each frame of the swinging motion. The figure on the right is a zoom of the figure on the left.

5.11). Only 17 frames that showed the lowest variability between their column values were selected. The standard deviation of each pixel over 17 frames was computed with the same method shown for the static acquisitions (figure 5.6) and the resulting standard deviation map is shown in Figure 5.12.

From these primary investigations, it was possible to conclude that in depth images captured with the Azure Kinect:

- The farther the object, the higher the standard deviation of depth pixels
- A lower accuracy is seen at the edges of the objects with respect to the centre
- When the object's speed increases, its area linearly decreases
- The faster the object is moving, the lower the depth pixels' accuracy

Before the methods comparison, the problem of pixel invalidation in the images of the depth sensor was examined also on the recordings of a single gait cycle. From the depth images

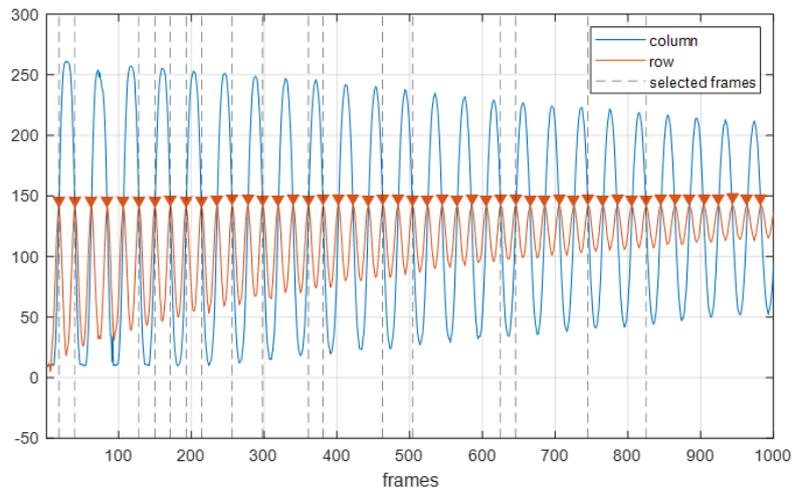


Figure 5.11: Column and rows coordinates of the pendulum extremity. The dotted lines represent the 17 selected frames in which the pendulum was in a vertical position

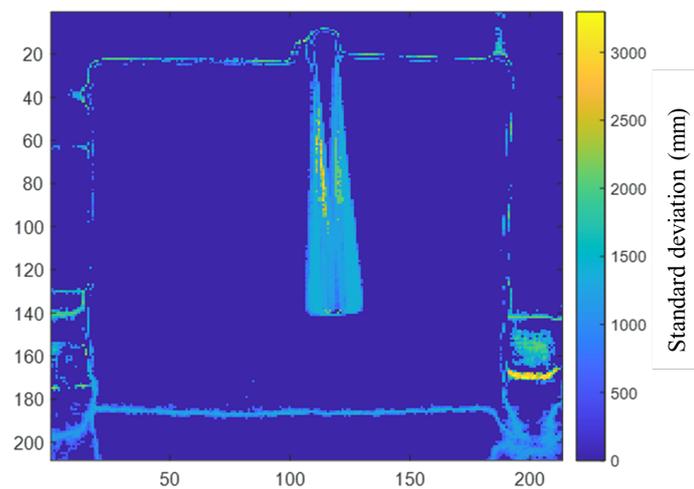


Figure 5.12: Standard deviation computed over each pixel for 17 frames of dynamic acquisition of the pendulum in a vertical position

captured during the swing phase of a typical gait cycle, it was observed that when the walking speed increases some information about the shank and the thigh depth values is lost, as shown in figure 5.13.

This phenomenon, which is particularly visible along the edges of the subject's lower limbs in the direction of progression, reduces the total area of the shank and affects its identification in our processing with the model-based method (described in Chapter 6).

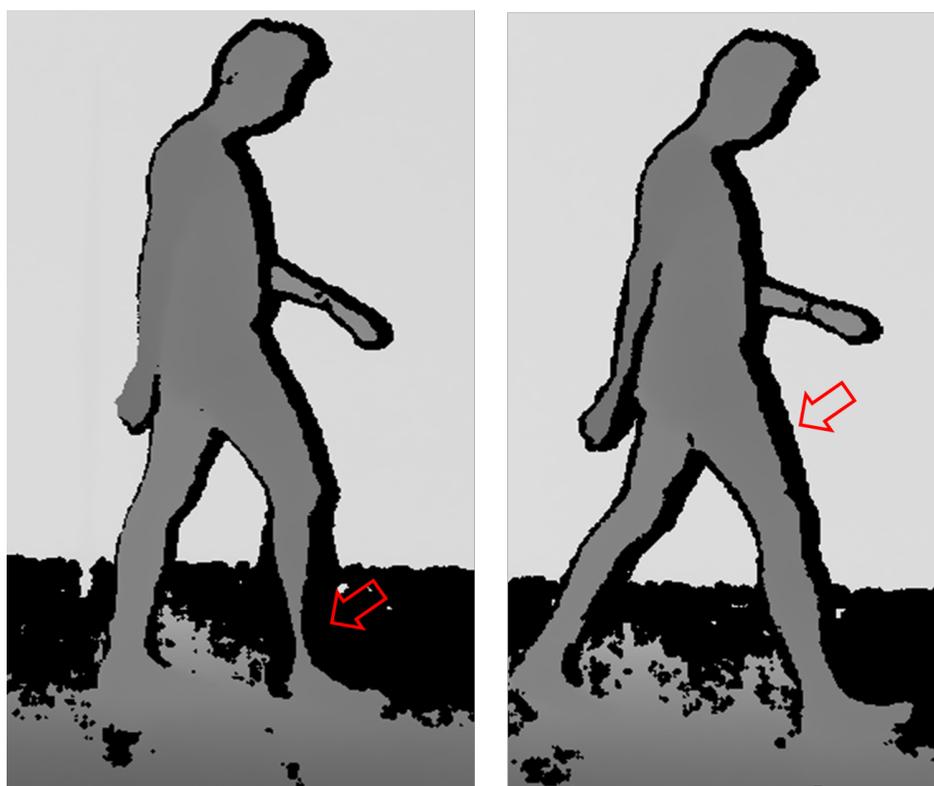


Figure 5.13: Illustration of invalidated pixels of the shank (left) and thigh (right) during the preliminary recordings of a gait cycle in the sagittal plane. The invalidated pixels mainly occur along the edges of the subject body and in the direction of progression

METHODS FOR JOINT KINEMATIC ESTIMATION

As mentioned in Chapter 2, this thesis project aims to compare two methods for 2D markerless gait analysis through a single RGB-Depth camera (Microsoft Azure Kinect). The two methods under investigation are: (i) a method based on a subject-specific kinematic model and (ii) a method based on a deep neural network.

6.1 Model-based Markerless Method (MLM)

The first method, implemented on markerless recordings from the Azure Kinect for the sagittal lower limb kinematics estimation, is based on a 2D clinical gait analysis protocol reported by Balta et al. 2020 [23]. The proposed method includes four main steps: the subject segmentation, the models definition, the lower limb body segments pose estimation and the lower-limb joint angles computation.

Subject segmentation

After the identification of the frames belonging to a single gait cycle, the subject's body must be identified in each RGB frame. All the steps of the subject segmentation are reported in figure 6.1.

The first step is to separate the subject's body from the background. For this purpose, an image containing only the background is intentionally captured before each acquisition. The background image is subtracted from the frame containing the subject and a thresholding method is implemented on the difference image.

The resulting difference image is converted to greyscale and the histogram of its grey values is computed. Based on the histogram of the difference image the optimal threshold that

separates the subject from the background is identified and applied to obtain a segmentation mask.

The segmentation mask, shown in the second figure of 6.1, now contains both the subject's real body and its shadow. To remove the residual shadow the depth image information is extracted and a thresholding method is applied. The selected threshold was 2500 mm, in this way all the pixels that belong to the background can be easily removed.

Some residual noise still visible in the segmentation mask is then removed by applying some morphological operations.

The subject was asked to wear a red sock on the right foot and a blue sock on the left foot to simplify the identification of the feet in the RGB frames.

The foot segmentation is achieved by implementing a simple red color filter for the right foot and a blue color filter for the blue foot. The obtained segmentation mask still contains some pixels belonging to the green carpet, to remove this residual noise the segmented feet were removed from the segmentation mask. At this stage the segmentation mask only contains small connected regions belonging to the green carpet that are removed through morphological operations to obtain the final segmentation mask.

Background
Subtraction



Thresholding
Method



Noise
Removal



Feet
Segmentation



Removal of
small connected
regions



Final
Segmentation

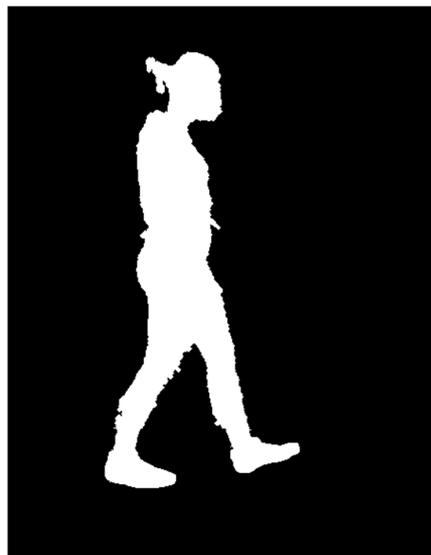


Figure 6.1: Block diagram of the subject segmentation step of the markerless method for the estimation of the lower limb joint kinematics.

Model Definition

Three subject-specific kinematic models for the foreground lower limb were defined from manually selected frames in a static upright position, the loading phase, and the swing phase (Figure 6.2). The kinematic model is made of three body segments: the foot, shank, and thigh. The three segments are linked by: the ankle joint over the lateral malleolus (LM), the knee joint over the lateral epicondyle (LE) and the hip joint over the great trochanter (GT).

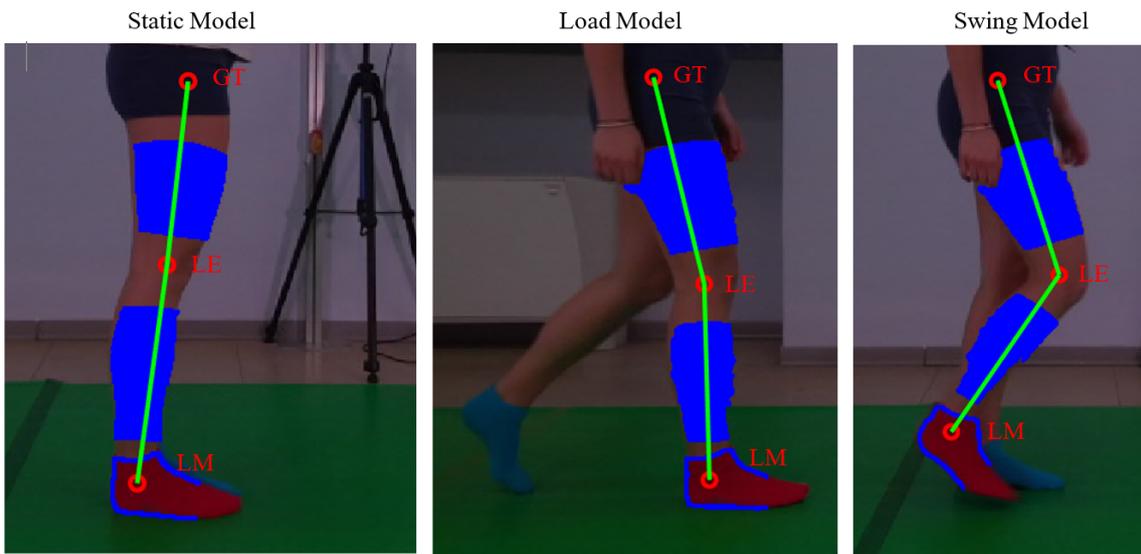


Figure 6.2: Illustration of the three subject-specific lower limb models in a static upstanding position, during the loading phase and the swing phases.

The lower limb model structure is shown in figure 6.3, it is defined starting from the three anatomical landmarks that are manually identified by a trained operator. The foot, shank and thigh models were defined by using a bottom-up approach as follows:

- Foot model, it is defined as the posterior part of the foot, from 0 to 90% of the previously segmented. The foot coordinate system is shown in figure 6.3, its origin is located at the extreme lower pixel of the foot edge. The x -axis positive direction is towards the toe.

$$foot(x, y) = \begin{cases} 1 & \text{if } 0 < I(x, y) < 90\% \text{ of the foot length} \\ 0 & \text{otherwise} \end{cases}$$

- Shank model, it is defined as the pixels of the subject segmentation mask ranging from the radii of 25% and 75% of the segment that starts from the LM and ends in the LE position centred in LM.

The origin of the shank coordinate system is the LM position.

$$shank(x, y) = \begin{cases} 1 & \text{if } 25\% < I(x, y) < 75\% \text{ of } L_{LM-LE} \\ 0 & \text{otherwise} \end{cases}$$

- Thigh model, it is defined as the pixels of the subject segmentation mask ranging from the radii of 25% and 75% of the segment that starts from the LE and ends in the GT position centred in LE. The origin of the thigh coordinate system is the LE position.

$$thigh(x, y) = \begin{cases} 1 & \text{if } 25\% < I(x, y) < 75\% \text{ of } L_{LE-GT} \\ 0 & \text{otherwise} \end{cases}$$

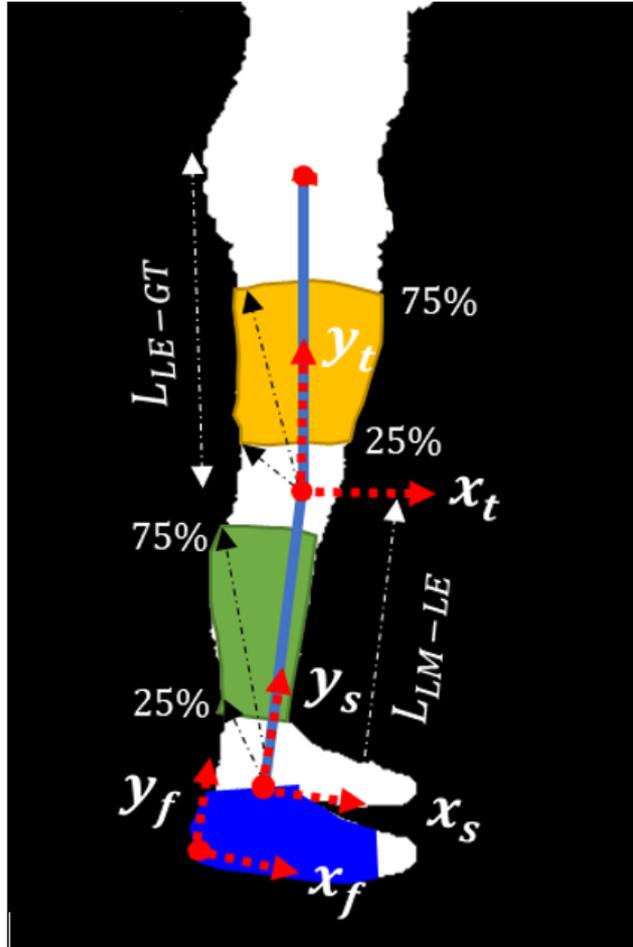


Figure 6.3: Illustration of the lower limb model structure with the foot, shank and thigh models and their coordinate systems.

Lower limb body segments estimation

In the second step of the markerless method (MLM), after the multi-segmental model's definition, the position and orientation of the lower limb body segments are estimated in each frame of the gait cycle through a bottom-up approach which is summarized in Figure 6.4.

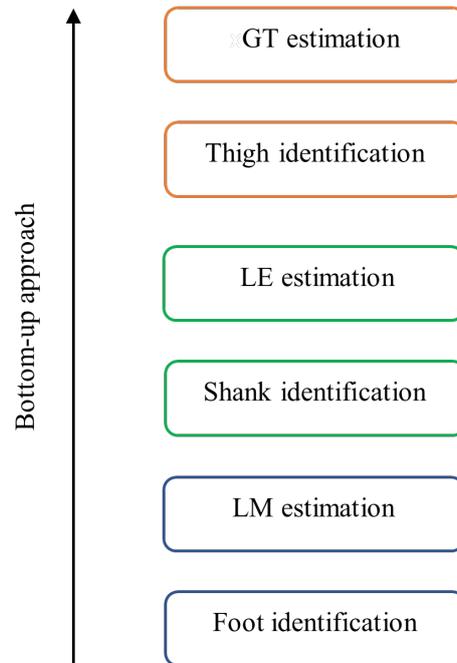


Figure 6.4: Overview of the bottom-up approach implemented for the lower limb body segment pose estimation.

Foot: The lateral malleolus position is estimated through model fitting starting from the foreground foot segmentation obtained through a colour filter. The posterior part of the foot is identified as explained in the model definition section and a transformation matrix that maximizes the superimposition between the foot in the i -th frame and the foot model is extracted using the iterative-closest point (ICP) method. The position of the lateral malleolus (LM) in each frame ($LM_{f,i}$) is calculated by multiplying the position of the LM in the model ($LM_{f,m}$) and the transformation matrix ($T_{f,i}$).

$$LM_{f,i} = LM_{f,m} \cdot T_{f,i}$$

Shank: The lateral epicondyle position is estimated through model fitting starting from the shank segmentation. To identify the shank in the subject segmentation mask a semicircular region centred in the LM position with a radius equal to the LM-LE segment length in the model is identified for each frame. Then, as shown in figure 6.5, two different cases may occur during the gait cycle:

- The foreground shank and the background shank are overlapped (e.g. during the late stance and the early swing phase). In this case, the foreground shank is identified as the sub-region with the maximum area.
- The foreground shank and the background shank are separated (e.g. during the early stance and late swing phase). In this case, the foreground shank is identified by relying on the depth image pixels. The threshold that best separates the foreground shank is identified on the envelope of the histogram of depth values based on the fact that the foreground thigh, being closer to the camera, represents the sub-region with the lower depth values and the higher pixels area as shown in Figure 6.6.



Figure 6.5: Shank region and lateral malleolus (LM) position identification in a frame in which the foreground and the background shank are separated (right) and overlapped (left).

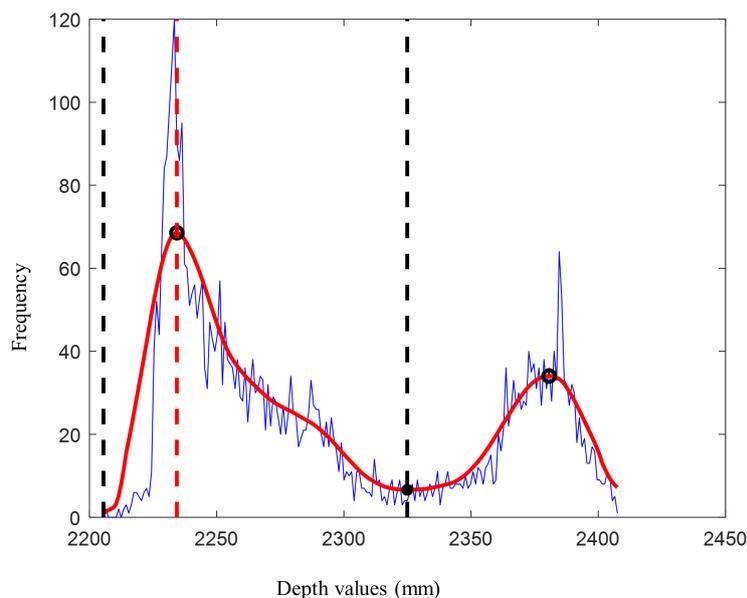


Figure 6.6: Envelope of the depth values histogram computed for the semicircular ROI centred in the LM of the foreground ankle with a radius equal to the LE-LM segment. The foreground shank is identified through the range of depth values between the two black dashed lines.

The transformation matrix that maximizes the superimposition between the shank in the i -th frame and the shank model is extracted using the iterative-closest point (ICP) method. The position of the lateral epicondyle (LE) in each frame ($LE_{s,i}$) is finally obtained by multiplying the position of the LE in the model ($LE_{s,m}$) and the transformation matrix ($T_{s,i}$).

$$LE_{s,i} = LE_{s,m} \cdot T_{s,i}$$

Thigh: The great trochanter position is estimated through model fitting starting from the thigh segmentation. The thigh is identified in the subject segmentation mask through a similar approach to the shank identification. A semicircular region centred in the LE position with a radius equal to the LE-GT segment length in the model is identified for each frame. Then, unlike the shank identification, the foreground and the background thighs are always overlapped. Thus the thigh identification is only based on the envelope of the histogram of depth values of the semicircular region. In certain frames, the hand of the subject may be also overlapped to the foreground thigh, but its depth values are lower than the foreground thigh so it can be separated based on the envelope of the histogram of depth values too (figure 6.7).

After the thigh identification, the transformation matrix that maximizes the superimposition between the thigh in each frame and the thigh model is extracted using the iterative-closest point (ICP) method. The position of the Great Trochanter ($GT_{s,i}$) is then obtained by multiplying the position of the LE in the model ($GT_{t,m}$) and the transformation matrix ($T_{t,i}$).

$$GT_{t,i} = GT_{t,m} \cdot T_{t,i}$$

The position estimation of the LM, LE and GT was conducted for the static model, the loading phase model and the swing phase model. Thus resulting in three joint trajectories estimations for the gait cycle.

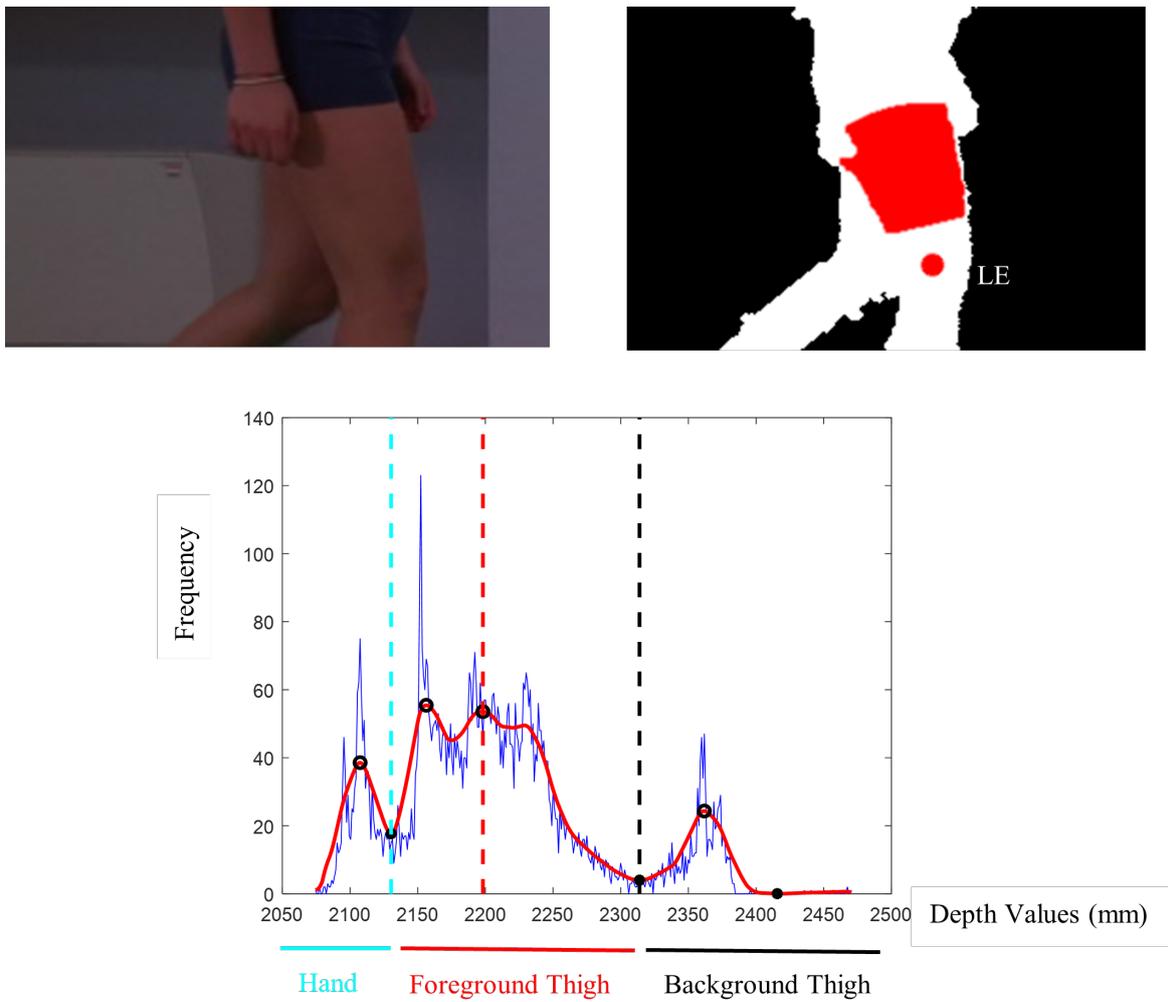


Figure 6.7: Thigh identification in an image where the hand, foreground thigh and background thigh are overlapped. RGB image (top left), thigh region identification and LE position (top right). Envelope of the depth values histogram computed for the semicircular ROI centred in the LE of the foreground knee with a radius equal to the LE-GT segment (bottom). The foreground thigh range is identified between the light blue and the black dashed lines.

Joint angles calculation

For each of the three joint position trajectories in every frame of the selected gait cycle, the joint angles trajectories during the gait cycle are computed as follows:

- The ankle joint angle is calculated as the inclination between the LM-TOE segment and the LM-LE segment minus 90 degrees. The LM-TOE segment is calculated starting from the LM position as the segment that best fits the foot segmentation.

$$\theta_{ankle} = \arccos\left(\frac{LE - LM}{|LE - LM|} \cdot \frac{T - LM}{|T - LM|}\right) - 90^\circ$$

- The knee joint angle is calculated as the angle between the LM-LE segment and the LE-GT segment

$$\theta_{knee} = \arccos\left(\frac{LE - GT}{|LE - GT|} \cdot \frac{LM - GT}{|LM - GT|}\right)$$

- The hip joint angle is calculated as the angle between the LE-GT segment and the vertical straight line passing through GT.

$$\theta_{hip} = \arccos\left(\frac{GT - LE}{|GT - LE|} \cdot (0, 1)\right)$$

Each gait cycle can be made of a different number of frames so the resulting angles were referred to as the percentage of the gait cycle (from 0% to 100%) through spline interpolation. As a result, three distinct joint angles (JA) curves are produced. By linearly weighting the static model (S), load model (L) and swing or flex model (F) joint angles based on the phases of the gait cycle a single resulting joint angle curve is calculated.

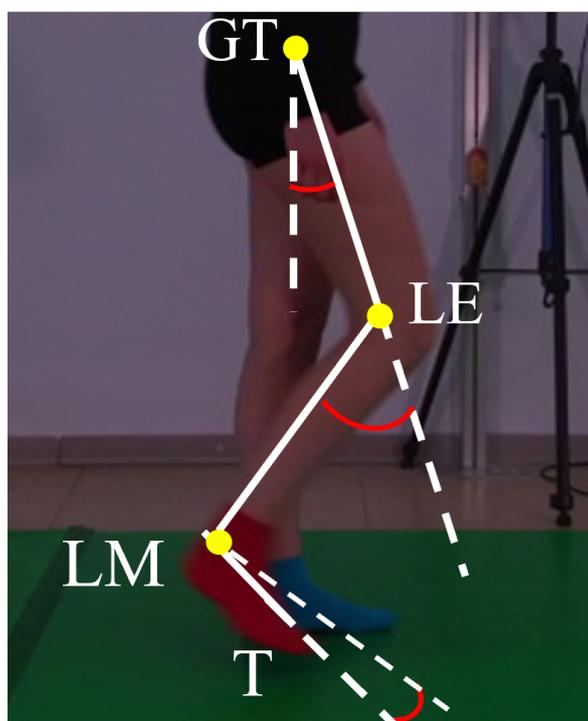


Figure 6.8: Illustration of the hip, knee and ankle angles of the foreground limb calculated starting from the GT, LE and LM 2D coordinates and the foot inclination (T) extracted through the MLM method

6.2 Body Tracking SDK

The Azure Kinect body tracking software development kit (SDK) [24] was developed by Microsoft and is capable to track the position of one or more human bodies in real-time. Each body in the field-of-view has its unique ID number, its segmentation map and its skeleton whose position and orientation can be extracted and visualized in real-time (Figure 6.9).

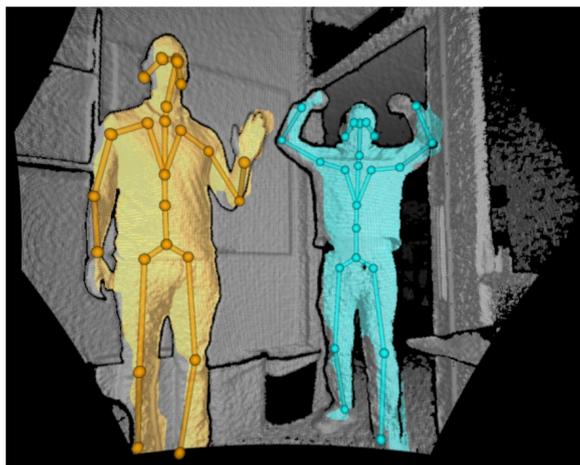


Figure 6.9: Illustration of real-time body tracking performed with the Azure Kinect SDK. Figure from: Microsoft Azure Kinect DK documentation

The skeleton consists of 32 joints which are shown in figure 6.11. For this study, only 9 joints belonging to the lower limbs were considered. The body tracking SDK provides the 3D position and orientation relative to the depth sensor reference frame for each joint as well as its confidence level.

The body tracker is a deep-learning-based algorithm which uses a Convolutional Neural Network (CNN) and a model-fitting algorithm to extract the skeleton 3D coordinates for each body in the FOV.

An overview of the architecture of the 3D Skeletal Tracking of Azure Kinect is represented in figure 6.10. Starting from the IR image captured by the camera, the body segmentation map and the extraction of the 2D body skeleton are performed through a CNN.

The CNN algorithm is capable to identify the 2D positions of the body joints starting from the input image and extracting a set of features which will be used to identify the different parts of the skeleton. The CNN algorithm was trained both on real images and on a large data set of synthetic images taken from videos of artificially simulated human bodies doing different movements. Since they are computer simulations the exact positions of the anatomical landmarks of interest are precisely known and the artificial intelligence algorithm can be successfully trained to recognize them. Another helpful factor of using synthetic images is that the network can be trained on many different images considering different movements,

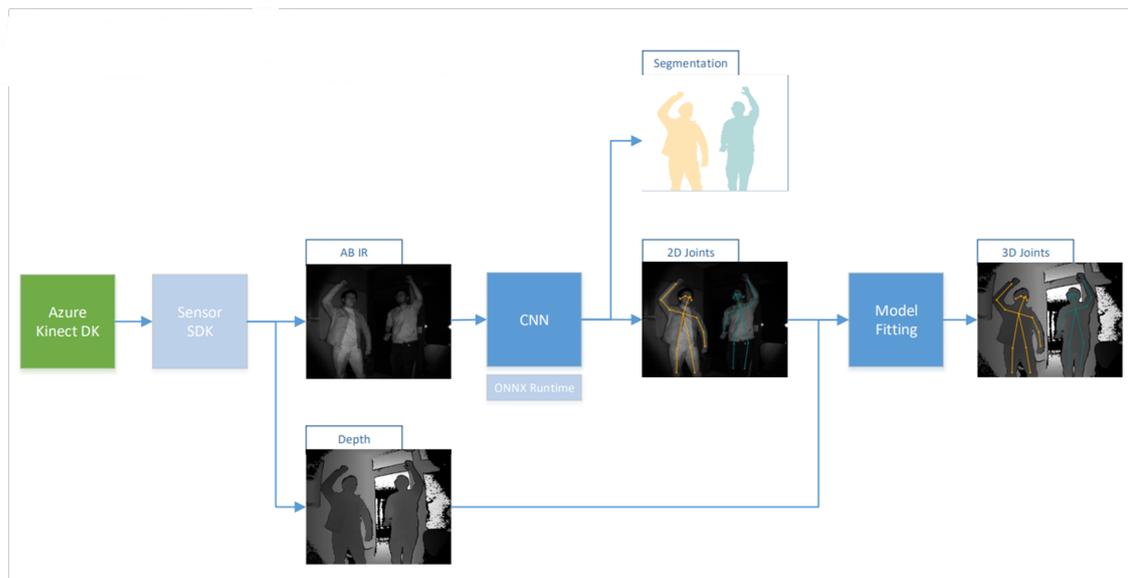


Figure 6.10: Illustration of the Body Tracking SDK system architecture. Figure from: 3D Skeletal Tracking on Azure Kinect

very different scenarios but also the physical variability of the human body.

The estimated joint coordinates in 2D are integrated with the distance information obtained from the depth image to compute the 3D joint coordinates and orientations through model fitting [24]. The model fitting method takes into account the rigid transformations of a real body, the temporal coherency, the anatomical joint limits and other regularization terms to compute the 3D coordinates of the joints' position, their orientation and their time information.

The most accurate estimations are obtained in the frontal view of the body with no occlusions in the FOV. The most challenging cases are the ones when the depth information is not highly reliable when the subject is wearing dark clothes which are IR absorbing, when there are occlusions and when the body is in a lateral position or can only be partially seen.

Body Tracking SDK Application An application for using the Body Tracking SDK can be developed using Windows or Linux platforms. A set of libraries is available for its implementation with C or C++ programming languages. To build an application that tracks the position and orientation of one or more bodies with a single Azure Kinect camera, first, it is necessary to download and install:

- The Azure Kinect Sensor SDK [27], which contains all the headers and libraries needed to operate with the Azure Kinect DK.
- The Body tracking SDK [28], which includes all the headers and libraries to build a track-

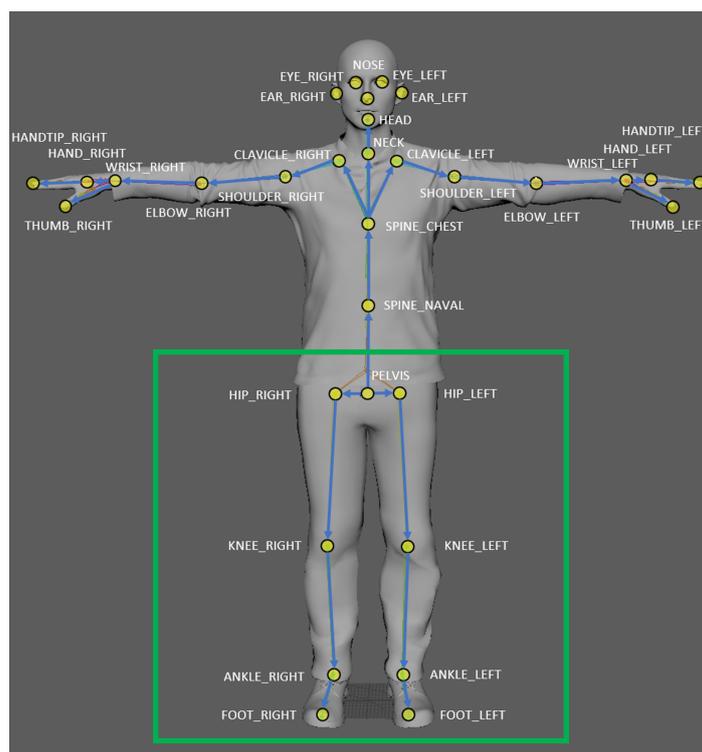


Figure 6.11: Illustration of the joint positions and connections provided by the Azure Kinect SDK. The rectangular box includes the joints under investigation in this study. Figure adapted from: Microsoft Azure Kinect DK documentation

ing application.

To create a simple application that performs the body tracking processing in real-time it is necessary to execute, in order, the following steps:

1. Open and configure the device. Different depth modes (e.g. OFF, NFOV, WFOV, binned or unbinned) and color resolutions (e.g. OFF, 720p, 1080p) in which the cameras will operate can be selected
2. Start the camera recording
3. Calibrate the sensor and then create a body tracker
4. Get the captures from the device and add them to the tracker input queue
5. Extract the results from the output queue, this operation gives a body frame structure as an output
6. Process the data, this means to access the body tracking information stored in the body frame structure which are:
 - The full body structures, which contain the number of recognized bodies and their body ID, the joint positions as 3D coordinates and the joint coordinate frames' orientations in quaternions

- The 2D body index map which contains the bodies segmentations, separated from the background
- The input capture from which the body tracking information was obtained



Figure 6.12: Illustration of the information contained in a body structure. Figure adapted from the Microsoft Azure Kinect DK documentation

7. Turn off the body tracker so that no more captures are added to the processing queue and release the body tracker.
8. Stop and close the device

MATERIALS AND METHODS

7.1 Experimental Protocol and Setup

The experimental protocol for the investigation of the gait cycle included the acquisitions of 5 different gait cycles for each side in 5 healthy subjects, who gave their informed consent. First, the measurements have been performed synchronously with the optical stereophotogrammetric system (Vicon Motion Systems [29]) and the Azure Kinect camera. Then the acquisitions have been performed only with the Azure Kinect as interferences were observed in the depth images from the Azure Kinect since the wavelength of the infrared sensor is the same as the SP system (850 nm). This issue will be better explained in the second section of this chapter. The experimental set-up included:

- Optical stereophotogrammetric system with 12 infra-red cameras (Vicon Vero) and 3 RGB cameras.
- One large green carpet
- One RGB-Depth camera (Azure Kinect) positioned at 85 cm height and approximately 2.5 m from where the subject was required to walk
- Two LED lamps are set to the maximum intensity and positioned at 120 cm from the Azure Kinect
- A lux meter to measure the illuminance in the FOV where the subject is walking, the measured result was 517 lux

The calibration of the Vicon system was performed in default condition with the camera strobe intensity and gain set to 1. This operation included three steps: the *mask cameras* operation, to remove any unwanted light reflections; the calibration of the cameras conducted by moving an active wand (Figure 7.2) in the capture volume to show the fixed lights to all the Vicon cameras; *Set volume origin* in which the active wand was placed on the floor to define the origin of the global reference system.

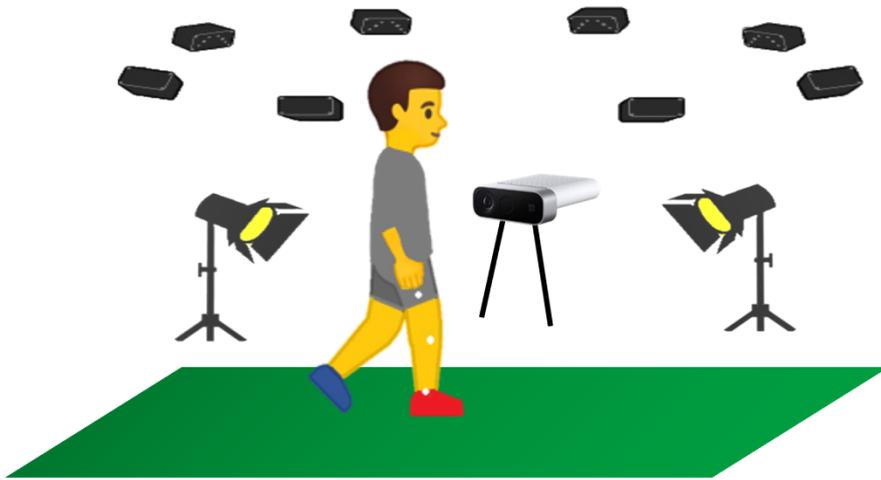


Figure 7.1: Illustration of the experimental set up



Figure 7.2: Active Wand for the optical stereo-photogrammetric system calibration

After the camera calibration, 16 reflective markers were placed on the participants' lower limbs with some asymmetries in the markers' disposition according to the lower limb plug-in gait 7.3. The subject was asked to wear shorts, one red sock on the right foot and one light blue sock on the left foot and to walk on the green carpet in the central area of the capture volume.

After the subject preparation, the subject is asked to stand in front of the RGB camera for one static acquisition using the Vicon system. The acquisitions were performed in the following order, first synchronously with the Vicon cameras and the Azure Kinect on:

1. One static upright posture
2. Five gait cycles with the right leg in the foreground
3. Five gait cycles with the left leg in the foreground

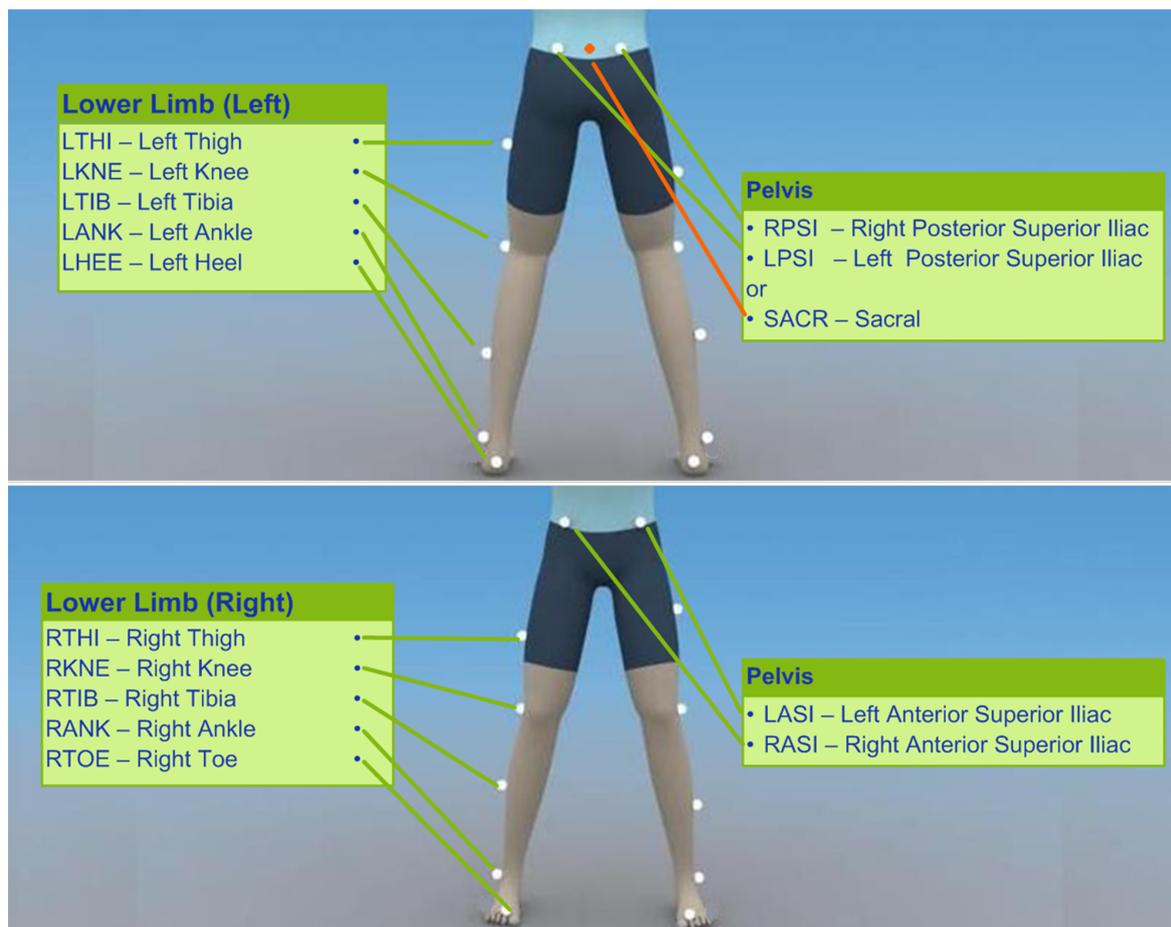


Figure 7.3: Illustration of the marker placement for the acquisitions with the Vicon system. Figures adapted from Vicon - lower boy plug-in-gait

The Vicon system was turned off and the markers were removed from the subject's limbs for the subsequent acquisitions which were conducted only with the Azure Kinect:

1. One static upright recording with the right leg in the foreground
2. One static upright recording with the left leg in the foreground
3. Five gait cycles with the right leg in the foreground
4. Five gait cycles with the left leg in the foreground

7.2 Infrared interference Issues

The acquisitions were not performed synchronously by the marker-based and markerless systems as interferences in the depth map reconstruction were observed in the Azure Kinect recordings. The wavelength of the Azure Kinect IR sensor is the same as the Vicon Vero system (850 nm) and this resulted in very poor quality depth images with many invalidated pixels whose value was either a zero or not-a-number (NaN). To reduce the interference the strobe intensity of the Vicon IR cameras was set to 0.08 and the gain was set to 4 after the system calibration. Nonetheless, black uninformative pixels were still present in the synchronous acquisitions, particularly in correspondence with the positions of the reflective markers. As can be seen in figure 7.4 the depth information about the foreground leg, which is the leg under investigation is very poor, this affected the joint centre's position estimations. Thus, the data processing for the markerless methods was made on different acquisitions that were performed without any markers and after the optical stereophotogrammetric cameras were turned off.

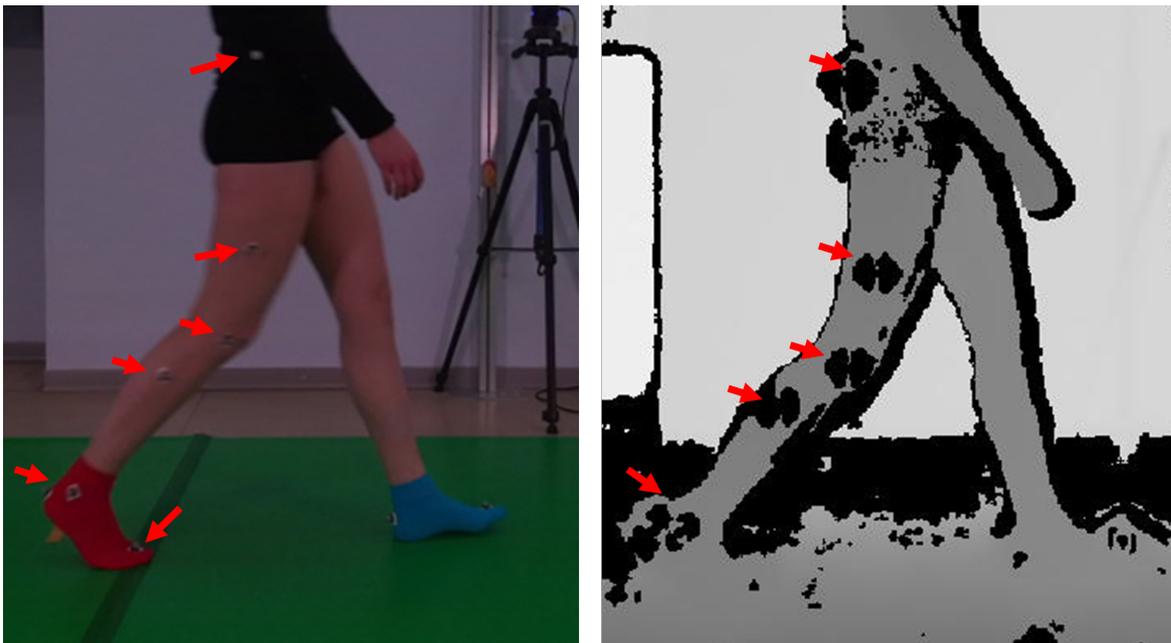


Figure 7.4: Illustration of the RGB (left) and the depth (right) images captured with the Azure Kinect during synchronous recording with the marker-based system. Black invalidated pixels can be seen particularly in correspondence of the reflective markers (red arrows).

7.3 Data Processing

The performances of the two methods (SDK and MLM) described in the previous chapter are compared in the lower limb joint angles estimation.

The joint angles on the markerless recordings are extracted through the SDK and MLM method while the joint angles on marker-based recordings are extracted through the Vicon-Nexus software.

Seven significant gait variables are identified from the joint angles. Finally, to perform the methods comparison the SDK and the MLM gait variables differences with respect to the MB gait variables are evaluated. The entire processing is summarized in figure 7.5.

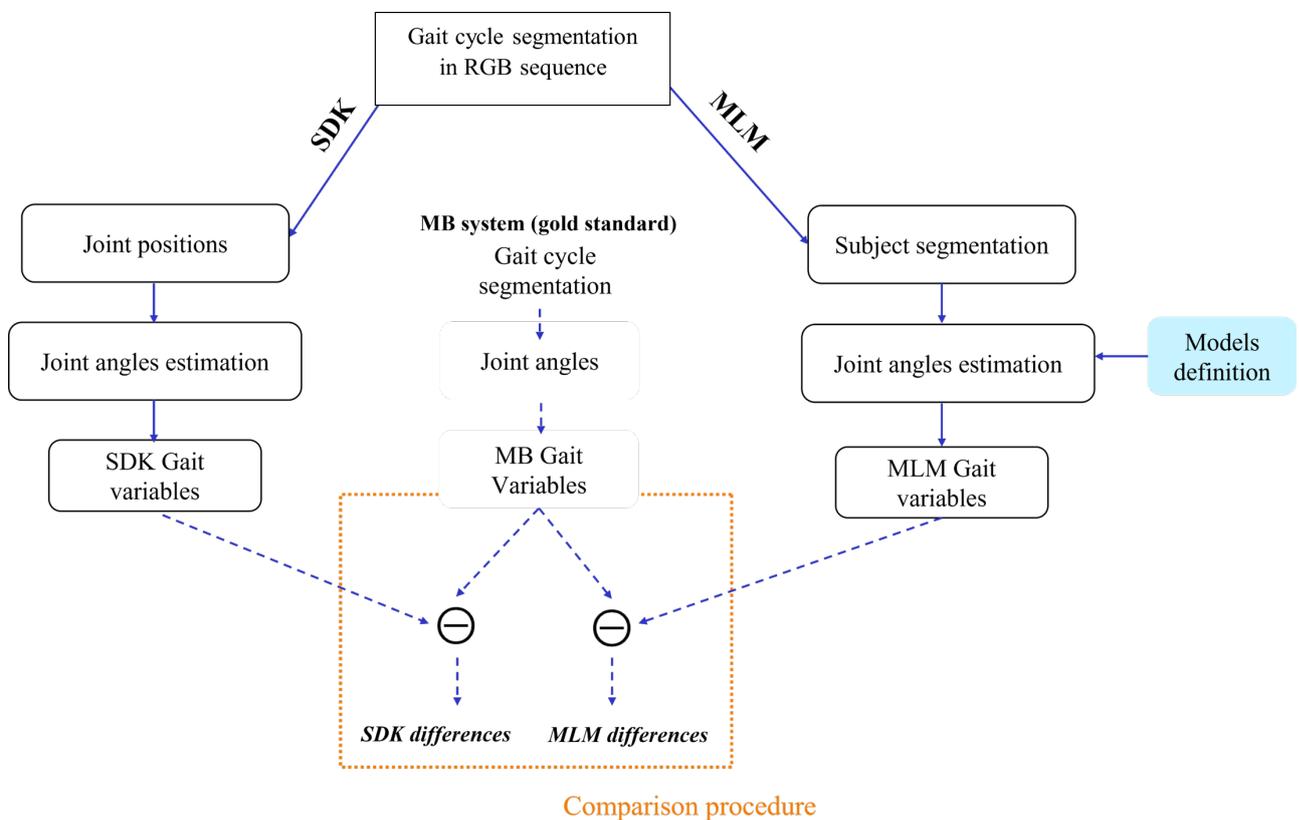


Figure 7.5: Overview of the methods.

Marker-based (MB) system

The gait analysis was conducted using the optical stereo-photogrammetric system made of 12 IR cameras (Vicon Vero) with a 100 frames/s frame rate. 16 reflective markers were attached to the subject's body according to the lower-limb plug-in gait model (Figure 7.3). Data were processed with the software Vicon Nexus to compute the Hip, Knee and Ankle joints angles in the sagittal plane.

An overview of all the steps performed for the extraction of the joint angles through the MB system in each trial is shown in Figure 7.6. First, the subject reconstruction was conducted for the static capture then each marker was manually labelled. To run the plug-in-gait processing functions, the participants' anthropometric parameters were entered. The measured parameters included the subject's body weight, height, leg length, knee width and ankle width. After the static capture processing, for each one of the dynamic captures the following steps were performed:

- Run of the *Reconstruct and Label* pipeline in which the markers are automatically identified in each frame.
- A single gait cycle was manually isolated by selecting the initial contact (IC) as the first floor contact of the foreground foot and the final contact (FC) as the subsequent foot-floor contact of the same foot.
- Run of the *Plug-in Gait Dynamic* pipeline. The Plug-in Gait directly computes the kinematic model's joint centres from the markers' 3D coordinates measured in each frame [30]. An example of the output of this step is shown in figure 7.8.
- Selection and export of the three outputs of interest in each gait cycle which are the:
 - Hip angle between pelvis and thigh.
 - Knee angle between thigh and shank.
 - Ankle angle between shank and foot.

The outputs included 5 arrays containing the above-mentioned angles for the left lower limb in each frame of a single gait cycle and 5 arrays for the right lower limb. Each gait cycle was made of a different number of frames, so the resulting angles were referred as to the percentage of the gait cycle (from 0% to 100%) through spline interpolation.

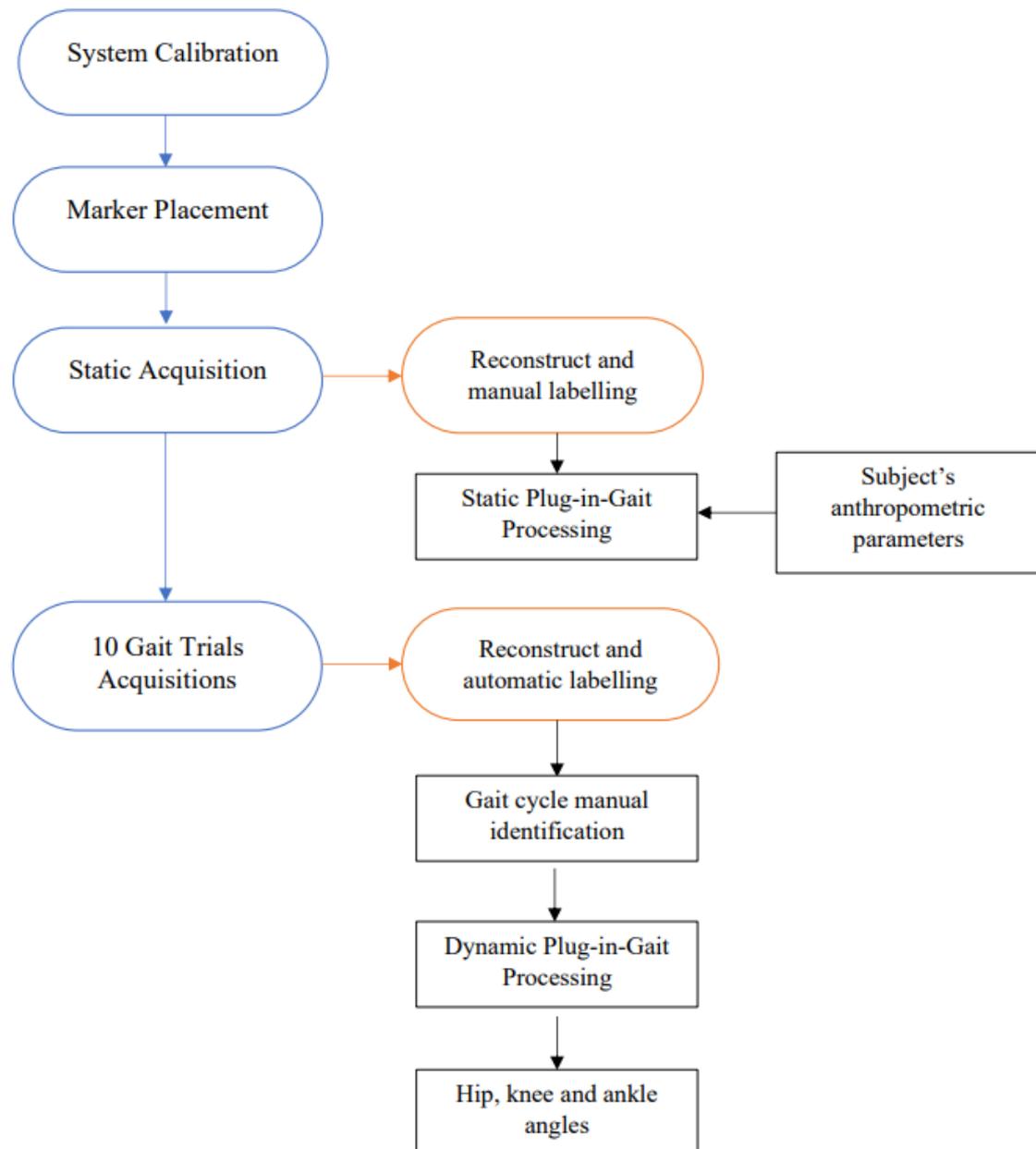


Figure 7.6: Illustration of the recording and processing procedure adopted for the extraction of the joint angles from the marker-based recordings.

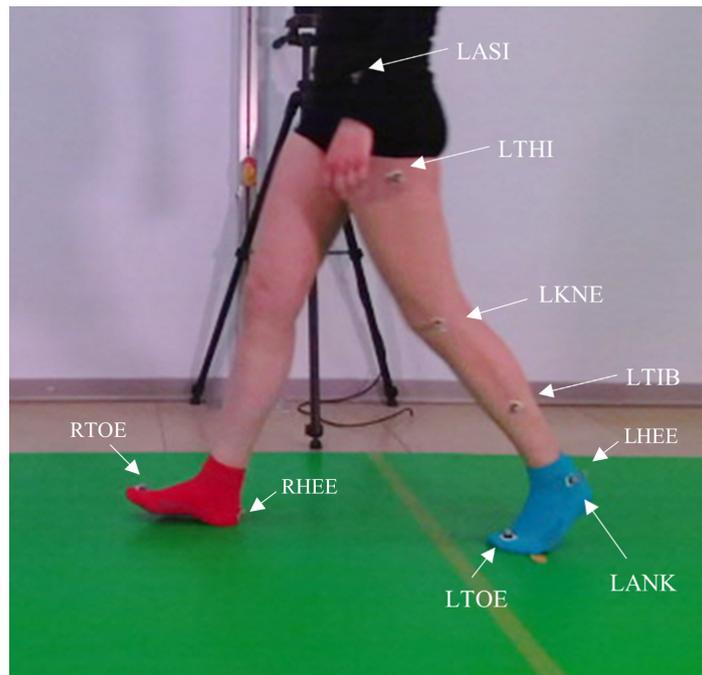


Figure 7.7: Example of a left trial recording with the marker-based system. Each marker relative to the anatomical landmarks is labelled according to the plug-in gait lower limb model.

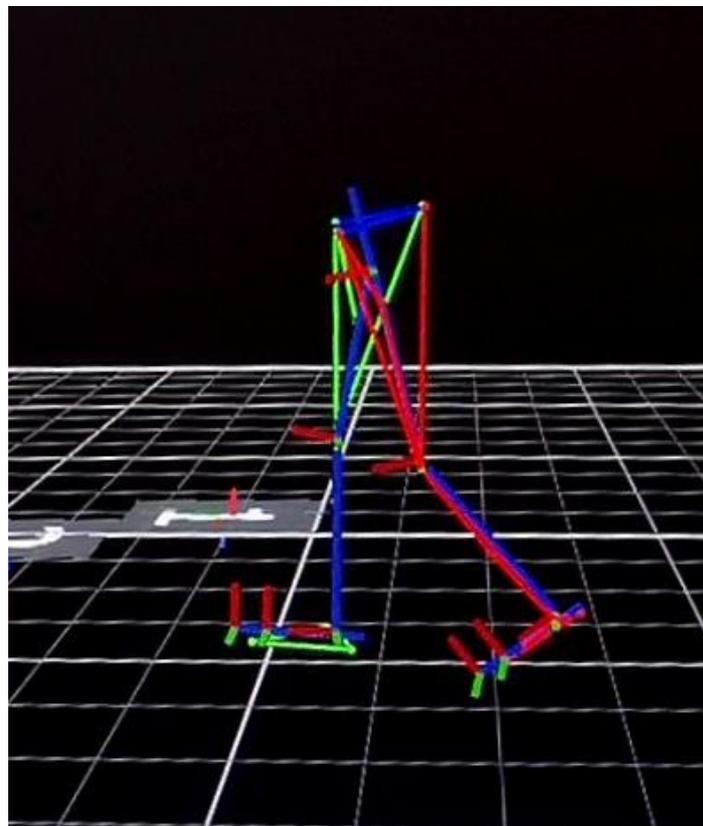


Figure 7.8: Example of the plug-in-gait modelling outcome with the Vicon-Nexus system.

Markerless Method (MLM)

The markerless method described in the previous chapter was implemented in MATLAB (release R2021b) for each of the 5 subjects in each trial (5 for the left lower limb and 5 for the right lower limb). The processing steps for each trial are listed below:

1. *Gait Cycle segmentation*: A single gait cycle is manually identified on the RGB images. The frame of the initial contact (IC) was identified as the first floor contact of the foreground foot while the final contact (FC) was identified as the subsequent foot-floor contact of the same foot, as shown in figure 7.9.

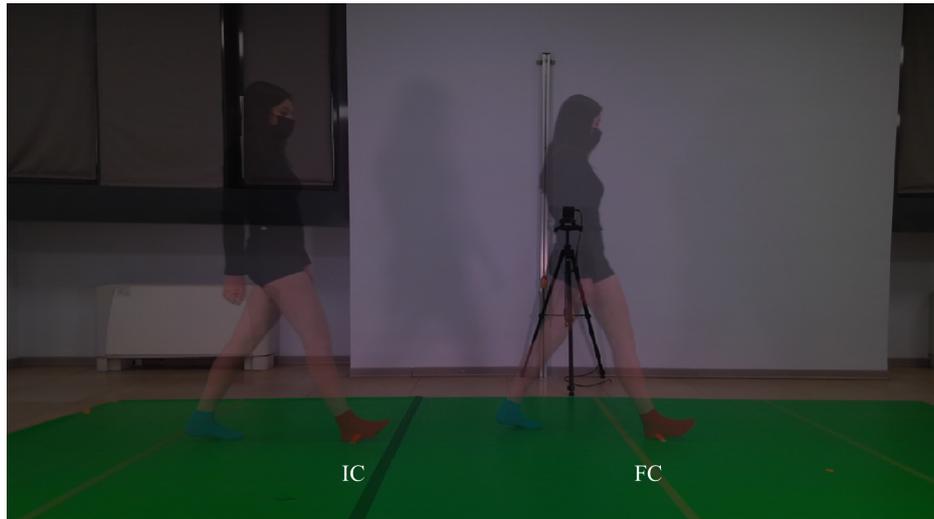


Figure 7.9: Examples of foot-floor contact in the initial contact (IC) and final contact (FC) frame selected for the identification of a single gait cycle.

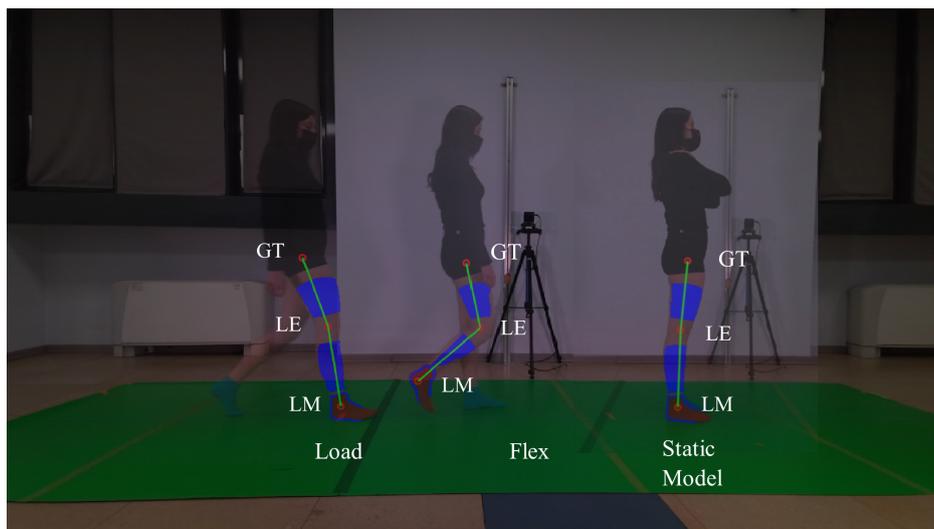


Figure 7.10: Examples of frames selected for the load model, flex (or swing) and static model definition for one of the right limb trials.

2. *Model Definition*: Three frames were manually selected for the creation of the subject-specific lower limb kinematic models (Figure 6.3). One frame was selected from the static acquisition of the foreground limb while from each dynamic trial one frame was selected during the load phase and one frame during the swing phase as shown in figure 7.10 5.5. This operation included the manual identification of the anatomical landmarks (LM, LE and GT) for each of the three models. The static model was selected either from the right and the left limb static recordings accordingly to the dynamic trial under investigation.
3. *Lower Limb joint kinematics estimation*: The MLM method described in the previous chapter was implemented to reconstruct the joint angles trajectories of the hip, knee and ankle joints in the sagittal planes based on the three subject-specific kinematic models. As a result, ten joint angles trajectories (5 for right and 5 for left limb), which are shown in the results chapter, are obtained for each subject.

Body Tracking SDK

The body tracking information from the SDK software was extracted from the video recordings of the Azure Kinect of the walking subject in *mkv* format by using the *offline_processor* sample from the body tracking samples [31] provided by Microsoft.

This algorithm enables to open a recording *mkv* file, run through the body tracking SDK ([28]) and store the results of joints coordinates in the 3D depth image space to a *json* file. The 3D joint position results were converted in the 2-dimensional colour image space, for this purpose, the function *k4a_calibration_3d_to_2d* provided in the Body Tracking SDK documentation, was implemented. From the results of the 32 joint positions in the colour image space, only the positions of the joints of interest were selected. Only the *HIP*, *KNEE*, *ANKLE* and *FOOT* coordinates for the limb in the foreground were selected, as shown in figure 7.11.

The lower limb joint angles were computed as follows:

- The ankle joint angle is calculated as the inclination between the ANKLE-FOOT segment and the ANKLE-KNEE segment minus 90 degrees.

$$\theta_{ankle} = \arccos\left(\frac{KNEE - ANKLE}{|KNEE - ANKLE|} \cdot \frac{FOOT - ANKLE}{|FOOT - ANKLE|}\right) - 90^\circ$$

- The knee joint angle is calculated as the angle between the ANKLE-KNEE segment and the KNEE-HIP segment

$$\theta_{knee} = \arccos\left(\frac{KNEE - HIP}{|KNEE - HIP|} \cdot \frac{ANKLE - HIP}{|ANKLE - HIP|}\right)$$

- The hip joint angle is defined as the angle formed between the KNEE-HIP segment and the vertical straight line passing through the HIP coordinate.

$$\theta_{hip} = \arccos\left(\frac{HIP - KNEE}{|HIP - KNEE|} \cdot (0, 1)\right)$$

Each gait cycle was made of a different number of frames so the resulting angles were referred to as the percentage of the gait cycle (from 0% to 100%) through spline interpolation. The SDK resulting angles were filtered with a 4th-order lowpass digital Butterworth filter with a cut-off frequency of 5 Hertz.

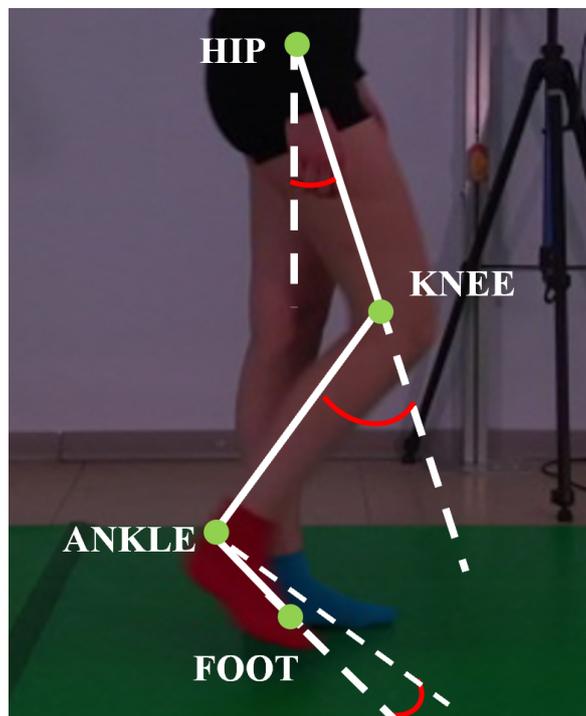


Figure 7.11: Illustration of the hip, knee and ankle angles of the foreground limb calculated starting from the HIP, KNEE, ANKLE and FOOT 2D coordinates extracted through the SDK method

Gait Variables

For assessing the performance of the methods, the knee, ankle and hip flexion/extension angles were estimated with the marker-based (MB), markerless model-based (MLM) and Body tracking SDK methods and seven gait parameters were extracted [32], [9] for each trial:

- K1: the knee flexion at the initial contact (0% of the gait cycle)
- K2: the knee maximum flexion during the loading response (0 - 40% of the gait cycle)
- K3: the knee maximum extension during the stance phase (25- 75% of the gait cycle)
- K5: the knee maximum extension during the swing phase (50 - 100% of the gait cycle)
- A3: the ankle maximum dorsiflexion during the stance phase (25 - 75% of the gait cycle)
- A5: the ankle maximum dorsiflexion during the swing phase (50- 100% of the gait cycle)
- H3: the hip maximum extension during the stance phase (25 - 75% of the gait cycle)

Results were compared by calculating the mean difference obtained from the SDK and the MLM method with respect to the marker-based gold standard (GS).

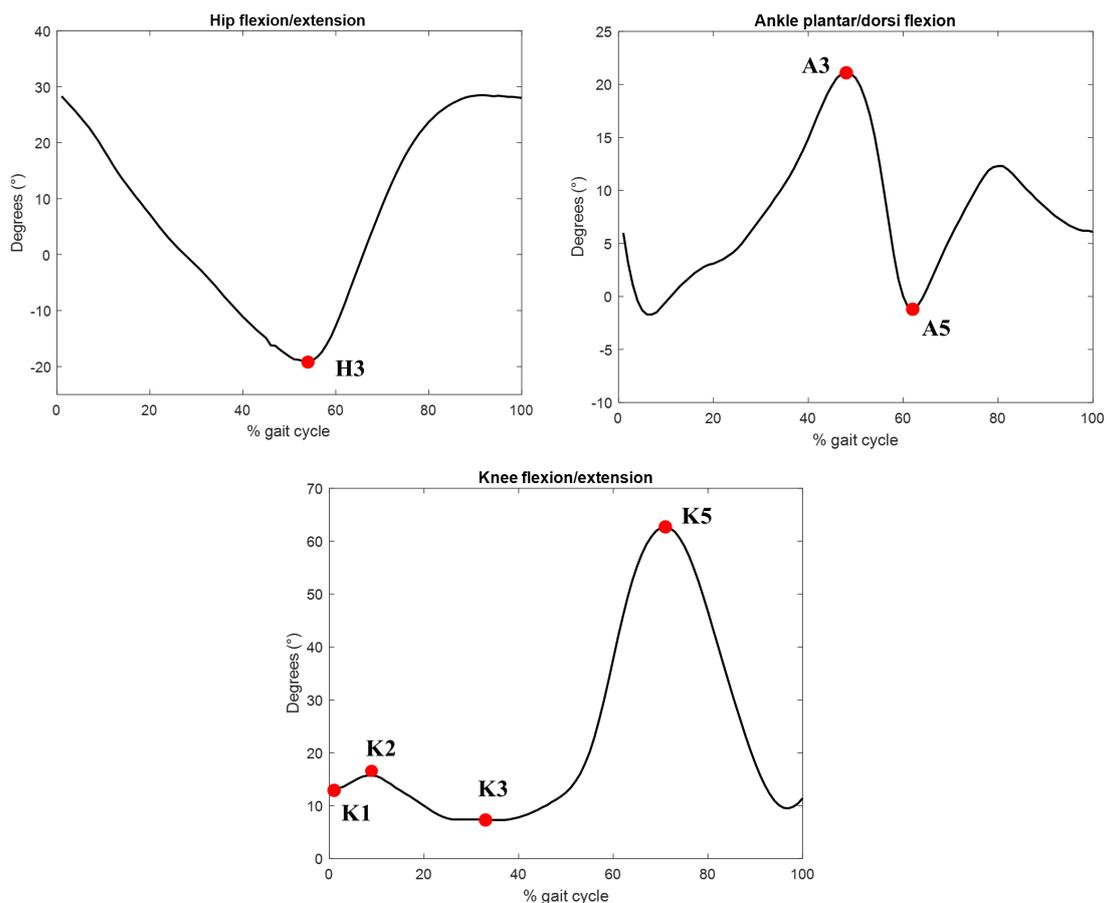


Figure 7.12: Illustration of the seven gait variable extracted from the hip, knee and ankle angles in the sagittal plane for the comparison of the methods.

RESULTS AND DISCUSSIONS

The results of the lower limb joints 2D coordinates estimated with the two merkerless methods (MLM and SDK) were superimposed to the color frames as shown in figure 8.1 and 8.2.



Figure 8.1: Illustration of the joint positions resulting estimations through the MLM method for some frames of a right trial.

After a visual inspection of the results obtained with the body SDK method, it was observed that in two out of five subjects there was some misidentification in the hip, knee, ankle and foot coordinates. In some frames the foreground limb was incorrectly identified as the background limb, an example is depicted in Figure 8.3. This issue was present in 0.5% and 2.3% of the total number of frames of the ten gait trials of the two subjects and inevitably affected the SDK joint angles estimations.

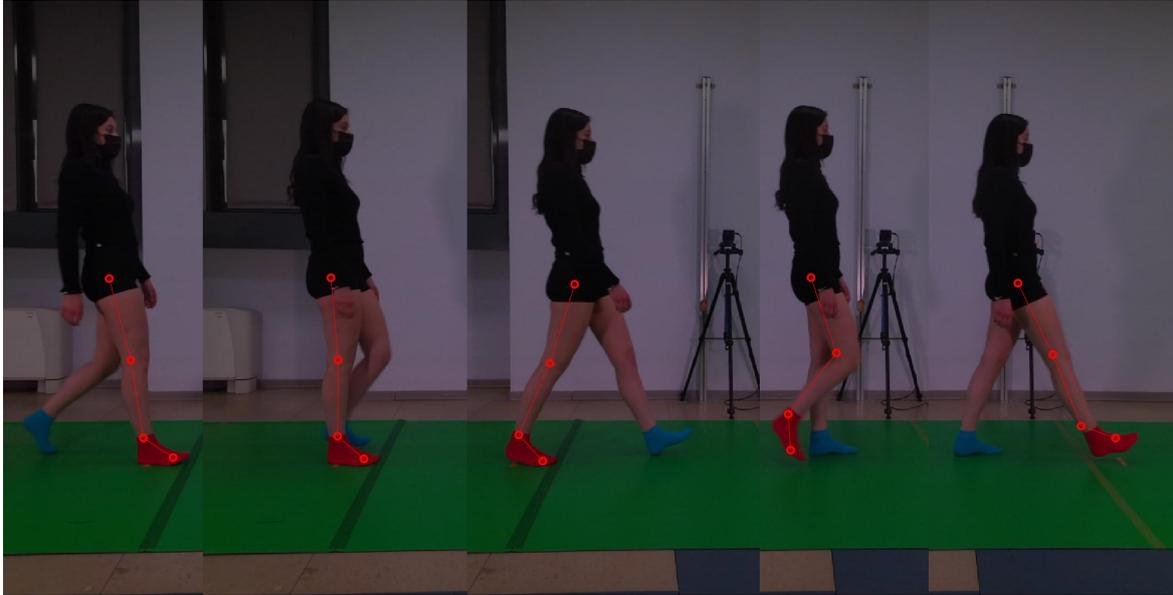


Figure 8.2: Illustration of the joint positions resulting estimations extracted through the SDK method for some frames of a right trial.



Figure 8.3: Example of wrong leg identification with the body tracking SDK method. In this frame the supposed left hip, knee, ankle and foot coordinates were misidentified as the right limb ones.

The resulting hip, knee and ankle angles in the sagittal plane estimated for ten gait trials for one subject are reported in figure 8.4 as a percentage of the gait cycle. The estimation was performed with the marker-based (MB) method and the two markerless methods (MLM and SDK). While the markerless method curves were estimated on the same gait trials for each subject, the marker-based curves were computed on ten different gait trials, unsynchronized with respect to the markerless ones due to the infrared interference issues mentioned in Chapter 7. The mean values and the standard deviation SD shown in table 8.2 result from the gait parameters calculated from the hip, knee and ankle angles in the sagittal plane averaged over 10 gait cycles per 5 participants.

The overall differences between the gait parameters estimated with the marker-based method and the gait parameters estimated with the body tracking SDK method ranged from -0.7° in the estimation of the knee max flexion in the load phase to 39.6° in the estimation of the ankle max dorsiflexion in swing phase. The differences between the gait parameters estimated with the marker-based method and the gait parameters estimated with the model-based markerless method (MLM) method ranged from -0.6° in the estimation of the knee max extension in the stance phase to -11.2° in the estimation of the ankle max dorsiflexion in stance phase.

The standard deviation of the gait parameters grand mean ranges from 2.5° to 4.8° for the MLM method and from 2.6° to 8.5° for the SDK method.

The variability was high for the ankle in the swing phase both for the MLM and the MB methods with SD values of 6.9° and 8.5° , respectively. The lower variability was found for the MB method in the knee parameter during the swing phase ($SD = 2.6^\circ$), for the MLM method in the ankle in stance phase ($SD = 2.6^\circ$), and the SDK method in the knee parameter at the initial contact ($SD = 2.5^\circ$).

		<i>ICC MLM</i>	<i>ICC SDK</i>
Knee	Initial contact	0.60	0.52
	Load	0.81	0.58
	Stance	0.31	0.48
	Swing	0.39	0.19
Ankle	Stance	0.60	0.37
	Swing	0.85	0.36
Hip	Stance	0.49	0.80

Table 8.1: Reliability of the model-based markerless (MLM) and body tracking SDK methods computed for each gait variable

The reliability of the methods is evaluated through the intraclass correlation coefficient (ICC) which is shown in figure 8.1. For each parameter p the ICC for the MLM and SDK methods were computed from the standard deviation SD of the MB values and the SD of

the MLM and SDK differences with respect to MB as follows:

$$ICC_{MLM,p} = \frac{SD_{MB,p}^2}{SD_{MB,p}^2 + SD_{MLM-MB,p}^2}$$

$$ICC_{SDK,p} = \frac{SD_{MB,p}^2}{SD_{MB,p}^2 + SD_{SDK-MB,p}^2}$$

The ICC values found for the MLM method range from 0.31 for the knee extension in the stance phase to 0.85 for the ankle flexion in the swing phase. For the SDK method, the ICC minimum is 0.19, obtained for the knee extension in the swing phase and the max is 0.80 obtained for the hip extension in the stance phase.

To compare the methods estimation the five number statistics for each gait parameter under investigation were evaluated over the average of ten gait trails per five subjects and it is shown in figure 8.5.

The MLM method showed large interquartile (IQR) ranges with respect to the MB and the SDK method in the max knee extension in the stance phase. The SDK method showed large interquartile (IQR) ranges for the knee parameters with respect to the MB and the MLM method in the max ankle flexion in the stance phase.

The overall distances of the MLM median values from the MB ones are lower than the SDK median distances from the MB ones for every gait parameter except for the max knee flexion at the initial contact and the max knee extension in stance. The symmetry in the boxplots of the gait variable is not highly present. For the MB method, only the data of the max ankle flexion in the stance phase and the max knee extension during swing showed symmetry. For the SDK method, only the data of the max hip flexion in the stance phase showed symmetry. And for the MLM method, only the data of the max ankle flexion in the swing phase showed symmetry but an outlier was found in the MLM max ankle flexion estimates.

Both the SDK and MLM methods provided good estimates for knee and hip kinematics while higher differences with respect to the marker-based system are provided for ankle kinematics, mainly for the SDK.

The largest differences between the SDK Body Tracking and the MB results can be observed especially in the values of the ankle angle during the gait cycle. This is mainly because the positions of the ankle and foot from the SDK are not a good representation of the actual inclination of the foot with respect to the subject's shank as shown in figure 8.6.

The resulting waveforms of the ankle angles from the SDK joint centres estimations were substantially different from the marker-based ones as shown in figure 8.7, particularly during the stance and swing phase where the lateral malleolus position was visibly misidentified.

Gait Variables (deg)	<i>MLM mean (SD)</i>	<i>SDK mean (SD)</i>	<i>MB mean (SD)</i>	<i>MLM Difference</i>	<i>SDK Difference</i>
Knee					
Initial contact	10 (5.2)	7.4 (2.5)	6.1 (4)	3.9	1.3
Load	18.2 (6)	13.5 (6)	14.2 (4.8)	4.0	-0.7
Stance	5.7 (5.9)	7.7 (3)	6.3 (3.1)	-0.6	1.4
Swing	64.5 (4.6)	59.7 (4.8)	64 (2.6)	0.5	-4.3
Ankle					
Stance	7.2 (2.6)	46.5 (4)	18.4 (3.2)	-11.2	28.1
Swing	-23.4 (6.9)	25.7 (3.6)	-13.9 (8.5)	-9.5	39.6
Hip					
Stance	-12.4 (3.8)	-16.8 (3.1)	-9.1 (4.6)	-3.3	-7.7

Table 8.2: Grand mean and standard deviation of the gait parameters over ten trials per five subjects computed for the marker-based (MB), model-based markerless (MLM) and body tracking SDK methods. Differences of the two markerless methods estimations with respect to the MB method.

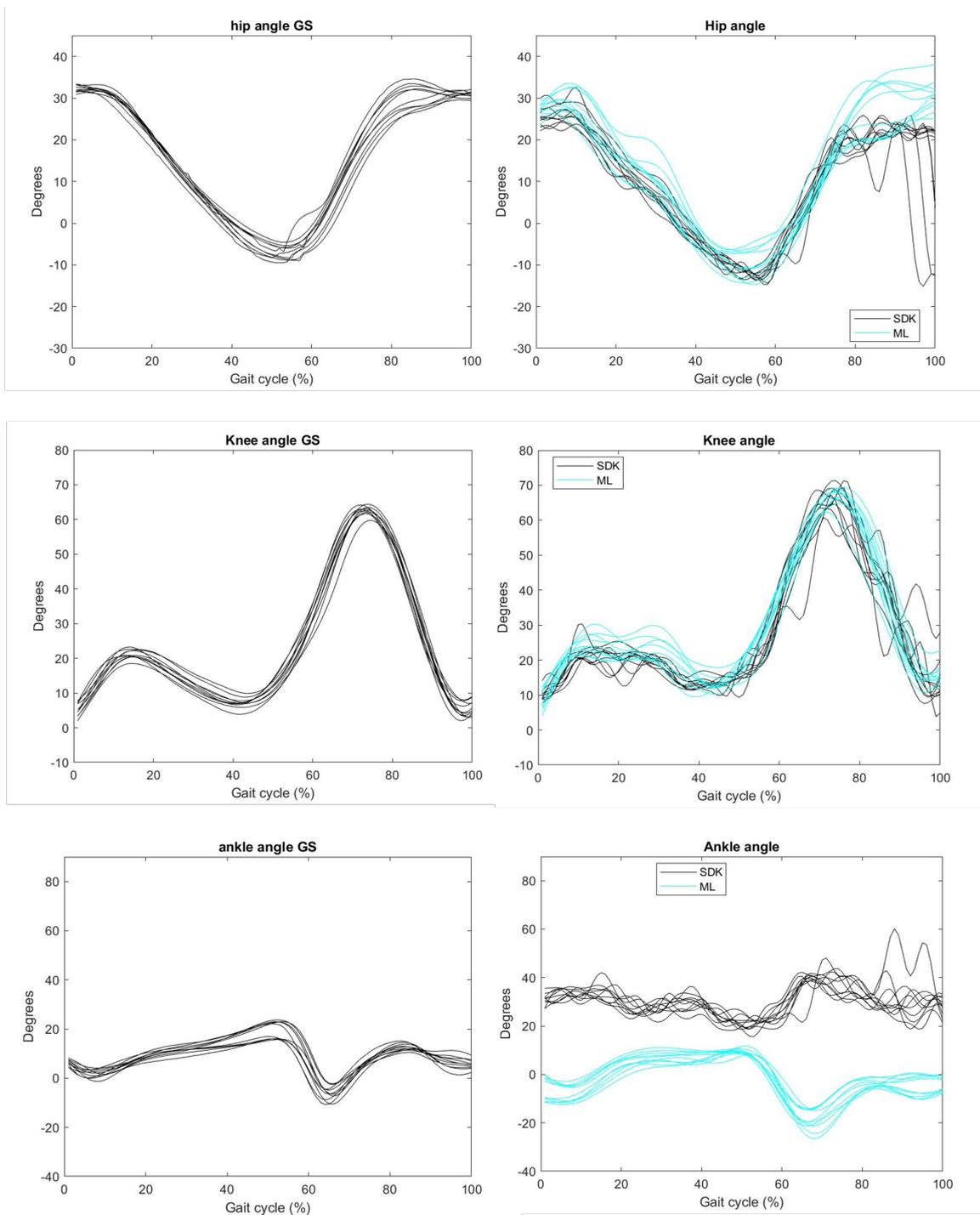


Figure 8.4: Illustration of hip, knee and ankle angles in the sagittal planes of ten gait trials of a single subject estimated with the marker-based (MB), model-based (MLM) and body tracking SDK methods. While the markerless angles result from synchronous recordings the MB angles were extracted from different recordings of the same subject with markers attached to the skin.

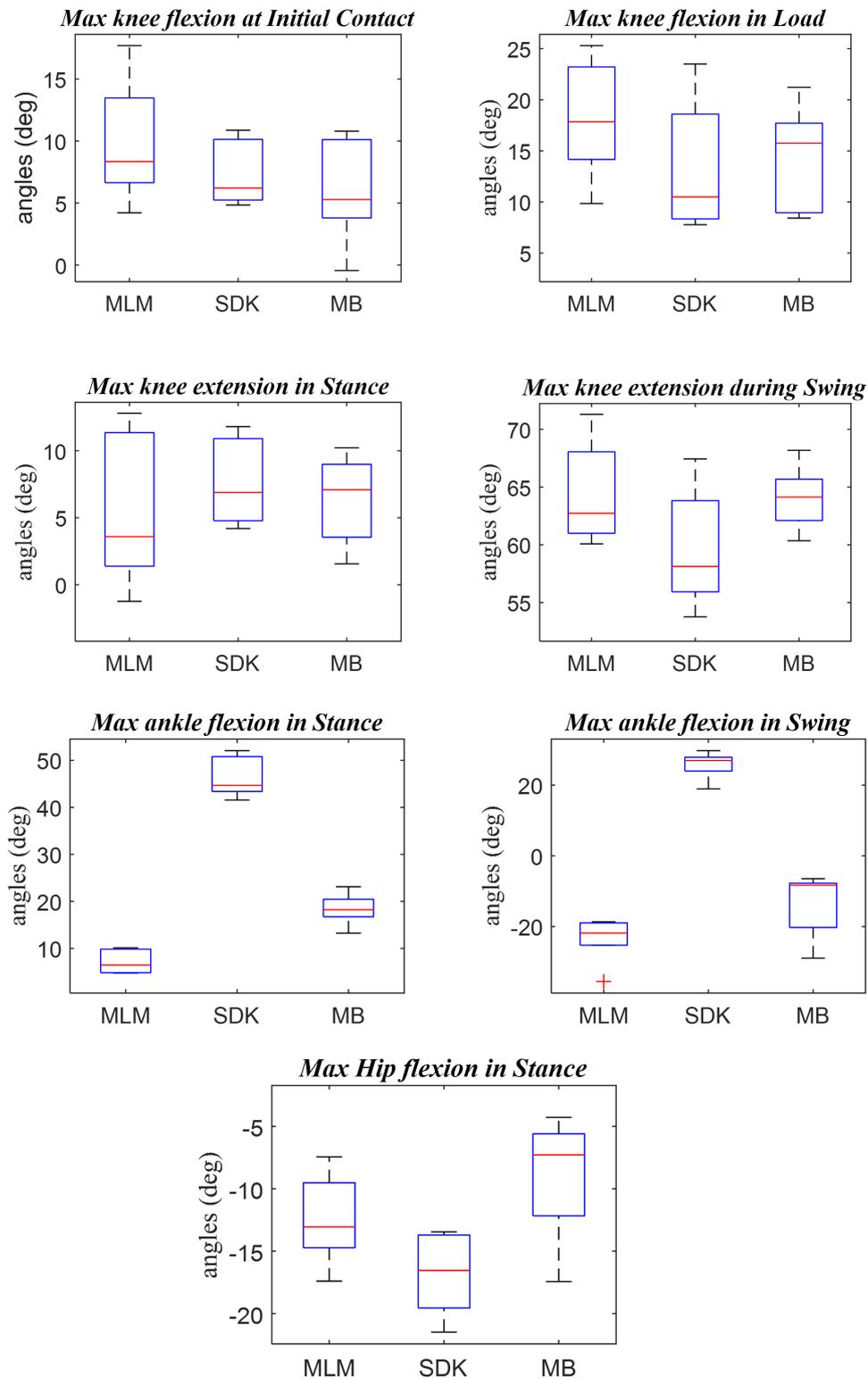


Figure 8.5: Five summary statistics for the seven gait variables extracted from the joint angles computed with the marker-based (MB), model-based markerless method (MLM) and body tracking SDK computed for each gait variable.

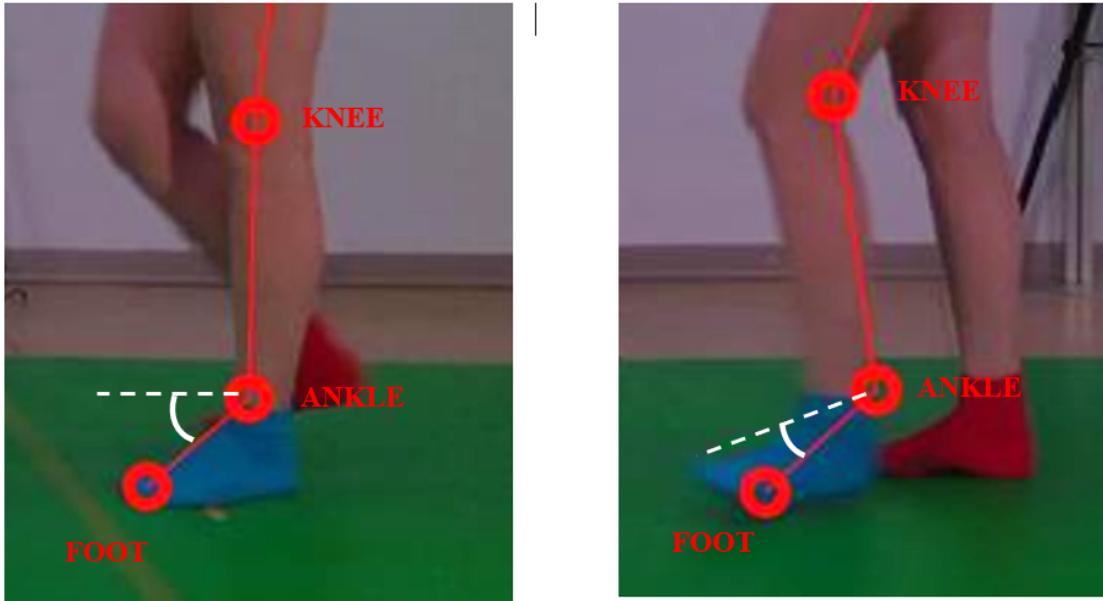


Figure 8.6: Examples of ankle angle evaluations from the coordinates extracted with the SDK method in two different phases of the gait cycle, during the stance (left) and swing (right) phase. The coordinates of the ankle, foot and knee are visibly misidentified resulting in incorrect ankle angle estimations.

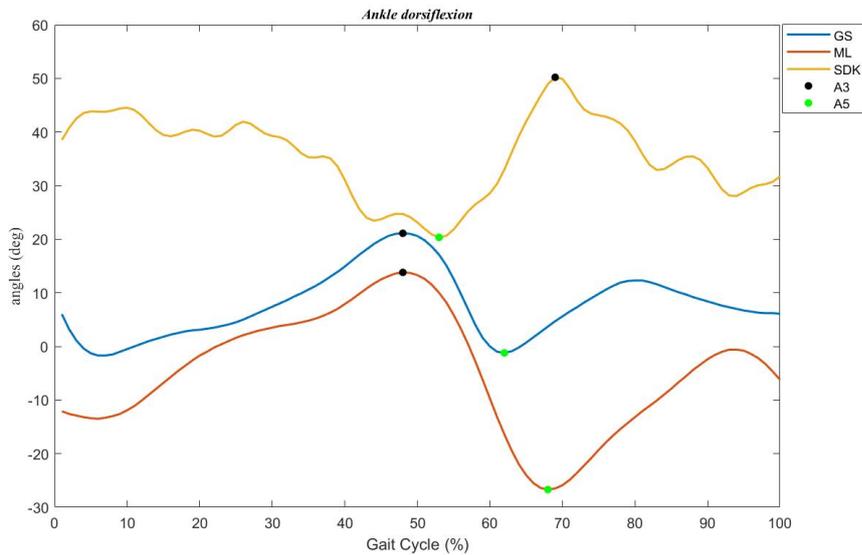


Figure 8.7: Illustration of ankle angle parameters estimation from the dorsiflexion angles extracted with the marker-based (MB), model-based (MLM) and body tracking SDK methods. The SDK ankle curve results in incorrect estimations of the ankle max flexion in the stance phase (A3) and swing phase (A5).

The SDK differences with respect to the MB method for the ankle in stance and swing phase were respectively 28.1° and 39.6° while the MLM method underestimates the MB values of 11.2° and 9.5° . These differences can be attributed to how the two methods define the foot segment for the ankle angle calculation. While the MLM method defines the foot segment starting from the LM position as the segment that best fits the foot segmentation, in the SDK method the foot was defined starting from the ankle coordinate to the toe coordinate.

The MLM performances were higher in the stance and swing phase of knee kinematics ($d < 0.8^\circ$) while SDK differences were lower during the first 40% of the gait cycle ($d < 1.4^\circ$). For what concerns the hip angles, both methods overestimate the hip flexion during the stance phase, the MLM showed good performances with a mean absolute difference with respect to the MB estimations of only 3.3° while the SDK mean absolute difference was 7.7° .

The overall performances of the SDK are also affected by the fact that in 2 out of 5 subjects the contralateral limb was misidentified as the foreground one, as shown in figure 8.3. This inevitably affected the calculation of the joint angles as during the gait cycle the joints of the two limbs have different inclinations, and while the foreground limb is in stance phase the background lower limb is in its swing phase.

Based on the ICC values, ML revealed moderate (0.5 - 0.75) to good (0.75 - 0.9) reliability [33] except for the knee parameters in stance and swing phase and the hip parameter. The SDK showed lower values of ICC with respect to the MLM except for the hip and knee variables in the stance phase.

CHAPTER 9

CONCLUSIONS

This thesis aimed to investigate whether motion tracking through the body tracking SDK integrated into the Azure Kinect DK could be employed to perform gait analysis for clinical purposes and to compare the performances of the above-mentioned SDK to an improved custom version of a 2D markerless method (MLM) based on a subject-specific kinematic model developed by Balta et al., 2020.

For this purpose, a gait analysis protocol for the evaluation of the lower limb joint kinematics was assessed. The investigations were conducted on five healthy subjects recorded during ten gait trails (5 for the left limb and 5 for the right limb) with the optical stereophotogrammetric system and the Azure Kinect. The data were not synchronously collected due to the infrared interference of the two systems which had the same IR working wavelength. For the optical SP data, the hip, knee and joint angles were computed with the marker-based MB method that included the estimations through the Vicon-Nexus software. For the Azure Kinect data, the lower limb joint angles were estimated with the MLM method based on three subject-specific kinematic models (in stance, load and flexion) and with the SDK method starting from the 2D joint coordinates of the hip, knee and ankle extracted with the body tracking software provided by Microsoft.

Seven significant gait variables were extracted from the hip, knee and ankle angles to evaluate the estimation performances. To compare the results obtained with the two ML methods the mean differences with respect to the MB system were computed.

The ML recordings were a lot faster than the MB ones since they did not require the marker placement procedure, moreover, they can be conducted also outside of gait laboratories and are considerably less expensive.

The MLM method was more time-consuming than the SDK one because it required the manual identification on the image of the joint centre's positions for the creation of the three subject-specific models in each gait trail. The processing time of the MLM method is larger than the body tracking SDK software, which provides the estimation of joint coordinates

also in real-time. Although, while the Body Tracking SDK is a "black box", the outcomes of the MLM method can be controlled to increase the accuracy of the estimations for clinical applications.

From a visual inspection of the resulting lower limb body tracking with the SDK method it can be concluded that the accuracy of the joint position estimation is visibly low in many frames when the motion tracking is performed in the sagittal plane. Moreover, the foreground limb was sometimes misidentified as the foreground one, affecting the performance of the method.

From the comparison with the MB method, which can be considered the gold standard for the evaluation of joint kinematics, both the SDK and the MLM method showed good performances in the estimation of the knee and hip angles. This preliminary investigation demonstrated that the SDK method provided higher differences than the MLM in the estimation of most of the gait variables under investigation. Both methods showed higher differences with respect to the MB method in the evaluation of the ankle parameters, especially the SDK.

The SDK method showed significant limitations in the evaluation of the ankle angle. The foot segment was defined starting from the ankle to the toe coordinate but the ankle's real inclination was not well represented through this method. In the MLM method, the foot segment was defined starting from the LM position as the segment that best fits the foot segmentation and resulted in visibly better performances in the evaluation of the ankle angles.

Considering the outcomes found so far it can be concluded that the SDK method implemented in this study cannot be employed to perform gait analysis for clinical purposes as it is because of the above-mentioned limitations in its performance.

From the comparison of the methods in this investigation, the MLM method demonstrated better performances in the evaluation of the lower limb joint kinematics.

In summary, the Azure Kinect demonstrated to be a useful, compact and low-cost technology for performing gait analysis out of laboratories. The huge potentialities of the SDK method are its fast implementation and its applicability in various conditions. The drawback of the SDK method is that you cannot have any access or control over the outputs of the joints' coordinates extracted with the Body Tracking SDK, which algorithm is a "black box".

The MLM method required longer processing times because of the definition of the models but because of its subject-specificity, it provides better estimations. For the implementation of the MLM, the subject is required to wear colored socks and walk on a colored carpet while the Body Tracking SDK algorithm was developed to perform in very different environments. The MLM method provided better results than the SDK method implemented in

this project in the evaluation of the lower limb joint kinematics, the highest differences with respect to the MB system were found for the ankle angles.

Compared to the SDK the outputs of the MLM method can be controlled and tuned to be more accurate. In this context, the accuracy of the joint centre's estimation with the MLM method could be further improved by implementing a 3D model based on the depth image information in future work. The outputs of the lower limb joint coordinates of both the SDK and the MLM method could be integrated to increase the applicability of the MLM method but also reduce its processing time. The specific estimation of the joint position from the SDK body tracking could be used based on their confidence level, which is a parameter provided by the body tracking software. Regarding the ankle and toe coordinates, the possible outcome could rely more on the MLM method which performs better in these cases.

As regards the dataset, one of the limitations of this study is the limited number of subjects. In future work, these investigations could be conducted for a larger dataset and in patients with altered gait functions, e.g., patients with cerebral palsy, to study the performances of the methods on unusual gait cycles and evaluate their clinical applicability.

Bibliography

- [1] W. Pirker and R. Katzenschlager. Gait disorders in adults and the elderly: A clinical guide. *Wiener klinische Wochenschrift*, 129, 10 2016. pages viii, 7
- [2] G. Wu, S. Siegler, P. Allard, C. Kirtley, A. Leardini, D. Rosenbaum, M. Whittle, D. D’Lima, L. Cristofolini, H. Witte, O. A. Schmid, and I. Stokes. ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—part i: ankle, hip, and spine. international society of biomechanics. *Journal of biomechanics*, 35:543–8, 05 2002. pages viii, 8, 9, 11
- [3] E. S. Grood and W. J. Suntay. A Joint Coordinate System for the Clinical Description of Three-Dimensional Motions: Application to the Knee. *Journal of Biomechanical Engineering*, 105(2):136–144, 05 1983. pages viii, 8, 10
- [4] M. P. Kadaba, H. K. Ramakrishnan, and M. E. Wootten. Measurement of lower extremity kinematics during level walking. *Journal of Orthopaedic Research*, 8, 1990. pages viii, 11
- [5] A. Cappozzo, U. Della Croce, A. Leardini, and L. Chiari. Human movement analysis using stereophotogrammetry: Part 1: theoretical background. *Gait & Posture*, 21(2):186–196, 2005. pages viii, 13
- [6] Neil J. Cronin. Using deep neural networks for kinematic analysis: Challenges and opportunities. *Journal of Biomechanics*, 123:110460, 2021. pages viii, 16
- [7] N. Nakano, T. Sakura, K. Ueda, A. Omura, L. and Kimura, Y. Iino, S. Fukashiro, and S. Yoshioka. Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras. *Frontiers in Sports and Active Living*, 2, 2020. pages viii, 16, 17
- [8] Microsoft. Azure Kinect DK hardware specifications. <https://docs.microsoft.com/en-us/azure/kinect-dk/hardware-specification>. Accessed May 2022. pages ix, 18, 20

- [9] E. Pantzar, A. Cereatti, G. Figari, N. Valeri, G. Paolini, U. Della Croce, A. Magnuson, and J. Riad. Knee joint sagittal plane movement in cerebral palsy: a comparative study of 2-dimensional markerless video and 3-dimensional gait analysis. *Acta Orthopaedica*, 89:656–661, 12 2018. pages 3, 17, 56
- [10] A. Castelli, G. Paolini, A. Cereatti, and U. Della Croce. A 2D Markerless Gait Analysis Methodology: Validation on Healthy Subjects. *Computational and Mathematical Methods in Medicine*, 2015, 04 2015. pages 3, 12, 14, 17
- [11] L. Wade, L. Needham, P. McGuigan, and J. Bilzon. Applications and limitations of current markerless motion capture methods for clinical gait biomechanics. *PeerJ*, 10, February 2022. pages 3, 12, 14, 16
- [12] S. Colyer, M. Evans, D. Cosker, and A. Salo. A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system. *Sports Medicine - Open*, 4:1–15, June 2018. pages 4, 14
- [13] Microsoft. About Azure Kinect DK. <https://docs.microsoft.com/en-us/azure/kinect-dk/about-azure-kinect-dk>. Accessed May 2022. pages 4, 18
- [14] D. Balta, M. Salvi, F. Molinari, G. Figari, G. Paolini, U. Della Croce, and A. Cereatti. A two-dimensional clinical gait analysis protocol based on markerless recordings from a single rgb-depth camera. In *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6, 2020. pages 4, 17
- [15] J. Perry and M.G. Benedetti. *Analisi del movimento*. Elsevier, 2005. pages 6
- [16] Kenton Kaufman and KaiNan An. Chapter 6 - biomechanics. In Gary S. Firestein, Ralph C. Budd, Sherine E. Gabriel, Iain B. McInnes, and James R. O’Dell, editors, *Kelley and Firestein’s Textbook of Rheumatology (Tenth Edition)*, pages 78–89. Elsevier, tenth edition edition, 2017. pages 8
- [17] A. Cereatti, F. Margheritini, M. Donati, and A. Cappozzo. Is the human acetabulofemoral joint spherical? *The Journal of bone and joint surgery. British volume*, 92:311–4, 02 2010. pages 10
- [18] L. Chiari, U. Della Croce, A. Leardini, and A. Cappozzo. Human movement analysis using stereophotogrammetry: Part 2: Instrumental errors. *Gait & Posture*, 21(2):197–211, 2005. pages 13

-
- [19] A. Leardini, L. Chiari, U. Della Croce, and A. Cappozzo. Human movement analysis using stereophotogrammetry: Part 3. soft tissue artifact assessment and compensation. *Gait & Posture*, 21(2):212–225, 2005. pages 13
- [20] U. Della Croce, A. Leardini, L. Chiari, and A. Cappozzo. Human movement analysis using stereophotogrammetry: Part 4: assessment of anatomical landmark misplacement and its effects on joint kinematics. *Gait & Posture*, 21(2):226–237, 2005. pages 13
- [21] Theia Markerless Inc. Theia 3D. <https://www.theiamarkerless.ca/>. Access May 2022. pages 14
- [22] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Real-time multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. pages 16
- [23] D. Balta, M. Salvi, F. Molinari, G. Figari, G. Paolini, U. Della Croce, and A. Cereatti. A markerless gait analysis protocol based on a single RGB-Depth camera: sensitivity to background changes. *GNB 2020 PROCEEDINGS*. pages 17, 28
- [24] Microsoft. Skeletal Tracking on Azure Kinect. <https://www.microsoft.com/en-us/research/project/skeletal-tracking-on-azure-kinect/>. Accessed May 2022. pages 17, 40, 41
- [25] C. S. Bamji, S. Mehta, B. Thompson, T. Elkhatib, S. Wurster, O. Akkaya, A. Payne, J. Godbaz, M. Fenton, V. Rajasekaran, L. Prather, S. Nagaraja, V. Mogallapu, D. Snow, R. McCauley, M. Mukadam, I. Agi, S. McCarthy, T. Xu, Z. and Perry, W. Qian, V. Chan, P. Adepu, G. Ali, M. Ahmed, A. Mukherjee, S. Nayak, D. Gampell, S. Acharya, L. Kordus, and P. O’Connor. IMpixel 65nm BSI 320MHz demodulated TOF Image sensor with 3m global shutter pixels and analog binning. In *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, pages 94–96, 2018. pages 19
- [26] Microsoft. Azure Kinect DK depth camera. <https://docs.microsoft.com/th-th/azure/kinect-dk/depth-camera>. Accessed May 2022. pages 19
- [27] Microsoft. Azure Kinect Sensor SDK. <https://docs.microsoft.com/en-us/azure/kinect-dk/sensor-sdk-download>. Accessed February 2022. pages 41
- [28] Microsoft. Azure Kinect Body Tracking SDK. <https://docs.microsoft.com/en-us/azure/kinect-dk/body-sdk-download>. Accessed February 2022. pages 41, 54

- [29] Vicon Motion Systems Ltd UK. Motion Capture Cameras. <https://www.vicon.com/hardware/cameras/>. Accessed May 2022. pages 44
- [30] Vicon Motion Systems Ltd UK. Modeling with Plug-in Gait. <https://docs.vicon.com/display/Nexus212/Lower+body+modeling+with+Plug-in+Gait>. Accessed May 2022. pages 49
- [31] Microsoft. Azure Kinect Body Tracking samples. <https://github.com/microsoft/Azure-Kinect-Samples/tree/master/body-tracking-samples>. Accessed February 2022. pages 54
- [32] M. G. Benedetti, F. Catani, A. Leardini, E. Pignotti, and S. Giannini. Data management in gait analysis for clinical applications. *Clinical biomechanics*, 13 3:204–215, 1998. pages 56
- [33] Terry Koo and Mae Li. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15, 03 2016. pages 65