

POLITECNICO DI TORINO

Master Degree Course in Biomedical Engineering



Master Degree Thesis

Two dimensional markerless gait analysis protocol for estimating the sagittal lower limb joint kinematics with a single RGB-D camera for clinical applications.

Supervisors:

Prof. Ing. Filippo Molinari

Prof. Ing. Andrea Cereatti

Candidate:

Dott. Diletta Balta

December 2019

A noi tre... una vera famiglia, praticamente una squadra

Symbols

GA	Gait Analysis
ML	Markerless
ICP	Iterative Closest Point
JI	Jaccard Index
RMSD	Root Mean Square Deviation
CP	Cerebral Palsy
ROM	Range of Motion
FoV	Field of View
ToF	Time of Flight
LM	Lateral Malleolus
LE	Lateral Epicondyle
GT	Great Trochanter
JCS	Joint Coordinate System
CCS	Cartesian Coordinate System
DoF	Degrees of Freedom
PMW	Progressive Weighted Mean

Contents

1	Clinical relevance and aim of the work	10
2	Introduction to gait analysis	13
2.1	<i>The gait cycle</i>	13
2.2	<i>Joint Kinematics</i>	15
2.2.1	Hip joint kinematics.....	15
2.2.2	Knee joint kinematics.....	16
2.2.3	Ankle joint kinematics.....	17
3	Optical tracking methods	19
3.1	<i>Optoelectronic stereophotogrammetry</i>	19
3.2	<i>2D-dimensional markerless gait analysis</i>	20
3.2.1	2D-dimensional markerless gait analysis – A literature review.....	21
4	2D markerless gait analysis protocol for clinical applications through the use of a single RGB-D camera	29
4.1	<i>Cycle segmentation</i>	31
4.2	<i>Human segmentation</i>	31
4.3	<i>Two dimensional orientations of the lower limb segments</i>	46
4.3.1	Participant-specific multi-segmental model.....	47
4.3.2	Extraction of the three body segments.....	49
4.3.2.1	Foot identification.....	49
4.3.2.2	Shank identification.....	51
4.3.2.3	Thigh identification.....	57
4.3.3	Joint center tracking trajectories during the gait cycle though the ICP algorithm.....	65
4.3.3.1	Iterative closest point (ICP) algorithm.....	67
4.3.4	Computation of lower limb joint angles.....	68
5	Experimental acquisitions	69
5.1	<i>Preliminary investigations: hardware selection</i>	69
5.1.1	Hardware description.....	69
5.1.1.1	Kinect II.....	69
5.1.1.2	Intel real sense depth camera D435.....	74
5.1.2	Hardware selection.....	78

5.2	<i>Experimental set up</i>	80
5.2.1	Two different setup conditions: with and without green background.....	80
5.2.2	Two different light conditions: two LED lamps vs one LED lamp	81
5.3	<i>Subject preparation</i>	82
5.4	<i>Calibration of the SP and ML systems</i>	83
5.5	<i>Measurement protocol</i>	84
6	Results and discussions	85
6.1	<i>Comparison between manual and automatic masks in the different setup conditions</i>	85
6.2	<i>Comparison between marker-based and 2D markerless kinematics</i>	89
6.2.1	Factors affecting the accuracy of the joint kinematics estimation	94
6.2.1.1	Misalignment between RGB and depth images	94
6.2.1.2	Accuracy of the subject segmentation: how errors in segmenting the subject affect joint angles estimation	97
7	General conclusions	99
7.1	<i>Limitations and future work</i>	101
	Appendix	102
	Bibliography	103

Abstract

The measure of lower limb joint kinematics is helpful in evaluating gait alterations. Nowadays, optoelectronic stereophotogrammetry gait analysis represents the gold-standard in clinical practice to quantitatively assess the human motion because it measures the position of the markers in the three-dimensional space with submillimeter accuracy. However, it is not without limitations.

Video-based markerless systems represent a low-cost powerful and promising alternative to 3D marker-based systems. In fact, the use of markerless techniques allows to make the experimental sessions faster and easier since it does not require the application of some markers on the skin of the patients. In addition, recent advancements in RGB-D camera (RGB camera with a depth infrared sensor) technology and video processing have paved the way to the development of a new generation of low-cost movement analysis systems.

The aim of this thesis project is to develop a two-dimensional markerless gait analysis protocol through the use of a single RGB-D camera to estimate the lower limb joint kinematics for clinical applications. The proposed protocol for the estimation of the joint angles consists of four parts:

- image segmentation to separate the participant from the background (Salvi and Molinari, 2018);
- definition of a multi-segmental model of the lower limb;
- joint center tracking in which the trajectories of the joint center of hip, knee and ankle are estimated by using the ICP algorithm (P. J. Besl and N. D. McKay, 1992);
- estimation of the joint kinematics: the orientations of the lower limb segments are determined on sagittal plane from joint center trajectories.

This thesis project is organized in four sections. The first one aims at implementing an automatic segmentation algorithm to separate the subject from the background. The image segmentation is a key component in several markerless gait analysis systems, but the state of the art introduces the need of a controlled environment including a coloured background and *ad-hoc* coloured clothes as socks or bodysuit. To simplify this step, the presence of a coloured background is required; if on one hand it standardises the experimental scenario, on the other hand it complicates the set-up conditions by increasing the number of required tools and the installation time. The subject segmentation is the most critical step because the automated choice of an optimal threshold to separate the subject from the background is not trivial if the latter is inhomogeneous or if it has similar colours to some body regions. A modified version of Salvi and Molinari, 2018 is implemented to segment the subject by introducing anatomical and biomechanical hypothesis. The human segmentation algorithm used in this thesis is a fully automatic thresholding method which is evaluated on the image histogram resulting from the subtraction between a frame containing the subject and a frame with only the background captured at the beginning of the trial. The choice of an automatic setting of the threshold value only on the basis of the image histogram makes this algorithm robust to variable backgrounds. The algorithm performances are validated by comparing the obtained segmentations with manual masks by calculating the Jaccard index.

The second part of this work aims at applying a modified version of the algorithm proposed in Pantzar-Castilla *et al.* to extract the foreground lower limb and to estimate the relative joint kinematics.

In the third part the experimental session conducted to validate the two methods is described. Prior to performing the experiments, a preliminary investigation about the quality of the images provided by two different RGB-D cameras is done to select the one which provides the higher quality images. The experimental session is executed with and without the homogeneous coloured background and with two

different light conditions to quantify how these factors affect the accuracy of both the segmentations and the joint angle estimates.

Finally, the resulting joint angles are compared with those obtained from a stereophotogrammetric system (only the sagittal plane is considered) for the different setup conditions by calculating the mean waveforms obtained from three trials. In particular, the offset between the mean waveforms and the RMS value of the waveforms difference without their offset are computed for each condition under analysis.

The first evidence, by analysing the segmentation results, is the worsening of the Jaccard index values when the segmentation is carried out without the background in all the considered conditions. In particular, the performances decrease of 6.6 % and 2.2% in one and two LED lamps light conditions, respectively.

The percentage decrease of the JI values, obtained without the green background, amounts to 5% passing from two additional LED lamps to only one. By analysing the offset of the mean joint kinematics waveforms, the hip and the ankle exhibit higher values than one of the knee. This could be explained by considering that the anatomical axes of hip and ankle defined within the stereophotogrammetric protocols are different from those defined for 2D video-based gait analysis. Instead in the case of knee joint, the reference axes are defined starting from the same anatomical landmarks in both protocols and this explains the lower offset values.

In general, by considering the RMSD values, the average errors for all the three joints considered are lower when the acquisitions are conducted with two LED lamps both with and without the green background. When the green background is present, the RMSD values with two LED lamps are slightly better with respect to one LED lamp only, while without the green background, the results obtained with two LED lamps are more accurate (about 2° on average). The RMSD values in ankle kinematics (3.8° on average) are lower than the other joints (5.6° for the hip joint and 6.4° for the knee joint, on average). To sum up, the main findings of this study highlight the possibility of conducting a markerless gait analysis acquisition without the need of a homogeneous background since with two LED lamps the decrease of the performances is neglectable (only 0.2° on average) with respect to the acquisitions with green background. This represents an excellent result since the experimental set-up is simplified in favour of both system transportability and installation.

The main source of errors which have arisen during this project can be identified in the spatial misalignment between the RGB and depth images which are directly reflected in the joint angle computation and in the presence of a highly inhomogeneous background regions (e.g. reflective glass window) which can be reduced when two LED lamps are used.

To conclude, this thesis is a part of a European project with a Swedish hospital center. In particular, in Sweden there is a preventive follow-up program for children with cerebral palsy to increase the chances of identifying children at risk of crouch gait. The 2D markerless gait analysis protocol proposed in Castilla et al was validated on 18 children with cerebral palsy by using a single RGB-D camera. We have collaborated with this Swedish hospital center to study the advantages and limitations of the adopted 2D gait analysis markerless protocol in order to improve the algorithm and especially the experimental scenario.

The first part of this thesis project was also aimed at conducting a comparative study of the performances of the most common segmentation algorithms in terms of Jaccard index. This work has been accepted on 'Gait and posture' as a conference proceedings of the Italian Society of the Clinical Movement Analysis (SIAMOC) held in Bologna in October 2019 and it is reported in Appendix.

1 Clinical relevance and aim of the work

Walking is an important part of daily activities; however, it is difficult to study all aspects of the mechanics of walking. The gait analysis is defined as the quantitative and qualitative study of the mechanical aspect of the human walking using instrumentation and observation.

The main features considered in the study of human movement are spatio-temporal parameters, ground reaction forces, joint kinematics and the energy expense [1].

The gait analysis (GA) is a valuable tool to characterize and monitor the locomotion alterations in many fields such as research, rehabilitation, sports biomechanics. One aspect of particular relevance is that the GA is very suitable in children with cerebral palsy (CP) to identify and quantify gait alterations [2]. In particular, the crouch gait is a difficult-to-treat deviation (Figure 1.1) and it is represented by excessive flexion of the ankle, knee, and hip during stance phase. The specific causes of this disease are different among the patients, but the primary biomechanical contributors are short hip flexors, short and/or spastic hamstrings, weak hip and knee extensors, weak ankle plantar flexors, and/or malrotation of the femur, tibia, and foot [3].



Figure 1.1: This picture illustrates the crouch gait, that affects children with CP, represented by excessive flexion of the ankle, knee, and hip during stance phase. Taken from: <https://opensim.stanford.edu/work/rehabilitation.html> (last accessed: 26/10/2019, 19:22).

After a literature review, it can be noted that crouch gait worsens with age due to increased body size and weight. S. Ounpuu *et al* (2002) [4] analysed the progression of gait in children with CP, by considering twenty-eight children. In particular, these patients had two gait analyses with no surgical intervention between the tests. The effect of age and growth were evaluated by using temporal and stride parameters, three-dimensional kinematics and clinical examination measures. The results showed the decrease of lower limb joints sagittal plane range of motion (ROM), the decline in timing of toe off, cadence and walking velocity.

Generally, in these patients with cerebral palsy, the gait functions decrease compared with a control group who had an orthopaedic intervention. E. E. Butler *et al* (2016) [3] analysed, annually, a young subject with CP by clinical motion analyses over an eight-year time period (Figure 1.2), from six to thirteen years of age.

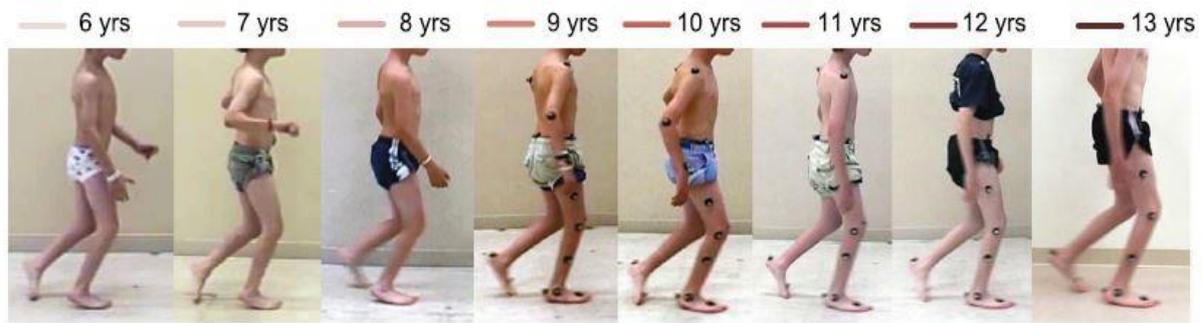


Figure 1.2: the progression of crouch gait over an eight-year time period. Taken from E.E. Butler et al (2016).

In conclusion, the crouch gait can ultimately limit the ability to ambulate and its monitoring in time is fundamental to decide whether the surgery is necessary.

A functional and low-cost evaluation becomes a key factor in the scenario as the one presented when a 3D analysis is not strictly necessary.

Nowadays, optoelectronic stereophotogrammetry gait analysis is a gold-standard in clinical practice for quantitative assessment of human motion, it provides accurate results, but it is not without limitations. The use of this technique is limited in specialized laboratories, the instrumentation is expensive and allows only the evaluation of movement in a confined space due to cameras. Moreover, it is time consuming because of the meticulous marker placement on the patient as well as the accurate calibration needed every time the cameras are moved.

Video-based markerless (ML) systems can represent a promising alternative to 3D marker-based systems. In fact, the use of ML techniques allows to make the experimental sessions faster and simpler since it does not require the application of fixtures on the skin of the patients [5].

Markerless methods are not widely available for biomechanical and clinical applications because nowadays the accuracy of the joint kinematics estimates is not sufficient to assess pathologies or gait alterations.

After a literature review, it can be noted that there are a lot of works that analyse the development of the use of a single camera because it is sufficient for the description of gait in 2D. In this way the space needed for performing the experimental sessions is reduced as well as the costs associated to the camera. In addition, recent advancements in RGB-D camera (RGB camera with a depth infrared sensor) and video processing have paved the way to the development of a new generation of low-cost movement analysis systems. RGB-D camera systems integrate depth assessment with traditional two-dimensional images, so in this way it is possible to have information about the distance of an object from the depth sensor. The depth information could be useful to identify different regions of the human body in order to choose the region of interest that in this specific application is the foreground lower limb.

The aim of the thesis project is the development of a 2-dimensional markerless (2D ML) gait analysis protocol using a single RGB-D camera to estimate the lower limb joint kinematics for clinical applications.

The proposed protocol for the estimation of the joint angles consists of four parts:

- image segmentation to separate the participant from the background;
- multi-segmental model to introduce a lower limb model;
- joint center tracking to estimate the trajectories of the hip, knee and ankle joint center estimated by using the ICP algorithm (P. J. Besl and N. D. McKay, 1992);
- estimation of the joint kinematics where the orientations of the lower limb segments are determined on sagittal plane from joint center trajectories.

This thesis project is organized in four sections.

The first one aims to implement an automatic segmentation algorithm to separate the subject from the background. The image segmentation is a key component in several current markerless gait analysis systems, but the state of the art introduces the need of a controlled environment including a coloured background and *ad-hoc* coloured clothes as socks or bodysuit. To simplify this step, the presence of a coloured background is required; on the one hand it standardises the experimental scenario, but on the other hand it complicates the set-up conditions by increasing the number of required tools and the preparation time. The subject segmentation is the most critical step because the automated choice of an optimal threshold to separate the subject from the background is not trivial if the latter is inhomogeneous or if it has similar colours to the body regions. A modified version of (Salvi and Molinari, 2018) is implemented to segment the subject by introducing anatomical and biomechanical hypothesis. The human segmentation algorithm used in this thesis is a fully automatic thresholding method which is evaluated on the image histogram resulting from the subtraction between a frame containing the subject and a frame with only the background captured at the beginning of the trial. The choice of an automatic setting of the threshold value only on the basis of the image histogram makes this algorithm robust to variable backgrounds. The algorithm relative performances are validated by comparing the obtained segmentations with manual masks by calculating the Jaccard index.

The second part of this work aims at applying a modified version of the algorithm proposed in Pantzar-Castilla *et al.* to extract the foreground lower limb and to estimate the relative joint kinematics.

In the third part the experimental session conducted to validate the two methods is described. Prior to performing the experiments, a preliminary investigation about the quality of the images provided by two different RGB-D cameras is done to select the one which provides the higher quality images. The experimental session is executed with and without the homogeneous coloured background and with two different light conditions to quantify how these factors affect the accuracy of both the segmentations and the joint angle estimates. The computed joint angles are finally compared with those obtained from a stereophotogrammetry system (only the sagittal plane is considered) for the different setup conditions. In the last part, there is an appendix in which my collaboration with a Swedish hospital center is described because this thesis is a part of a European project with this center. In particular, in Sweden there is a preventive follow-up program for children with cerebral palsy to increase the chances of identifying children at risk of crouch gait.

The 2D markerless gait analysis protocol proposed in Pantzar-Castilla *et al.* was validated on 18 children with cerebral palsy by using a single RGB-D camera. I have been collaborated with this Swedish hospital center to study the advantages and limitations of the adopted 2D gait analysis markerless protocol in order to improve the algorithm and especially the experimental scenario.

2 Introduction to gait analysis

2.1 The gait cycle

The gait cycle represents the functional unit in gait analysis. Considering the lower limbs each normal gait cycle begins and ends with two consecutive impacts of the leading limb's heel with the ground, the so-called *heel strike* (or *initial contact*) event. Each gait cycle can be also subdivided into two main periods: the *stance* and the *swing* [6]. During *stance* (60% of the gait cycle, on average) the foot is in contact with the floor and it supports the whole or a part of the body weight, while during *swing* (40% of the gait cycle, on average) the foot is no longer in contact with the ground because the leg is swinging in the air for limb advancement [6].

Figure 2.1 illustrates the different phases [7] that make up the *stance* and the *swing* period of a normal level ground walking, furthermore, a description of these phases is provided below.

Stance

- *Load Response*: this phase starts with *heel strike* event that marks also the beginning of normal gait cycle, as stated before. The *load response* phase is characterised by flexed hip, nearly fully extended knee (about 5 degrees) and the ankle in the position of dorsiflexion at the moment of *heel strike*, then from 0% to 10% of gait cycle the knee is progressively flexed to attenuate the shock until the whole foot is in contact with the ground (*foot flat* event). During the *load response* the leading limb is rapidly loaded so it takes over the body weight [6]. In this phase both feet are in contact with the ground (*double support*). *Load response* ends when the opposite limb begins the swing period.
- *Mid Stance*: this phase takes place between 10% and 30% of gait cycle and it is characterised by an extension of both hip and knee while the ankle is dorsiflexed, the opposite limb is in the swing phase and it passes the stance limb, so the body is moved forward. The *mid stance* is the first phase of gait cycle where the whole-body weight is supported by the stance leg only (*single support*). *Mid stance* ends with *heel off* (or *push off*) event where the heel loses contact with the floor.
- *Terminal Stance*: it takes places between 30% and 50% of gait cycle, during this phase the knee increases its extension and then begins a slight flexion. *Terminal stance* is the phase in which the body is propelled forward while the limb advances on the forefoot used as a fulcrum. It ends with the *heel strike* of the opposite limb.
- *Pre-Swing*: it takes place between 50% and 60% of gait cycle, this phase is characterised by an increasing flexion of both ankle and knee, while the hip decreases its extension. The push caused by the propulsion results in the loss of contact between toes and ground (*toe-off* event) that marks the end of this phase and the *stance* period.

Swing

- *Initial-Swing*: it arises between 60% and 73% of gait cycle, in this phase the foot is no longer in contact with the floor and the limb is accelerated forward thanks to the hip and knee flexion. The ankle results to be only partially dorsiflexed.

- *Mid-Swing*: it takes place between 73% and 87% of gait cycle. In this phase the swinging limb crosses the line of body gravity by means of a further hip flexion while the knee is extended, and the ankle continues its dorsiflexion.
- *Terminal-Swing*: this phase, which takes place between 87% and 100% of gait cycle, is the last phase of swing period. During *terminal swing* the swinging limb is decelerated and stabilized by muscles in preparation for the forthcoming *heel strike* which marks the end of the gait cycle. In this phase the hip is flexed, the knee is extended, and the ankle is in the position of dorsiflexion.

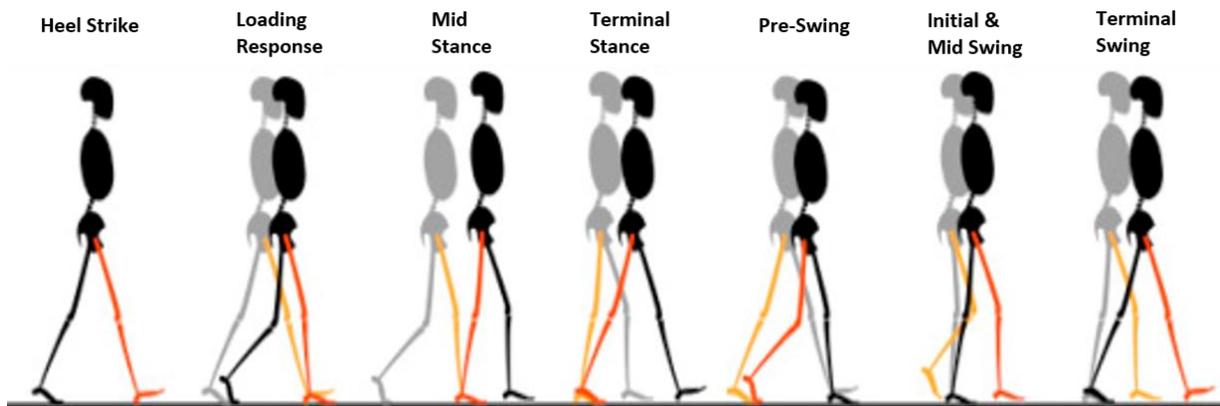


Figure 2.1: In this picture the gait phases and events of a normal level ground walking are shown on a normalized time scale. Adapted from Mohammed et al (2016).

Focusing on one lower limb a four-level coding classification of the gait cycle based on gait events can be made as suggested by Agostini et al. (2014) [6]. In fact, during normal walking the majority of gait cycles presents the following sequence of events: heel contact (H), flat foot contact (F), push-off (P) and swing (S). All cycles that follow this pattern (HFPS) are classified to be typical.

These gait events can be detected by foot-switches sensors attached to the sole of the foot that allow for a direct measure of the contact with the floor. Figure 2.2 shows the foot-switches state (the black disc indicates that the switch is closed, so the region of the sole in which is mounted is in contact with the floor) and the corresponding phase of the gait cycle.

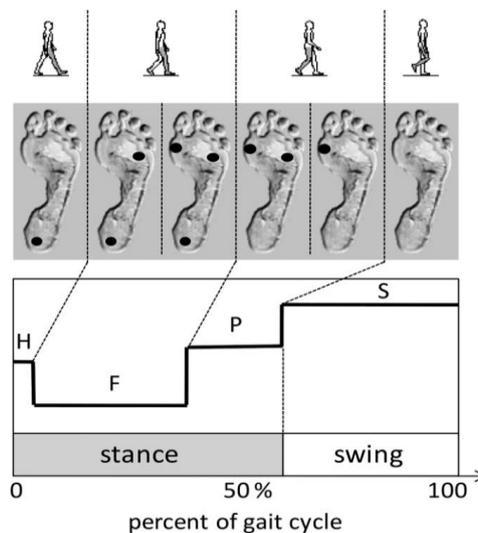


Figure 2.2: 'Foot switches'. Gait phases: Heel contact (H), flat foot contact (F), push-off or heel off (P), swing (S). Taken from Agostini et al (2014).

2.2 Joint Kinematics

2.2.1 Hip joint kinematics

The hip joint connects the acetabulum of the pelvis and the head of femur. The acetabulum is a cup-like depression located on the inferolateral aspect of the pelvis. The head of femur is hemispherical and fits completely into the concavity of the acetabulum.

The model adopted to model the hip joint in this context is the spherical hinge. The centre of the hinge coincides with the centre of rotation of the femur relative to the pelvis which is made to coincide with the geometrical centre of the acetabulum and, in turn, with the centre of the femoral head. The spherical hinge is characterized by three rotational DOFs: Fl/Ex, Ab/Ad and In/Ex rotation.

Several proposals can be found in the literature but the International Society of Biomechanics (ISB) recommends the use of the "Joint Coordinate System" (JCS) proposed by Grood and Suntay in 1983 [8]. This JCS allows to report the motions of the hip joint in clinically relevant terms, moreover it is widely accepted in biomechanics community [8]. Figure 2.3 illustrates the JCS.

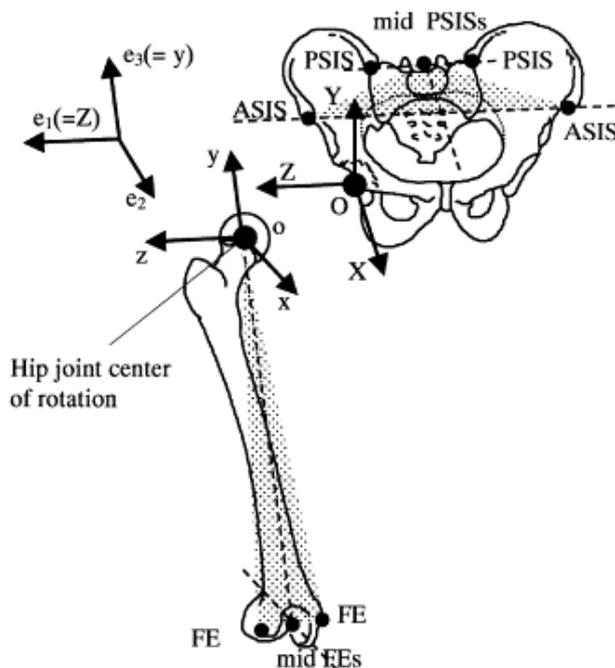


Figure 2.3: Pelvic coordinate system (XYZ), femoral coordinate system (xyz), and the JCS for the right hip joint. Taken from D. Bookwater et al (2002).

The hip JCS is composed by three rotation axes, e_1 , e_2 and e_3 . Two of the axes, e_1 and e_3 are named body fixed axes because they are embedded in pelvis and femur respectively, while e_2 is called floating axis and it is defined as the cross product of e_3 and e_1 , so it is the mutual perpendicular to them.

Grood and Suntay suggested to use capitalized letters X, Y, Z to denote the pelvis CCS with unit base vectors I, J, K and lower-case letters x , y , z to indicate the femoral CCS with unit base vectors i , j , k . The e_1 and e_3 body fixed axes of the JCS coincides with the Z-axis of the pelvic coordinate system and y-axis of the right (or left) femur coordinate system, respectively.

From a clinical point of view, the body fixed axis e_1 corresponds to the motion of flexion-extension.

The internal-external rotation of the femur, instead, occurs on the body fixed axis e_3 .

Finally, the adduction-abduction motion takes place on the floating axis e_2 .

Although the three rotations take place simultaneously during on normal level walking, as shown by Figure 2.4, the greater motion is flexion-extension that occurs on the sagittal plane only.

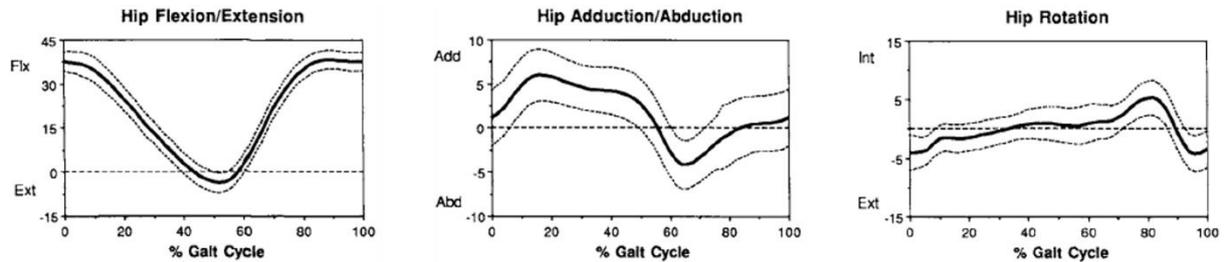


Figure 2.4: Mean (solid line) and one standard deviation (dotted line) of the hip angles during walking of normal adults. Adapted from: Kadaba (1990).

2.2.2 Knee joint kinematics

The knee connects the femur and the tibia bones, and it is one of the most complex and largest joint in human body. For the sake of completeness, the knee joint is actually composed by tibiofemoral and patellofemoral joints.

The knee is referred to be a modified-hinge joint with six degrees of freedom (DoF) because it exhibits three rotations and three translations [9].

A definition of one knee joint coordinate system is needed in order to describe the three rotation movements that occur during movement. Also in this case, the International Society of Biomechanics (ISB) recommends the use of the "Joint Coordinate System" (JCS) proposed by Grood and Suntay in 1983 [10]. This JCS allows to report the motions of the knee joint in clinically relevant terms, moreover it is widely accepted in biomechanics community [10]. Figure 2.5 illustrates the JCS.

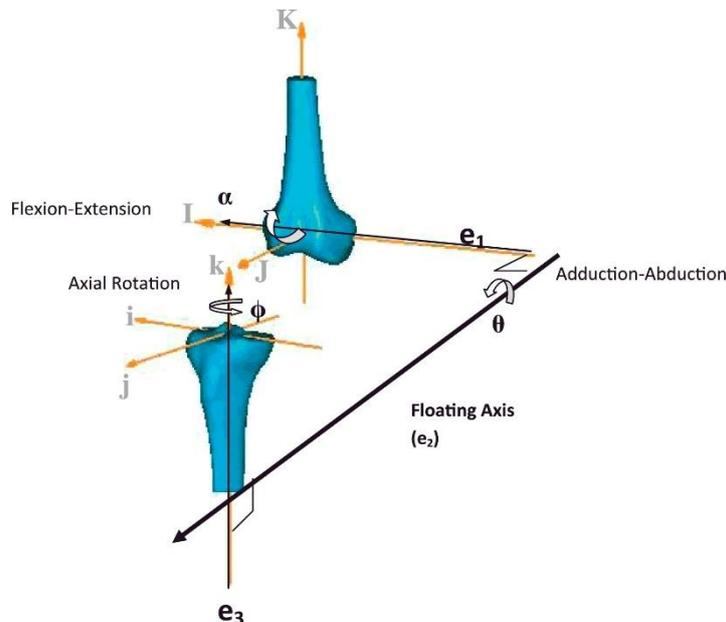


Figure 2.5: The Grood and Suntay Joint Coordinate System for the right knee. I, J, K is the base for the femur CCS, i, j, k for the tibia while e_1 , e_2 , e_3 are the axes of JCS. Taken from: Dabirrahmani and Hogg (2017).

The knee JCS is composed by three rotation axes, e_1 , e_2 and e_3 .

Two of the axes, e_1 and e_3 are named body fixed axes because they are embedded in femur and tibia respectively, while e_2 is called floating axis and it is defined as the cross product of e_3 and e_1 , so it is the mutual perpendicular to them.

For each bone a fixed Cartesian Coordinate System (CCS) is defined. Grood and Suntay suggested to use capitalized letters X, Y, Z to denote the femoral CCS with unit base vectors I, J, K and lower-case letters x , y , z to indicate the tibial CCS with unit base vectors i , j , k . The e_1 and e_3 body fixed axes of the JCS coincides with I and k of the femoral CCS and tibial CCS respectively.

From a clinical point of view, the body fixed axis e_1 corresponds to the motion of flexion-extension and it is also perpendicular to the femoral sagittal plane. The internal-external rotation of the tibia, instead, occurs on the body fixed axis e_3 . Finally, the adduction-abduction motion takes place on the floating axis e_2 .

Although the three rotations take place simultaneously during on normal level walking, as shown by Figure 2.6, the greater motion is flexion-extension that occurs on the sagittal plane only. As assessed in [11] many researchers measured the range of motion in this plane during level walking. In Figure 2.6 it is possible to observe that a fully or near fully extension can be found at *heel strike* (about 5°) and during *terminal stance*. The maximum flexion, instead, can be observed in *initial-swing* and it amounts to approximately 60° , while during *load response* phase a local maximum of flexion angle near 20° can be noticed.

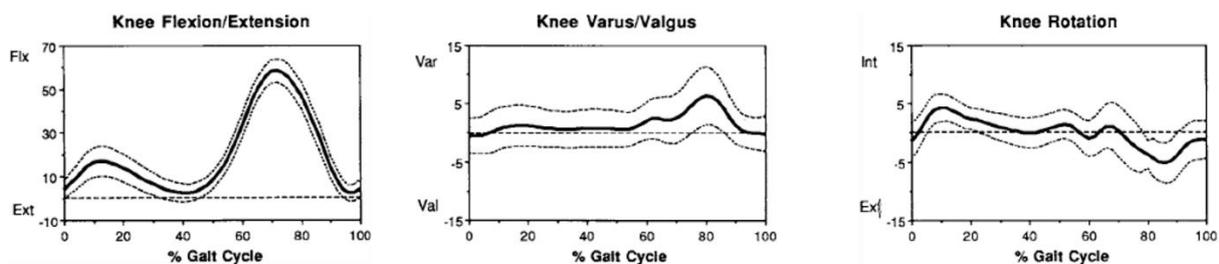


Figure 2.6: Mean (solid line) and one standard deviation (dotted line) of the knee angles during walking of normal adults. Adapted from: Kadaba (1990).

2.2.3 Ankle joint kinematics

The ankle joint is composed by the talocrural and the subtalar joints. The Talocrural joint (TC) connects shank and talus while Subtalar joint (ST) connects talus and foot.

The TC joint can be described as a hinge joint with the axis running just distal to the tips of the malleoli. Viewing from medial to lateral, the axis is directed posteriorly and inclined downward. The ST joint exhibits hinge-like rotation around a single oblique axis that approximately passes into the calcaneus posteriorly and the talus anteriorly.

Also in this case, the International Society of Biomechanics (ISB) recommends the use of the "Joint Coordinate System" (JCS) proposed by Grood and Suntay in 1983 [8]. This JCS allows to report the motions of the ankle joint in clinically relevant terms, moreover it is widely accepted in biomechanics community [8]. Figure 2.7 illustrates the JCS.

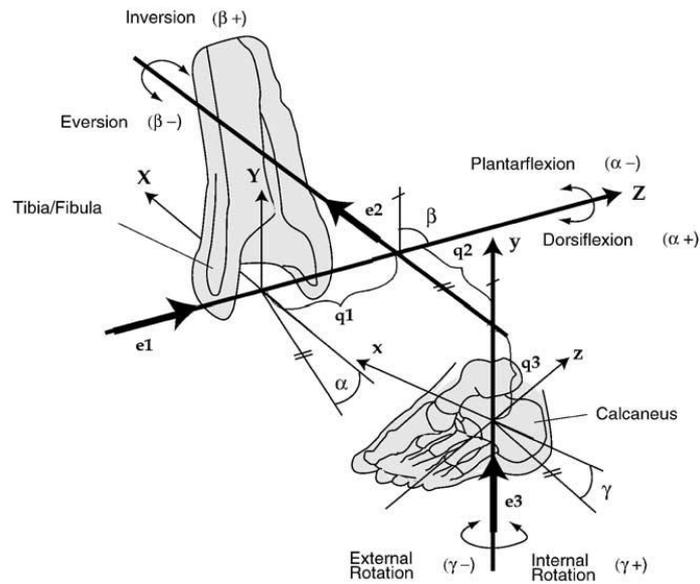


Figure 2.7: The JCS for the right ankle joint complex. Taken from D. Bookwater et al (2002)

The ankle JCS is composed by three rotation axes, e_1 , e_2 and e_3 . Two of the axes, e_1 and e_3 are named body fixed axes because they are embedded in tibia/fibula and calcaneus respectively, while e_2 is called floating axis and it is defined as the cross product of e_3 and e_1 , so it is the mutual perpendicular to them. Grood and Suntay suggested to use capitalized letters X, Y, Z to denote the tibia/fibula CCS with unit base vectors I, J, K and lower-case letters x , y , z to indicate the calcaneus CCS with unit base vectors i , j , k .

The e_1 and e_3 body fixed axes of the JCS coincides with the Z-axis of the tibia/fibula coordinate system and y-axis of the right (or left) calcaneal coordinate system, respectively.

From a clinical point of view, the body fixed axis e_1 corresponds to the motion of dorsiflexion (positive) or plantarflexion (negative).

The internal-external rotation, instead, occurs on the body fixed axis e_3 .

Finally, the inversion-eversion motion takes place on the floating axis e_2 .

The greater motion is dorsi-plantar flexion that occurs on the sagittal plane only as shown in Figure 2.8.

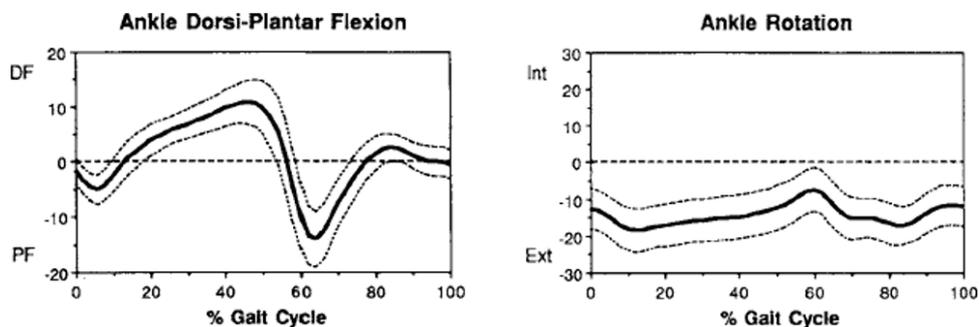


Figure 2.8: Mean (solid line) and one standard deviation (dotted line) of the ankle angles during walking of normal adults. Adapted from: Kadaba (1990).

3 Optical tracking methods

3.1 Optoelectronic stereophotogrammetry

Optoelectronic stereophotogrammetry goes back to Murey and Muybridge (19th century) considered to be the pioneers of this technique, in particular Jules Etienne Murey (1830 - 1904) was the first researcher that apply the photography in the study of human gait [12].

The principle on which stereophotogrammetry is based consists in the following steps. The first one comprises the acquisition of 2D images, after proper calibration of cameras, to identify (labelling) and extract the bi dimensional position of markers attached to the body segment in the image planes. To achieve this task some image processing techniques can be used such as linear thresholding, morphological operators and so on. The cameras involved in this step amounts from a minimum of two (to reconstruct the 3D) up to about 50 typically. The second step consists in the transition from 2D to 3D in order to obtain the 3D correspondences of each marker in a common coordinate system. Once the 3D coordinates are estimated it is possible to define a local reference frame for each body segment, then the joint pose can be obtained.

The last step consists in fusing the previous information with a human model to calculate the desired quantities such as trajectories, angular velocity, acceleration and joint kinematics of the skeleton segments in which the markers are positioned.

The optical system used can be either passive (as Vicon) or active (as Optotrak). Passive markers have in general spherical balls coated with a retroreflective material to reflect IR light coming from sources mounted on the cameras. The use of IR light allows to reduce the crosstalk with visible light by means of IR pass filter placed over the camera lens. Active markers are made up of IR LEDs placed on body segments. Only one LED at time is activated so the correspondence of a marker on the image is much more reliable. With this system the light has to cover the distance between camera and marker only once instead of twice as in the passive case, so it allows more distances and volumes available for acquisitions.

Furthermore, system based on different coloured LED markers are possible making it easier the labelling operation because each colour corresponds to a precise point of the body segment.

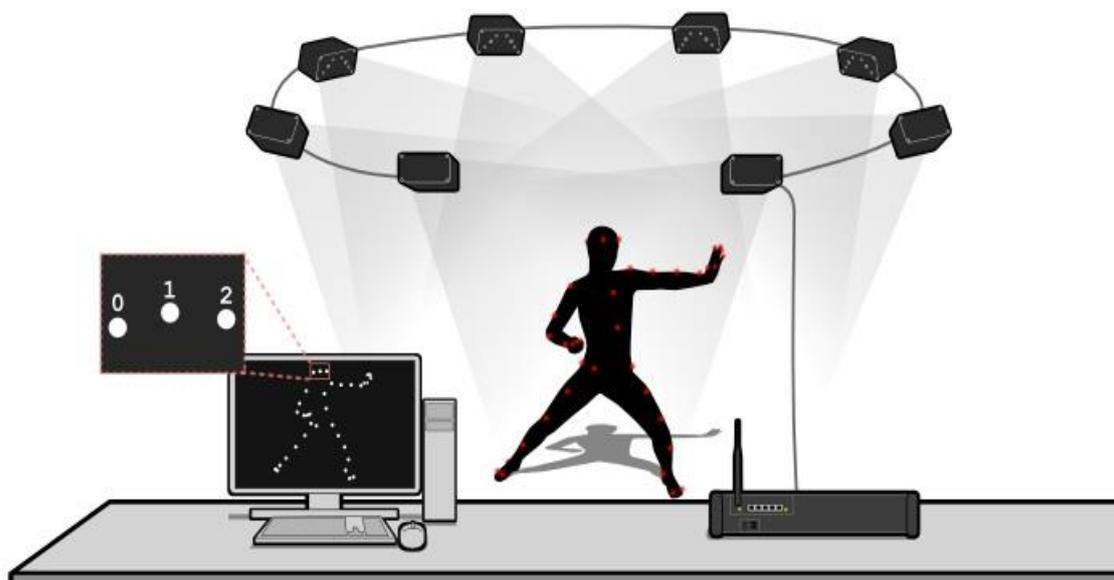


Figure 3.1: Example of motion capture system. Taken from http://www.phasespace.com/images/products/impulse_splash.jpg

The disadvantages of the active marker are that they require a power supply embedded on LEDs, moreover if each LED is activated at the same frame rate as the camera then the effective frame rate is divided by the number of markers. Both passive and active system suffers from the problem of trajectory gap and swap caused by marker occlusion that can occur when a light path is blocked during movement. The drawbacks of the stereophotogrammetry is that its availability is restricted in specialised laboratories with a confined capture volume, so only few number of gait cycle can be analysed and there is not the possibility to monitor the patients during their daily activities. Other limitations include the systems cost of the system and the time required for the calibration of the cameras and for the preparation of the patients. The measures can be performed only in the presence of an expert technician. Nevertheless, optical stereophotogrammetry provides very accurate results (errors in position are estimated to be about 1mm) and it represents a gold standard in gait analysis.

3.2 2D-dimensional markerless gait analysis

Motion capture is an efficient approach to study the biomechanics of the lower limb and traditionally it has been used to diagnose the patho-mechanics related to the musculoskeletal diseases [13]. The motion analysis by means of optoelectronic stereophotogrammetry has been widely recognized as a useful clinical tool, but the routine clinical use of gait analysis is severely limited by the applicability of existing technology to clinical problems and by the costs and the length of time for data collection, processing (in fact the computation of the marker trajectories is not fully automated, and often requires the labelling and gap-filling from an expert operator) and data interpretation [14].

For this reason, a non-invasive and markerless gait analysis system represents a new important advancement in human motion capture because this approach reduces patient preparatory time, by eliminating the use of markers and the problems related (e.g. markers falling off during the analysis). Moreover, the costs associated to the markerless setup are much lower than those of the stereophotogrammetry setup which involves the use of multiple cameras. However, markerless methods are not largely widespread for clinical applications because the related accuracy is still lower than stereophotogrammetry. Recent advancements in the field of computer vision could boost the use of markerless technology for biomechanical and clinical applications.

One of the most challenging and difficult steps is the video acquisition and representation of human movement. For these reasons, specific systems are required to capture the images. The extraction of information from the images in the field of the computer vision is still challenging for clinical applications [14]. There are different cameras on the market i.e. Kinect II, Intel RealSense D435, Asus Xtion2 3D-Sensor, which differ in terms of acquisition and reconstruction of the image. Also, the approach to elaborate the images are different for the gait analysis in terms of implemented algorithms, models and specific regions investigated (upper limb, lower limb or whole body).

Regarding the implemented algorithms, they are divided into:

- Model based method: this approach includes a priori human model with kinematic and anatomic information and the model is matched to 2D image planes or 3D representations. The model consists in cylinders, stick figures, CAD models;
- Model free method: this approach is independent of the use of a priori model, but it captures skeleton features by representing the human motion in form of stick-figure through the use of Isomaps, Laplacian Eigenmaps and medial axis transform to obtain a pose-invariant graph for extracting joint kinematics from 3D representation.

This project thesis implements a model free method.

3.2.1 2D-dimensional markerless gait analysis – A literature review

Mundernann *et al* (2006) [14] develops a markerless protocol by using multiple cameras in order to obtain a 3D measurements of human motion for clinical applications. This method employs an articulated iterative closest point (ICP) with soft joint constraints to perform a human movement analysis in visual hull sequences (a 3D representation of dynamic sequences achieved by multiple images) as shown in Figure 3.2. By implementing soft joint constraints for tracking human body segments, a more correct matching between the models is obtained through the optimisation of an objective function.

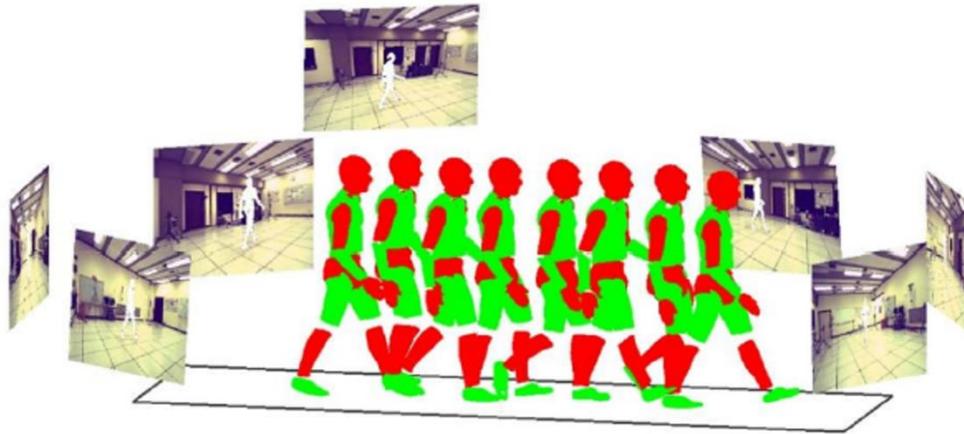


Figure 3.2: Selected visual hulls. Taken from Mundernann *et al* (2006).

In this study, an articulated body composed by fifteen body segments and fourteen joints was created by using a full body laser scan with markers affixed to the subject's joint as shown in Figure 3.3.

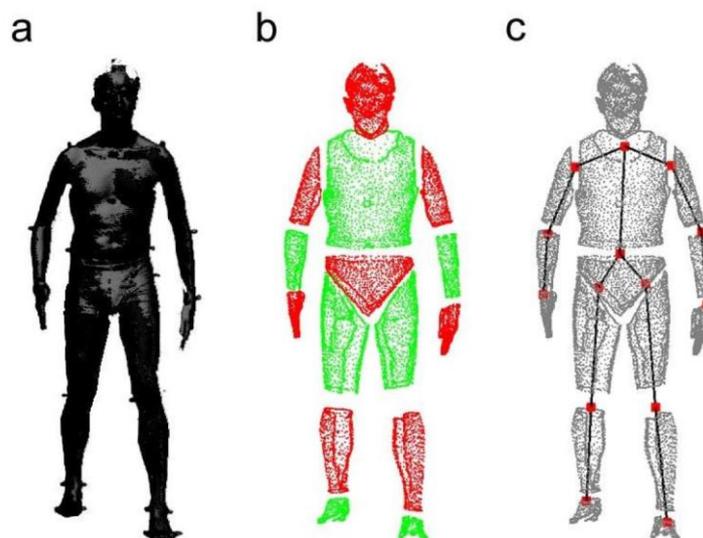


Figure 3.3: (a) Laser scan (b) Body segments (c) Joint centers. Taken from Mundernann *et al* (2006).

The articulated ICP algorithm is a modified version of the original ICP algorithm [15], [16]. This algorithm proposes to match an articulated model in a dynamic sequence of visual hulls.

The articulated model M is represented as:

- a set of points on the surface $\{p_1 \dots p_p\}$;
- a set of rigid segments $\{s_1 \dots s_p\}$;
- a set of joints $\{q_1 \dots q_q\}$.

Each visual hull V is represented as a set of points $\{v_1 \dots v_N\}$.

The subject's pose was matched to the first frame in the dynamic sequence of the gait cycle. The matching between the models brings the model M and model V through a series of rigid transformations T_j for each segment s_j .

The articulated ICP algorithm is divided into two steps:

- each point p_i is associated to its nearest neighbour $v_{s(i)}$ of the visual hull point V . A set of transformation T is computed to obtain the alignment of a set of corresponding pairs $(p_i, v_{s(i)})$;
- an objective function of the transformation variables is defined as:

$$F(T) = H(T) + G(T) \quad (3.1)$$

The first term ensures that the corresponding point are aligned, while the second term ensures that the joints are preserved.

$$H(r, t) = w_H \sum_{i=1}^P || R(r_{s(i)}) p_i + t_{s(i)} - v_i ||^2 \quad (3.2)$$

where:

- t_j is a 3x1 translation vector;
- r_j is a 3x1 rotation vector.

$$G(r, t) = w_G \sum_{(i,j) \in Q(M)} || R(r_i) q_{i,j} + t_i - R(r_j) q_{i,j} - t_j ||^2 \quad (3.3)$$

where:

- $q_{i,j}$ is the joint;
- w_G is a parameter that allows greater movement at the joint in order to improve the matching of the model M to the visual hull V .

Leu *et al* (2011) [17] propose a robust markerless vision-based human gait analysis able to extract gait features in order to estimate joint kinematics.

A gait analysis system must require functional clinical requirements to be useful in gait rehabilitation:

- easy experimental setup;
- easy to use by clinicians;
- low cost experimental setup.

This work is divided into five parts as shown in Figure 3.4.:

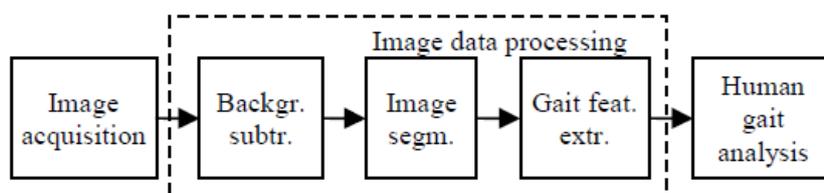


Figure 3.4: Block-diagram of proposed gait analysis protocol. Taken from Leu *et al* (2011).

- Image acquisition: in this works two cameras are used in order to obtain sagittal and frontal analysis as shown in Figure 3.5.

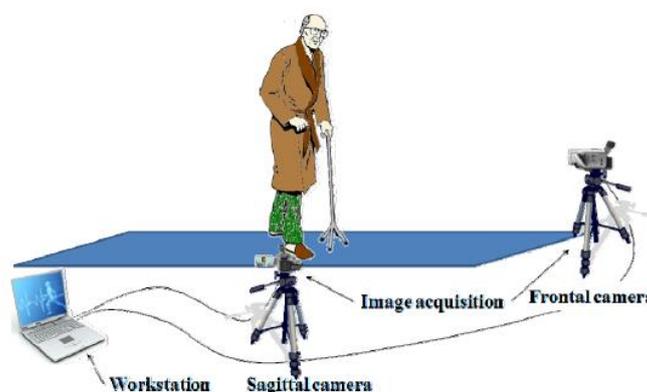


Figure 3.5: image acquisition with two cameras. Taken from Leu *et al* (2011).

- Background subtraction: this is a preliminary step in order to obtain the separation of the subject in the selected image. In particular, each frame of the video, containing the subject, is subtracted from the image containing only the background as:

$$r_{x,y,c} = abs(b_{x,y,c} - f_{x,y,c}) \quad (3.4)$$

where:

- x, y are image coordinates;

-
- c is the color channel;
 - b is the image containing only the background;
 - f is the image containing the subject.
- Image segmentation: the state of the art provides for a controlled experimental setup e.g. the presence of a coloured background to obtain the subject segmentation because this is a difficult step and it represents a key component of the majority of markerless gait analysis protocol [18]. A threshold algorithm is implemented to separate the subject from the background as presented in (3.5).

$$s_{x,y} = \begin{cases} 0, & \text{if } p_{x,y} \leq Th \\ 1, & \text{if } p_{x,y} > Th \end{cases} \quad (3.5)$$

where:

- x, y are image coordinates;
- p is the pixel value;
- Th is the threshold.

In order to choose the correct threshold, a two-dimensional entropy is evaluated as presented in (3.6).

$$S = - \sum_{i=0}^8 p_{(0,i)} \log_2(p_{0,i}) \quad (3.6)$$

where:

$$p_{(0,i)} = \frac{\text{number of segmented pixels surrounded with } i \text{ segmented pixels}}{\text{number of segmented pixel in the image}} \quad (3.7)$$

The pair $(0, i)$ represents the segmented pixel 0 surrounded with i segmented pixel in its eight-pixel neighbourhood.

The optimal threshold is the value that permits the minimum value of 2D entropy, because if the pixels are well connected (low entropy), the majority of pixels will be surrounded by eight pixels (except the border pixels) and this causes the 2D Entropy value to be small.

Gait feature extraction: The images obtained by segmentation are the input images for feature extraction. The extracted features are joint angles of upper and lower limbs, in particular the torso angle, right and left thigh angles and right and left shank angles both for sagittal and frontal plane as shown in Figure 3.6.

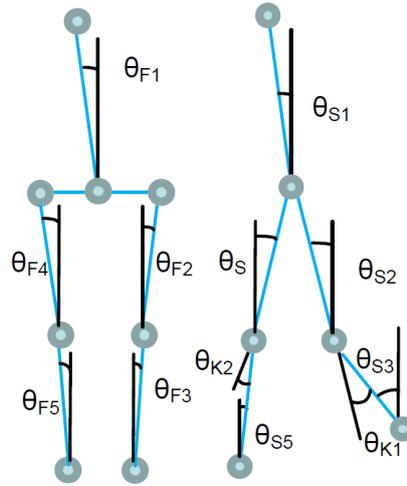


Figure 3.6: Extracted features in frontal and sagittal plane, respectively. Taken from Leu et al (2011).

To extract these features vertical and horizontal projections (only for the frontal plane) are defined as the number of segmented pixels in each row and as the number of white pixels in each column, respectively.

As shown in Figure 3.7, the vertical projection is used to determine the location of the neck joint, in particular the y coordinate corresponds to the minimum of the vertical projection while the x coordinate is represented by a middle point of the segmentation in the y row of the segmentation. The horizontal projection is used in order to identify arms regions as the right and left ‘hill’ of the horizontal projection. Other joints are extracted by using statistical anatomical measurements of the body segments [19], in particular:

- The hip joint is the center of the circles with $r = \text{length of the thigh}$;
- The knee joint is the middle of the arcs that represents the intersection between the circles and the segmented body parts identified as the leg.

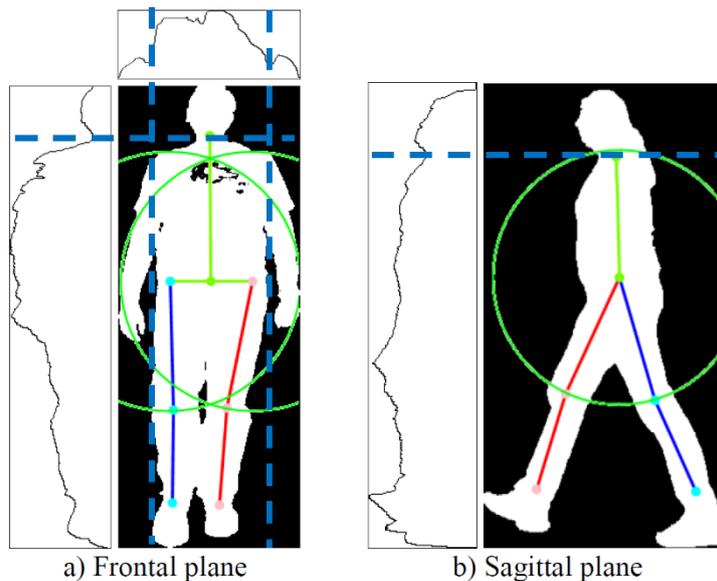


Figure 3.7: Coordinates of joints in segmented image. Taken from Leu et al (2011).

Also Goffredo *et al* (2008) [20] proposed a new method to estimate the lower limbs pose based on anatomical studies about human body anthropometric proportions [21]. This method is silhouette-based because the first step of this approach is the segmentation of the subject and the following extraction of his silhouette represented by (x_{sil}, y_{sil}) .

The vertical positions of hip, knee, ankle are defined as:

$$y_{hip} = \min(y_{sil}) + 0.5 \cdot H \quad (3.8)$$

$$y_{knee} = \min(y_{sil}) + 0.75 \cdot H \quad (3.9)$$

$$y_{ankle} = \min(y_{sil}) + 0.90 \cdot H \quad (3.10)$$

where H is the subject's height.

\mathbf{X}^* is defined as the set of x coordinates where $\mathbf{S}(\mathbf{X}^*, y_{knee}, t^*_\delta) = 1$ and $\mathbf{T}^* = [t^*_1, t^*_2, \dots, t^*_\delta]$ is the subset of frames where the legs do not overlap. These frames are identified as the frames in which \mathbf{X}^* forms 2 single connected components larger than $0.02H$ pixels.

The y coordinates belonging to the two shanks are defined as:

$$\mathbf{y}_{shin} = [y_{knee}, y_{knee} + 1, \dots, \min(y_{ankle}, y_{overlap} - 1)] \quad (3.11)$$

where $y_{overlap}$ includes the possibility of feet overlapping.

Then, the regions of images S_{shin_l} containing only the shanks are selected from the subject's segmentation and the right and left shanks ($l=1$ and $l=2$, respectively) are defined by the coordinates:

$$\begin{cases} x_{shin_l}(t^*_\delta) = [x_{1_l}, x_{2_l}, \dots, x_{s_l}] \\ y_{shin_l}(t^*_\delta) = y_{shin} \end{cases} \quad (3.12)$$

where

$$x_{s_l} = \frac{\sum_{j=1}^C x_j \cdot S_{shin_l}(x_j, y_s, t^*_\delta)}{\sum_{j=1}^C S_{shin_l}(x_j, y_s, t^*_\delta)} \quad (3.13)$$

The shins are linearly approximated to obtain $\mathbf{p}_l(x_{shin_l}, t^*_\delta)$.

The knee angle at frame t^*_δ is:

$$\alpha_l(t^*_\delta) = \pi - \arctan[p_{l1}(t^*_\delta)] \quad (3.14)$$

The hips position is computed with coordinates:

$$[x_{hip}, y_{hip}] = [x_1 + ((x_p - x_1) / 2), y_{hip}] \quad (3.15)$$

Subsequently, the finer hips pose estimation $q_l(t * \varphi)$ is obtained with a linear approximation of the thighs.

The hip angle at frame $t * \varphi$ is:

$$\gamma_l(t * \varphi) = \pi - \arctan [q_{l1}(t * \varphi)] \quad (3.16)$$

The extraction algorithm for shanks and thighs is shown in Figure 3.8.

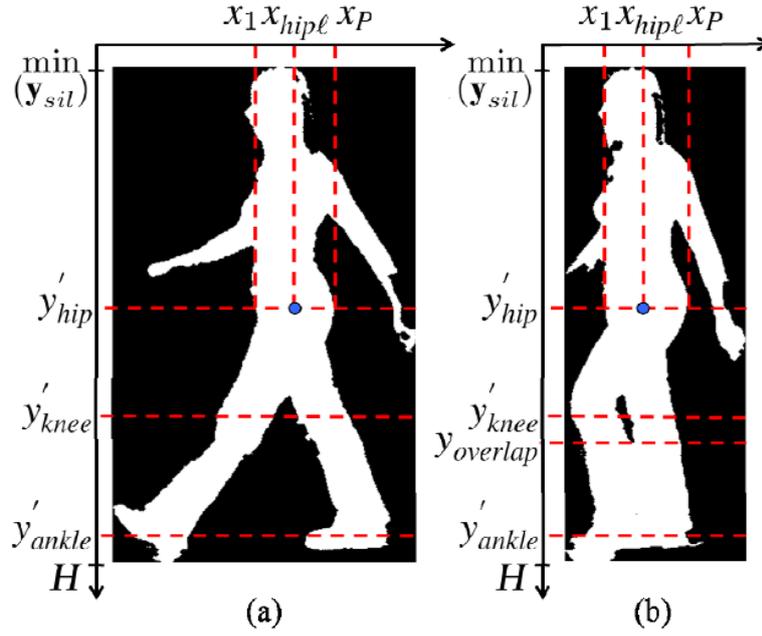


Figure 3.8: the extraction algorithm for shank and thigh for non-overlapped (a) and overlapped (b) feet. Taken from Leu et al (2011).

Metha et al (2017) [22] and Girshick et al (2011) [23] proposed two machine learning methods to estimate 3D human poses from a single camera RGB and a single camera RGB-D, respectively.

The first one implements a deep learning method by using a monocular RGB camera. In particular, this approach aims to obtain a 3D skeleton motion capture in real time and for this reason two steps are implemented:

- A convolution neural network (CNN) to predict joint positions both in 2D and 3D by reducing training time by avoiding the need to perform expensive bounding box computation as required in others deep learning methods;
- Model-based kinematic skeleton fitting to match the predicted 2D and 3D joint positions with a skeleton by optimizing a function in a least squares sense.

A general algorithm overview is shown in Figure 3.9.

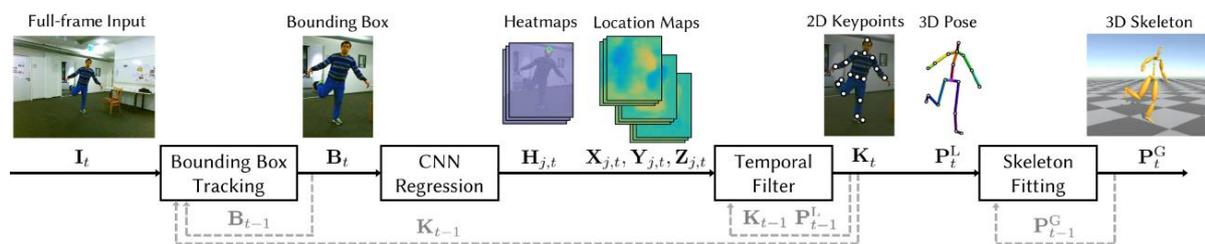


Figure 3.9: A general algorithm overview of Metha et al (2017). Taken from Metha et al (2017)

Girshick *et al* (2011) [23] presented a new approach to estimate general activity human pose through depth images by using a RGB-D camera as a Kinect sensor. In particular, this is a regression-based approach that allows to predict a 3D joint positions by using a regression forest defined as an ensemble of decision trees.

This method predicts human pose without the segmentation of the body but by using the raw depth image, so there is not the problem of occlusion.

4 2D markerless gait analysis protocol for clinical applications through the use of a single RGB-D camera

As outlined in Chapter 1, the aim of the thesis project is the development of a two-dimensional markerless gait analysis protocol through the use of a single RGB-D camera to estimate the lower limb joint kinematics for clinical applications.

The proposed protocol is a modified version of the protocols described in [2] and [5] to satisfy three functional requirements in clinic:

- the inability to use a machine learning approach because clinical applications i.e. Cerebral Palsy, Parkinson or stroke do not allow the availability of a large image data set fundamental to train different types of network [22],[23] for machine and deep learning methods;
- the choice not to automatically identify the anatomical landmarks because this approach neglects the subject-specific characteristics, possibly leading to joint angles estimation inaccuracies. In fact, this method may not be appropriate for different populations of pathological subjects who deviate from the proposed models. However, the collaboration of an expert operator is required only at the start of the data processing to identify the anatomical landmarks in the references frames for creating the models. This step takes only few seconds in a clinical field where it is not required to process a large quantity of images;
- the use of simple and low-cost setup with only one camera to introduce a new generation of simple markerless gait analysis beneficial to an extensive monitoring program for gait diseases.

The proposed protocol for markerless gait analysis consists in the three main parts described below.

- Cycle segmentation: two consecutive heel strikes are manually identified to determine the gait cycle to be analysed. Furthermore, the load phase and the swing phase are identified.
- Subject segmentation: a modified version of (Salvi and Molinari, 2018) subject by introducing anatomical and biomechanical hypothesis. The human segmentation algorithm used in this thesis is a fully automatic thresholding method which is evaluated on the image histogram resulting from the subtraction between a frame containing the subject and a frame with only the background captured at the beginning of the trial.
- Two-dimensional joint kinematics estimation: the implementation of a modified version of Pantzar-Castilla *et al* (2018) to extract the foreground lower limb and to estimate the lower limb joint sagittal kinematics through a participant-specific multi-segmental model which allows to generalize the proposed method also for subjects with gait alterations.

More specifically, the block diagram of the implemented steps is shown in Figure 4.1.

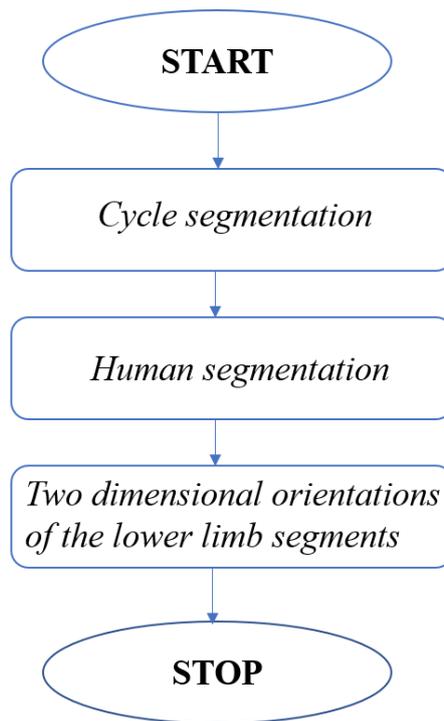


Figure 4.1: General overview of the proposed protocol for estimating the sagittal lower limb joint kinematics

An experimental session in which the subject is asked to walk on a green walkway was also conducted for three reasons:

- to compare the image quality from Kinect II (no longer in production today) [2], and Intel RealSense D435, to avoid blurred images obtained by the acquisition with Kinect 2 caused by a fixed time of exposure;
- to test how the light condition affects the quality of the estimated joint kinematics. These conditions may influence the quality of the segmentation since the automatic choice of a threshold in low light is difficult because of the low image contrast;
- to test how the presence of a green background behind the subject have an impact on the quality of the segmentation and consequently on the estimate of the joint kinematics.

All the implemented steps are described in detail in the following paragraphs. In chapter 5, the experimental session and the obtained results will be presented.

4.1 Cycle segmentation

In this first step of the proposed protocol, two consecutive heel strikes are manually identified to determine the gait cycle to be analysed. The choice is to select, among the different gait cycles captured for each trial (e.g. two or three, it depends on the stride length), the central one to reduce the distortions introduced by the camera at the edges of the images. Furthermore, the load and the swing phase are also identified. to estimate the lower limb joint kinematics.

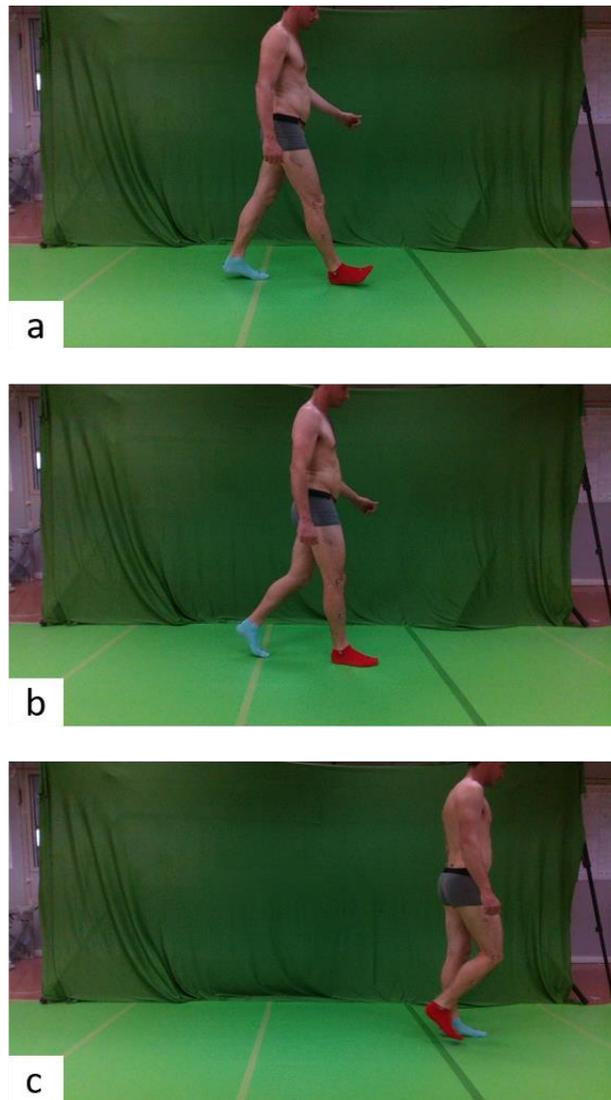


Figure 4.2: Cycle segmentation: Heel strike (a), Load phase (b), Swing phase

4.2 Human segmentation

The segmentation is the second step of this two-dimensional markerless gait analysis protocol because it allows to separate the subject from the background to analyse only the region of the image containing the participant. The segmentation becomes fundamental since the subject mask permits to obtain a 2D lower limb model used to estimate the joint kinematics.

The state of the art of markerless gait system introduces the need of a controlled environment e.g. a green background [2] in order to simplify the segmentation algorithm which is based on the removal of the green color from the image by implementing a color filter.

The human segmentation algorithm used in this thesis is a fully automatic thresholding method because it is based on the automatic setting of a threshold obtained from the evaluation of the histogram of the image resulting by the subtraction between a frame containing the subject and a frame representing only the background captured at the beginning of the trial. For this reason, this algorithm can be applied in every type of experimental set up condition because the threshold is chosen only on the basis of the image histogram.

The only requirement for the subject is the presence of two different coloured socks (in this case red for the right foot and blue for the left) used to easily identify the feet.

The segmentation algorithm is divided into five parts as shown in Figure 4.3.

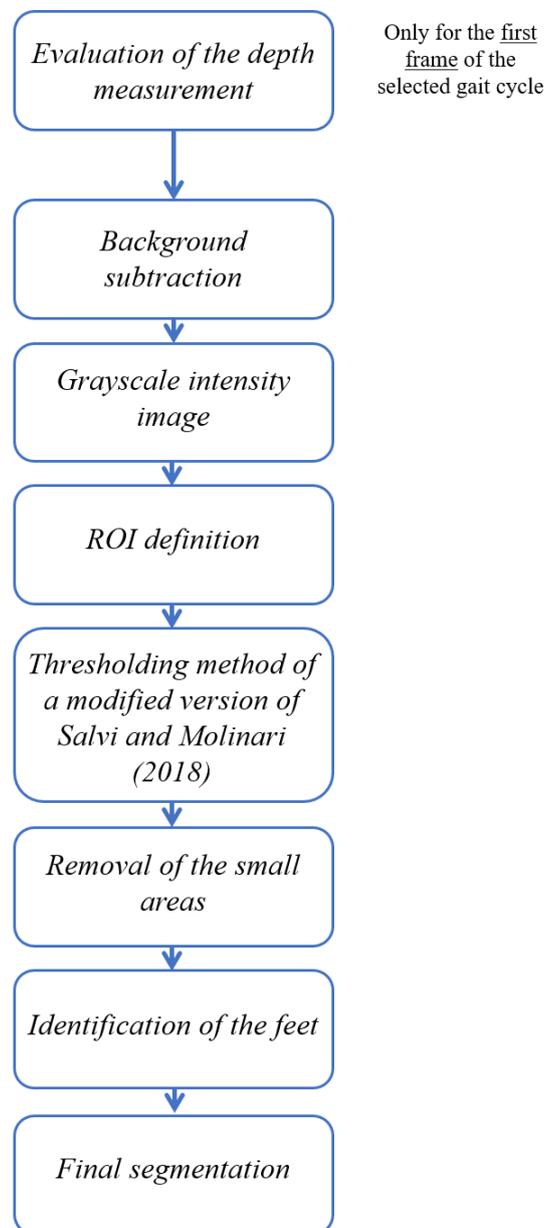


Figure 4.3: General overview of 'Human segmentation algorithm'

Each single step of the segmentation algorithm is detailed below:

- Evaluation of the depth measurements: by using the depth image (Figure 4.4), it is possible to remove the regions of the RGB image that are outside of a define range of depth measurements which includes the subject. This range is defined in the first frame by evaluating the depth measurement of the selected gait cycle.
The sum of the depth values along rows and columns are evaluated as defined in equation (4.1) and (4.2)

$$c = \sum_{i=1}^N Depth(i, j) \quad (4.1)$$

where:

- a. c is a vector containing the sum of the depth values along the rows;
- b. $Depth$ is the depth image;
- c. i and j are the image coordinates;
- d. N is the number of the rows in the depth image;

$$r = \sum_{j=1}^M Depth(i, j) \quad (4.2)$$

where:

- a. r is a vector containing the sum of the depth values along the columns;
- b. $Depth$ is the depth image;
- c. i and j are the image coordinates;
- d. M is the number of the columns in the depth image;

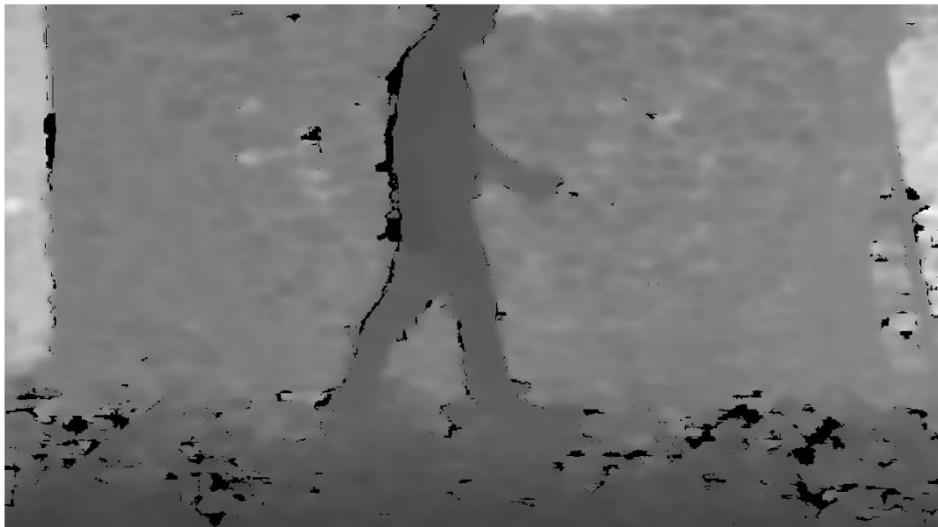


Figure 4.4: Depth image acquired by using the depth sensor from Intel RealSense D435. The subject is closer to the camera than the background. For this reason, the subject is characterized by the lowest depth values (darkest colours) in the image.

The subject is characterized by the lowest depth values in the depth image while the highest ones represent the background. For this reason, the minimum of each vector r and c is calculated in order to obtain the x and y coordinates of the lowest depth values in the image called r_{min} and c_{min} as shown in Figure 4.5 and Figure 4.6.

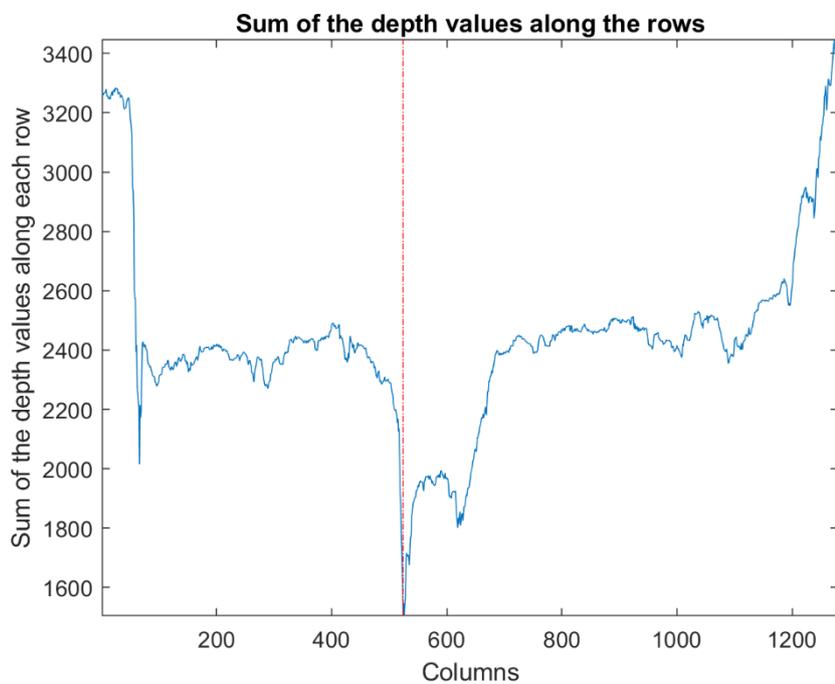


Figure 4.5: Vector c containing the sum of the depth values along the rows. The red line is the position of the c_{min} (the column containing the lowest depth value).

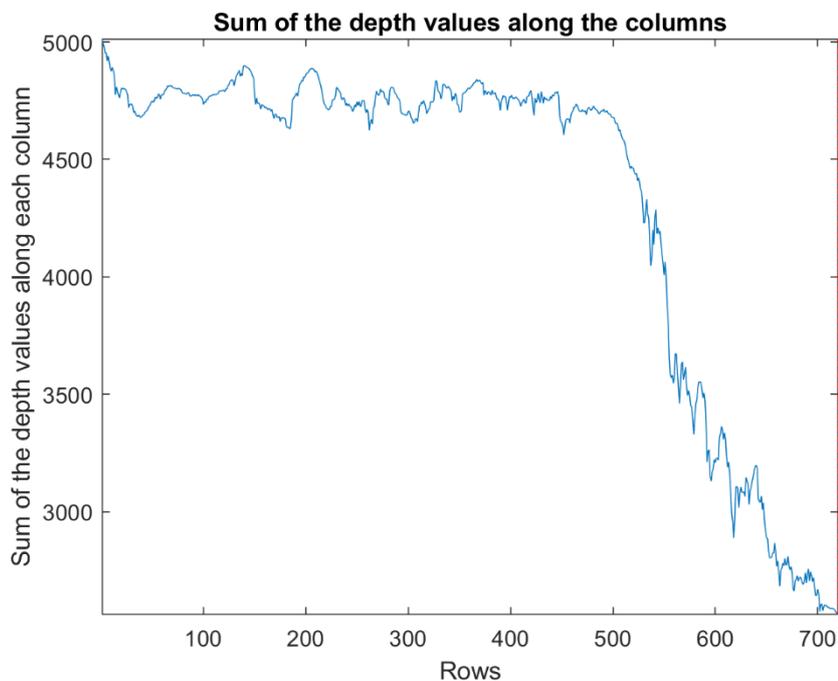


Figure 4.6: Vector r containing the sum of the depth values along the columns. The red line is the position of the r_{min} (the row containing the lowest depth value).

In this way, the inferior limit of the depth values range is obtained by considering the depth value corresponding to r_{min} and c_{min} defined as:

$$d_{min} = Depth(r_{min}, c_{min}) \quad (4.3)$$

The superior limit of the depth values range is defined by considering that:

1. The maximum value of step width is about 10 cm;
2. The width of the body trunk is about 50 cm;
3. The position of the contralateral arm of the subject is around 1 m, by considering the swing of this on frontal plane.

For these reasons, the superior limit of the depth values d_{max} is approximately 3 meters.

$$range = [d_{min}, d_{max}] \quad (4.4)$$

After removing the depth values outside the range, the resulting image $Depth$ is shown in Figure 4.7.



Figure 4.7: Depth image obtained after the removal of the depth values outside the depth values range.

In order to obtain the final RGB images, the intersection between $Depth$ and I is implemented, and the result is shown in Figure 4.8.

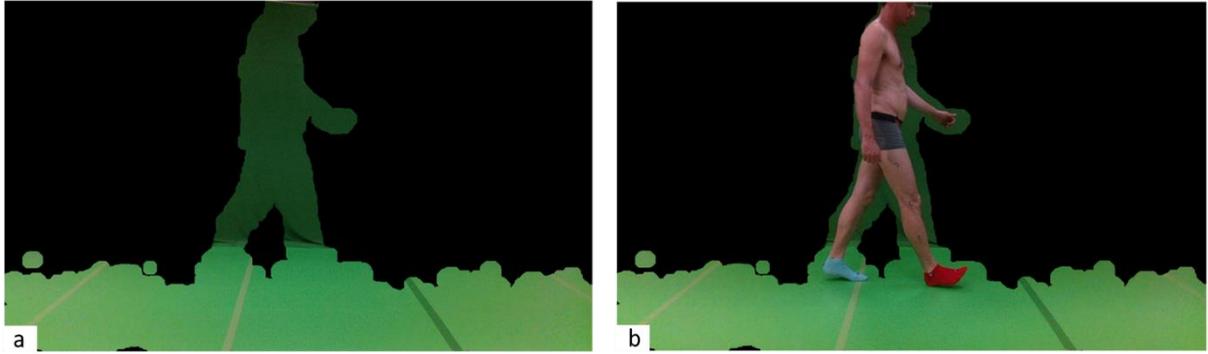


Figure 4.8 Background image (a) and image containing the subject (b) after the removal of the depth values outside the depth values range.

It is possible to notice that the depth value range includes also the green carpet under the feet of the subject because this region of the background presents the same depth values of the subject.

- Background subtraction: this is a preliminary step to obtain the separation of the subject in the selected image. In particular, each frame of the video, containing the subject, is subtracted from the image containing only the background as:

$$D_{x,y,c} = abs(B_{x,y,c} - F_{x,y,c}) \quad (4.5)$$

where:

- x, y are image coordinates;
- c is the color channel of RGB images;
- B is the background image;
- F is the image containing the subject.
- D is the difference image.

This step is fundamental because in difference image D the intensity of the pixels belonging to the subject are higher than the intensity of the ones belonging to the background in order to simplify the automated choice of a threshold to separate the subject from the background, as shown in Figure 4.9.

Ideally, the difference between the two considered images (B and F) should be equal to zero in correspondence to the pixels belonging to the background, since the intensity of these pixels is exactly the same between B and F . This condition is difficult to achieve because of the noise which inevitably affects the images.



Figure 4.9: Difference image D . It is possible to notice that the intensity of pixels belonging to background is lower than the intensity of ones belonging to the subject.

- Transformation of the difference image into the grayscale intensity image: this step permits to convert true RGB color to the grayscale intensity image. To obtain a grayscale intensity image (Figure 4.10) the equation is implemented:

$$D = \sqrt{D_{x,y,c1}^2 + D_{x,y,c2}^2 + D_{x,y,c3}^2} \quad (4.6)$$

where:

- $c1$ is the red channel of RGB images;
- $c2$ is the green channel of RGB images;
- $c3$ is the blue channel of RGB images;

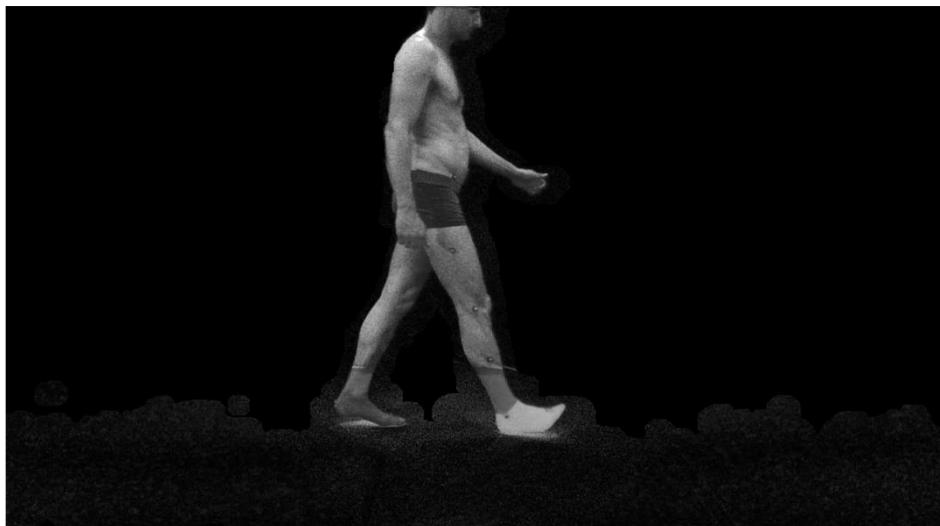


Figure 4.10: Difference image D in grayscale.

-
- Region of interest (ROI) definition: a region of interest is a rectangular area containing the subject. This step is implemented to remove the remaining pixels belonging to the background inside the range of the depth measurements. The difference image D is evaluated in order to obtain this area. In particular, the sum of the pixel intensities along rows is evaluated as defined in equation (4.7):

$$c = \sum_{i=1}^N D(i, j) \quad (4.7)$$

where:

- a. c is a vector containing the sum of the depth values along the rows;
- b. D is the depth image;
- c. i and j are the image coordinates;
- d. N is the number of the rows in the depth image;

The majority of the pixels in the difference image D belong to the background so the mode m of the vector c is calculated to define a threshold Th for identifying a ROI which contains the subject as follows

$$Th = 3 \cdot m \quad (4.8)$$

The mode has been multiplied for a small constant value because the baseline presents some peaks, due to the noise of the image, that must be excluded in the ROI of the subject. The ROI is defined as in equation (4.9) and the results are shown in Figure 4.11 and Figure 4.12.

$$ROI_{x,y} = \begin{cases} 0, & \text{if } p_{x,y} \leq Th \\ 1, & \text{if } p_{x,y} > Th \end{cases} \quad (4.9)$$

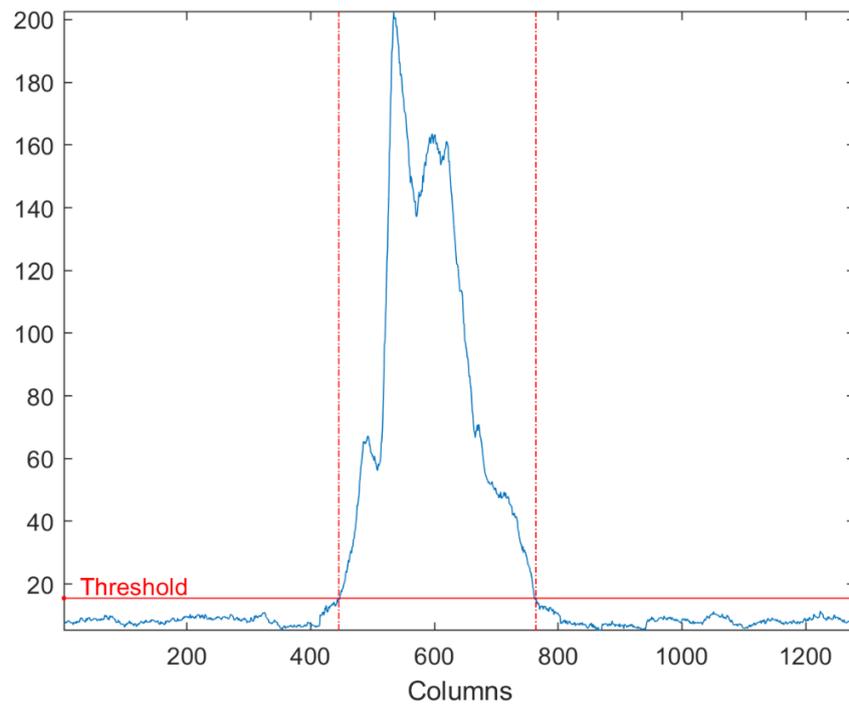


Figure 4.11: Vector r containing the sum of the pixel intensities along the rows. The horizontal red line represents the threshold Th used to obtain the ROI containing the subject (vertical red lines).

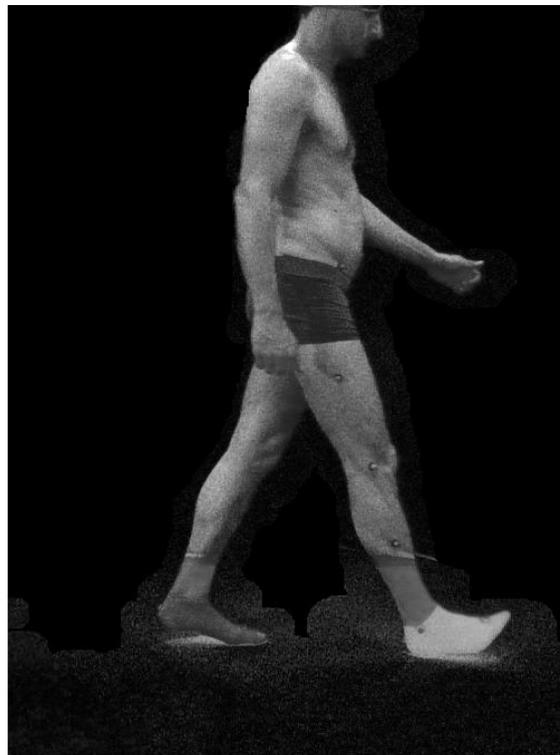


Figure 4.12: The image obtained after the ROI definition.

-
- Automatic thresholding algorithm: a modified version of Salvi and Molinari (2018) [24] is implemented in order to obtain an automatic threshold by implementing the steps shown in Figure 4.13.

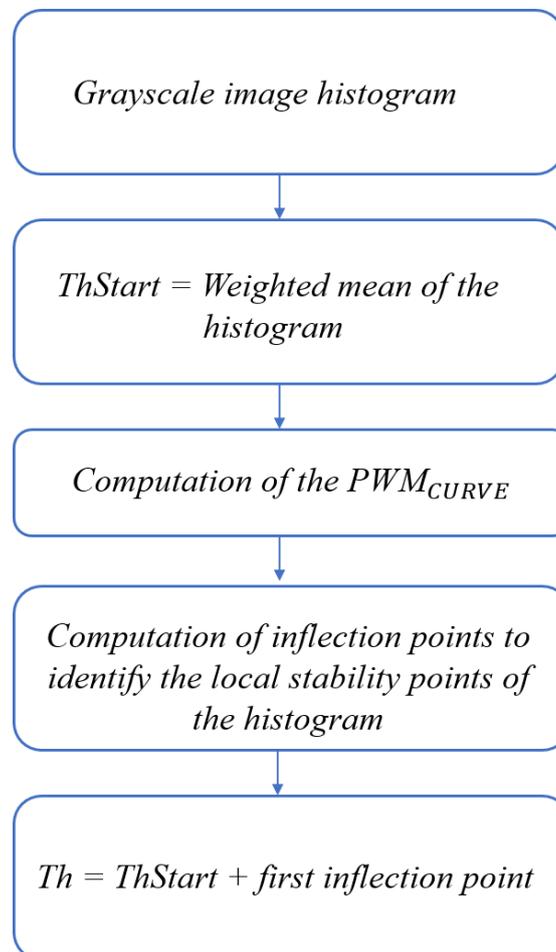


Figure 4.13: General overview of the modified version of Salvi et Molinari (2018)

The grayscale image ROI is composed by pixel intensities represented by integer numbers between 0 and N, where N = 255.

The image histogram is calculated, as shown in Figure 4.14.

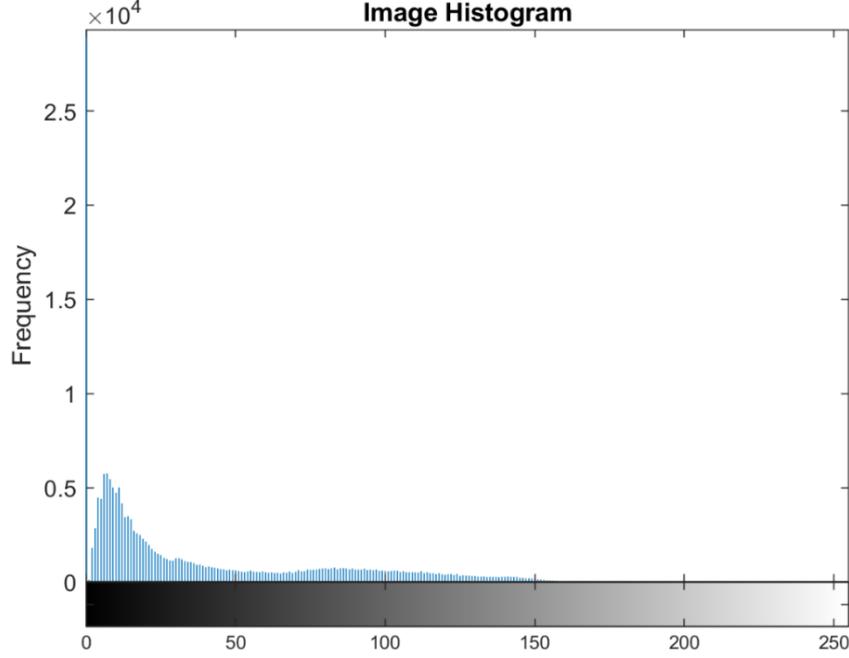


Figure 4.14: Image histogram of the grayscale image ROI. The histogram is composed by N+1 classes and it represents how many times each grey level occurs.

The histogram is a distribution composed by N + 1 classes and it represents graphically how many times each grey level occurs.

The weighted mean of the histogram is computed to obtain a threshold as follows:

$$WM = \frac{\sum_{i=0}^N w_i x_i}{\sum_{i=0}^N w_i} \quad (4.10)$$

where:

- a. w_i is the histogram count for the i th class;
- b. x_i is the bin location for the i th class.

This value represents the initial threshold called $ThStart$ as shown in Figure 4.15.

Subsequently, The progressive weighted mean (PWM_{CURVE}) of the histogram is computed as defined in equation (4.11)

$$PWM_{CURVE} = \frac{\sum_{i=ThStart}^N w_i x_i}{\sum_{i=ThStart}^N w_i} \quad (4.11)$$

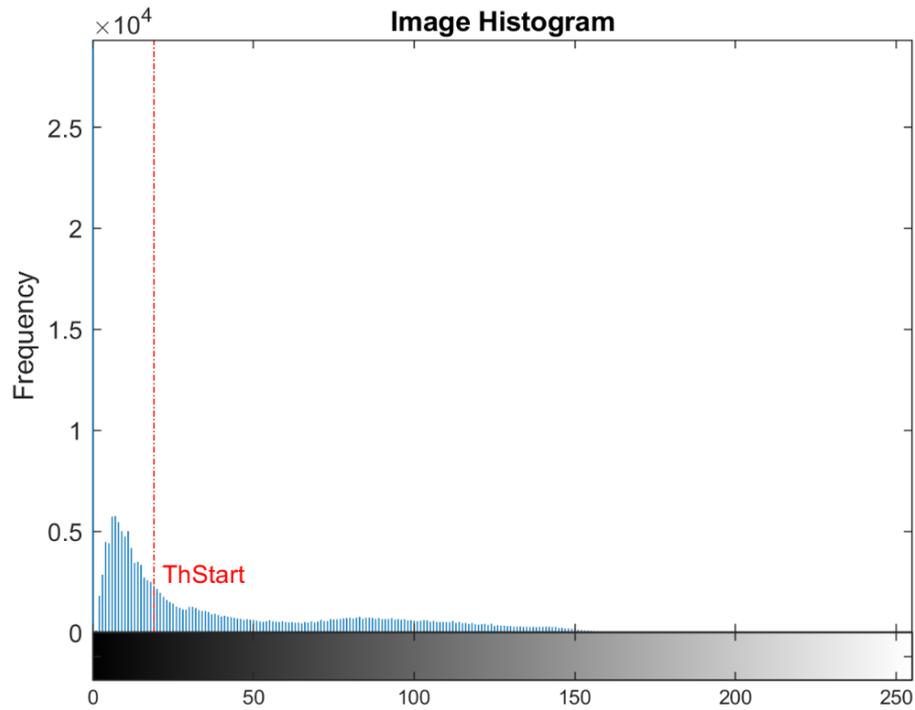


Figure 4.15: Image histogram of the grayscale image ROI. The red line represents the initial threshold called Th_{Start} .

The PWM_{CURVE} is evaluated for each class of the histogram as the weighted mean of all the grayscale histogram values up to that class as shown in Figure 4.16. The trend of this curve represents the histogram one, for this reason relevant characteristics on the color distribution of the image can be extracted by using this graph.

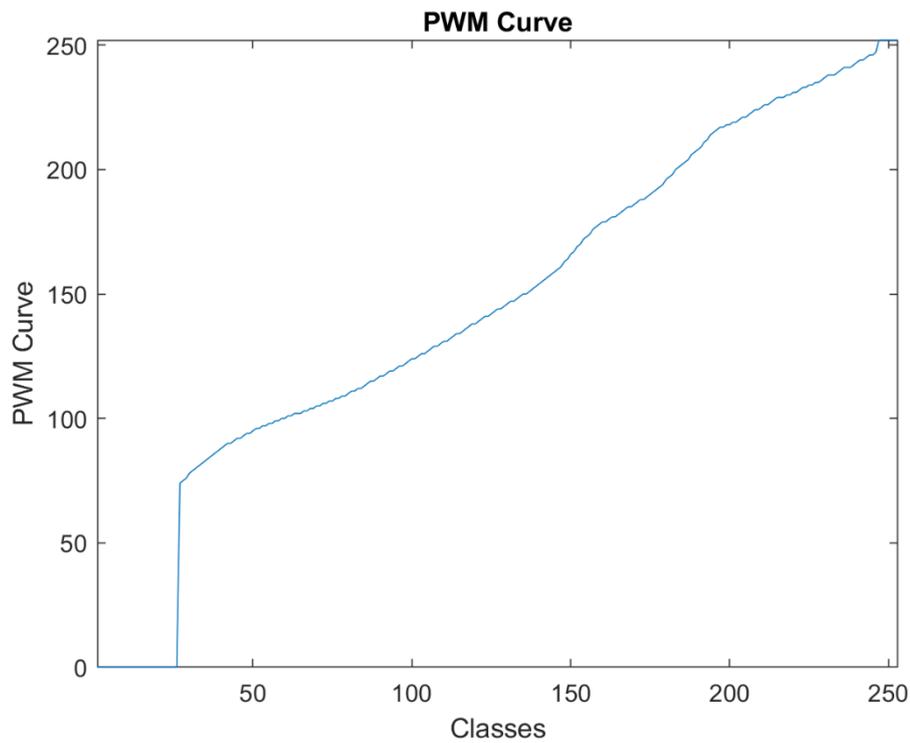


Figure 4.16: PWM_{CURVE}

In particular, if there are significant color variations from a certain point on the histogram with respect to the distribution that precedes it, here we can expect to see a change of concavity in the PWM_{CURVE} . Inflection points of PWM_{CURVE} may be potential threshold values for performing subject segmentation as they represent local stability points of the grayscale histogram.

For this reason, the inflection points of the PWM_{CURVE} are calculated. In the reference article, a method for the nuclei segmentation is explained and for this reason it is necessary to segment multiple objects within the image, therefore all the inflection points are calculated and the one corresponding to the highest median area is identified.

In this specific application, since there is always a subject, it was decided to change this step by taking the intensity value corresponding to the first inflection point.

The final threshold is defined as:

$$Th = ThStart + Inf_point \quad (4.12)$$

where Inf_point is the intensity value of the first inflection point. In this way, it is possible to identify the local stability, representing by the inflection point, closer than $ThStart$.

The final segmentation S is defined as follows:

$$S_{x,y} = \begin{cases} 0, & \text{if } p_{x,y} \leq Th \\ 1, & \text{if } p_{x,y} > Th \end{cases} \quad (4.13)$$

The final segmentation S obtained by implemented the modified version of Sal *et al* (2018) [24] is shown in Figure 4.17.

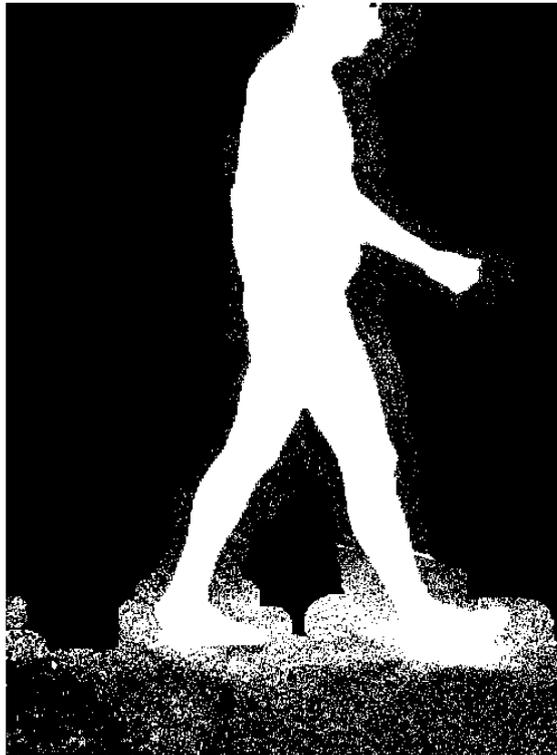


Figure 4.17: The final segmentation S obtained by implemented the modified version of Sal *et al* (2018)

It is possible to notice, due to the image noise, the presence of some small areas, primarily near the feet, which are outside the analysed subject.

- Removal of the small areas: it is important to identify the higher area which includes the subject in order to remove other small areas. For this reason, opening of S from a structural element B is computed as follows:

$$S' = S \circ B = (S \ominus B) \oplus B \quad (4.14)$$

where \ominus and \oplus are the erosion and expansion operations, respectively.

The final segmentation S' , after the application of these morphological operators, is shown in Figure 4.18

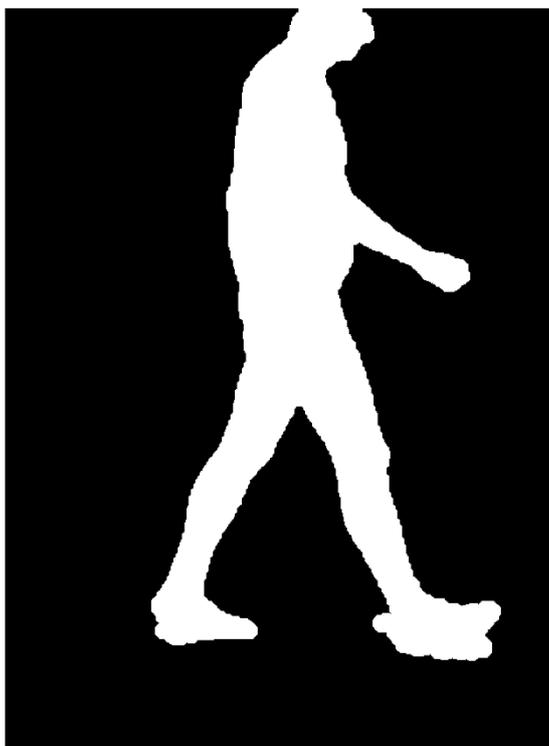


Figure 4.18: The final segmentation S' obtained after the application of these morphological operators

The opening of S from a structural element B removes the smaller objects of the structural element preserving the background. It is possible to notice that most of the small areas have been removed, however the feet are not yet well defined because around them there are regions belonging to the background (in particular belonging to the green carpet). For this reason, the identification of the feet may be refined.

-
- Identification of the feet: the subject wears coloured socks (red per right foot and blue for left foot) because these will be useful to identify the ankle joint (as it will be discussed in subchapter 4.3.2.1). The feet are identified (Figure 4.19) through a color filter applied in RGB and HSV (Hue saturation value) space.



Figure 4.19: Identification of the feet. The red lines represent the right foot (red) and the left foot (blue).

Subsequently, the identified feet are removed from the mask in order to create holes and to remove the areas around the feet belonging to the carpet as shown in Figure 4.20.

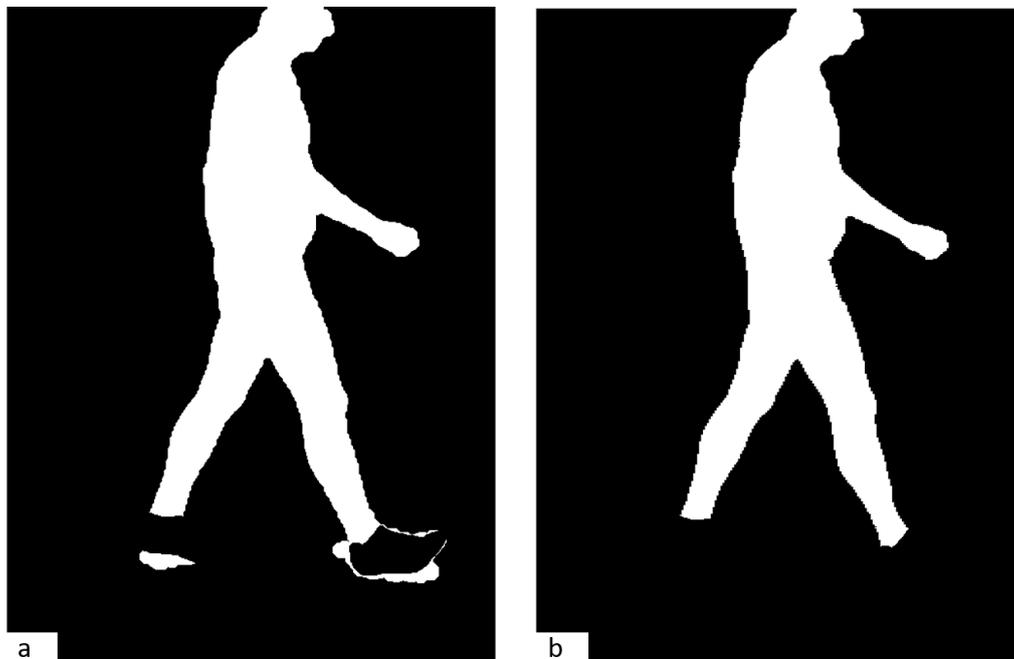


Figure 4.20: The identified feet are removed from the mask (a). The areas around the feet are removed because smaller than the other regions of the mask (b).

The final segmentation, obtained by the implementation of all the steps described, is shown in Figure 4.21.

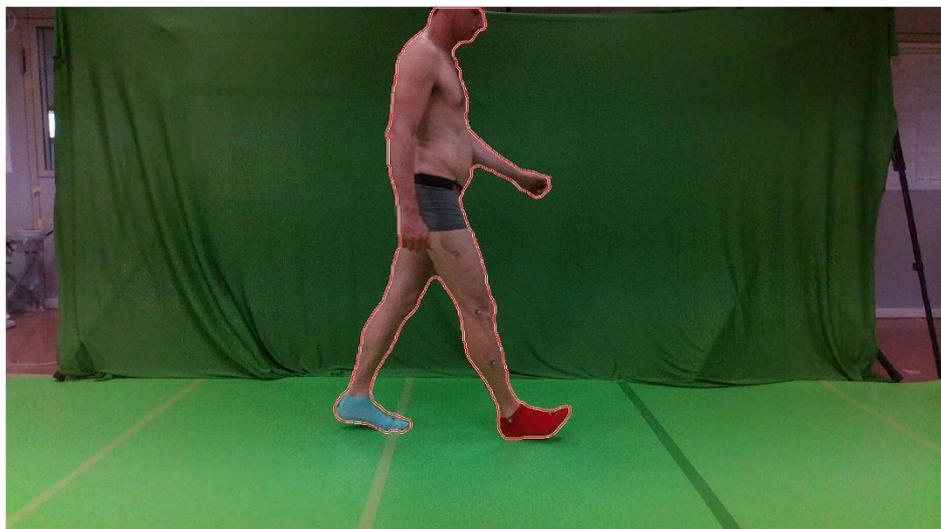


Figure 4.21: The red line represents the final segmentation obtained by the implementation of all the steps described in 'Human segmentation'.

4.3 Two dimensional orientations of the lower limb segments

The proposed method for estimating the sagittal joint kinematics of the lower limb is a model-based method because includes a priori human model with kinematics and anatomic information and the model is matched to 2D image planes.

The implemented algorithm is composed by four different sections:

- participant-specific multi-segmental model (calibration phase) which is necessary to introduce a lower limb model composed by three body segments (foot, shank and thigh) and their joint centres;
- extraction of the three body segments from the depth image and the correspondent masks. In this phase the foot, shank, and thigh are identified using the depth image acquired by the depth sensor;
- joint center tracking in which the trajectories of hip, knee and ankle joint centres are estimated by using the Iterative Closest Point (ICP) algorithm;
- computation of lower limb joint angles: the orientations of the lower limb segments are determined from joint centre trajectories.

4.3.1 Participant-specific multi-segmental model

The two-dimensional lower limb kinematics model was composed of three body segments (foot, shank, and thigh) connected by three hinges representing the ankle joint centred at the lateral malleolus (LM), the knee joint centred at the lateral epicondyle (LE), and the hip joint centred at the great trochanter (GT). The chosen kinematic model has six degrees of freedom: foot position and inclination in the sagittal plane (two degrees of freedom for position and one for inclination) and joint angles at the ankle, knee and hip (three degrees of freedom). Right and left sides were treated independently.

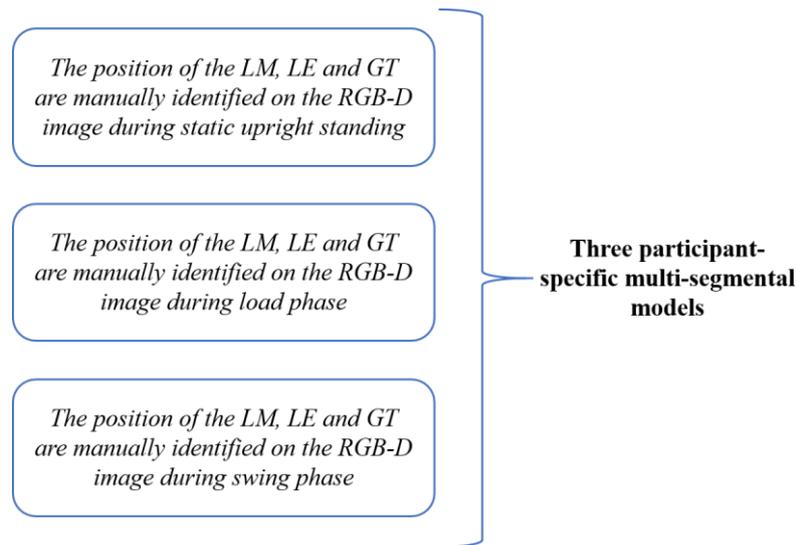


Figure 4.22: Three participant-specific multi-segmental models

It should be noted that, while the participant walks within the field of view of the RGB-D camera, the images of each body segment undergo both shape and size changes from one frame to the next. These changes are due to concurrent deformations of the body segments caused by the presence of soft tissue artefacts (Cereatti et al., 2017 [25]) and to variation of the two-dimensional images caused by parallax effects, variations in the relative distance between the RGB-D camera and the participant, and uncontrolled subject rotations out of the sagittal plane (e.g. pelvic rotation), as shown Figure 4.23.



Figure 4.23: Variation of the two-dimensional images caused by the parallax effects due to the relative distance between the RGB-D camera and the participant. In particular, when the subject is close to the extremities of the FOV of the camera (b) there is a variation of shape and size of the body segments. This variation is absent when the subject is in the center of the FOV of the camera.

To compensate for these artefacts, three distinct participant-specific lower limb models (foot–shank–thigh) are defined for the estimation of the knee joint angles. The first participant-specific lower limb model is defined from the RGB-D image recorded during static upright standing, as shown in Figure 4.24.

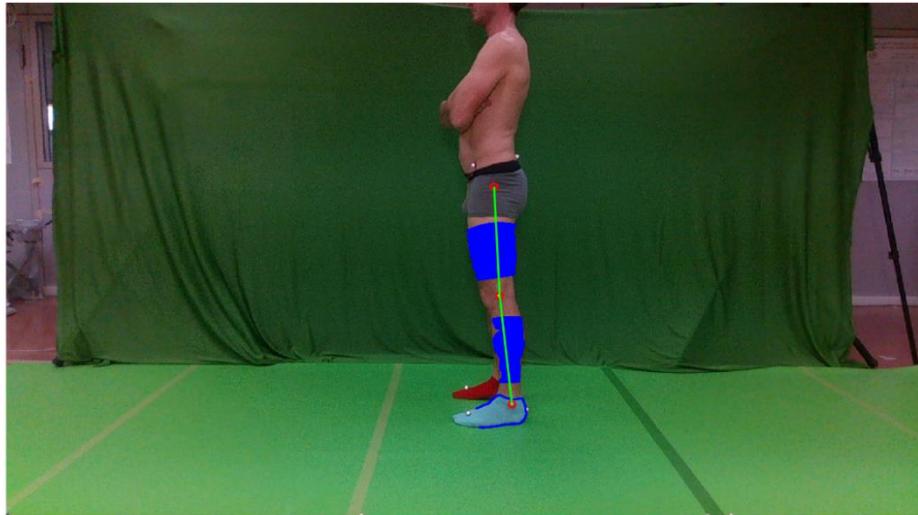


Figure 4.24: Participant specific lower limb model during static upright standing.

The positions of LM, LE, and the greater trochanter (GT) were manually identified on the RGB-D image by an operator using a mouse click on the relevant markers. The foot model is defined from the segmented image as the posterior half of the foot contour, including the LM position. The shank model is defined as the middle portion of the shank segment (between the 25% to 75% of the segment length, defined as the distance between LM and LE) including both LM and LE positions. Similarly, the thigh model is defined as the middle portion of the thigh segment (between 25% and 75% of the segment length, defined as the distance between LE and GT) including both LE and GT positions. The other two participant-specific kinematic models are defined following the same procedure from the RGB-D images recorded for the percentage of the gait cycle for which the participant exhibits the maximal knee flexion angles in accordance with the knee loading response and the swing phases, respectively as shown in Figure 4.25 and Figure 4.26.

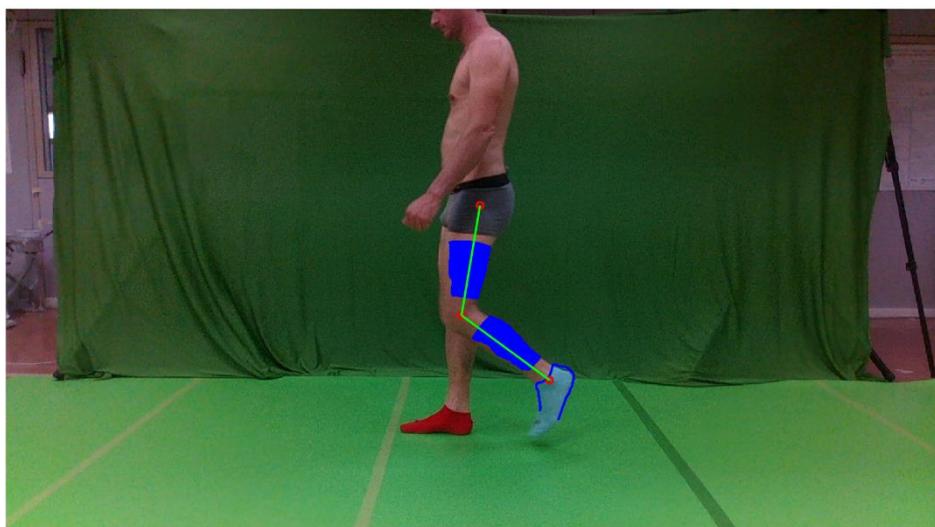


Figure 4.25: Participant specific lower limb model during flexion upright standing.



Figure 4.26: Participant specific lower limb model during load upright standing

4.3.2 Extraction of the three body segments

For the joint centre tracking, the preliminary and fundamental step is the extraction and the identification of the three body segments (foot, shank and thigh) which compose the lower limb model.

4.3.2.1 Foot identification

The algorithm used for the 'Foreground foot identification' is shown in Figure 4.27.

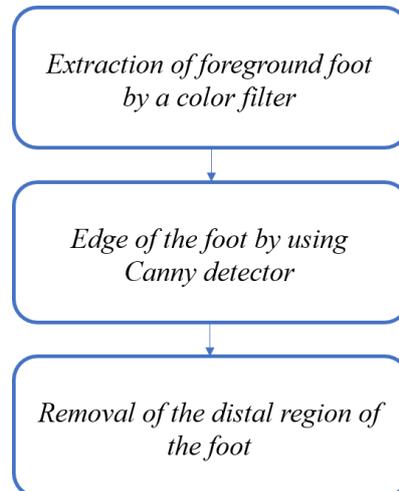


Figure 4.27: Block diagram of 'Foreground foot identification'.

The foreground foot is extracted by using a colour filter (red for the right side and blue for left side) as shown in Figure 4.28.



Figure 4.28: Foreground foot identification. The yellow surface represents the red foot

Subsequently, the edge of the extracted foot is computed by means of Canny's edge detector, and the distal region of the foot is removed as shown in Figure 4.29.

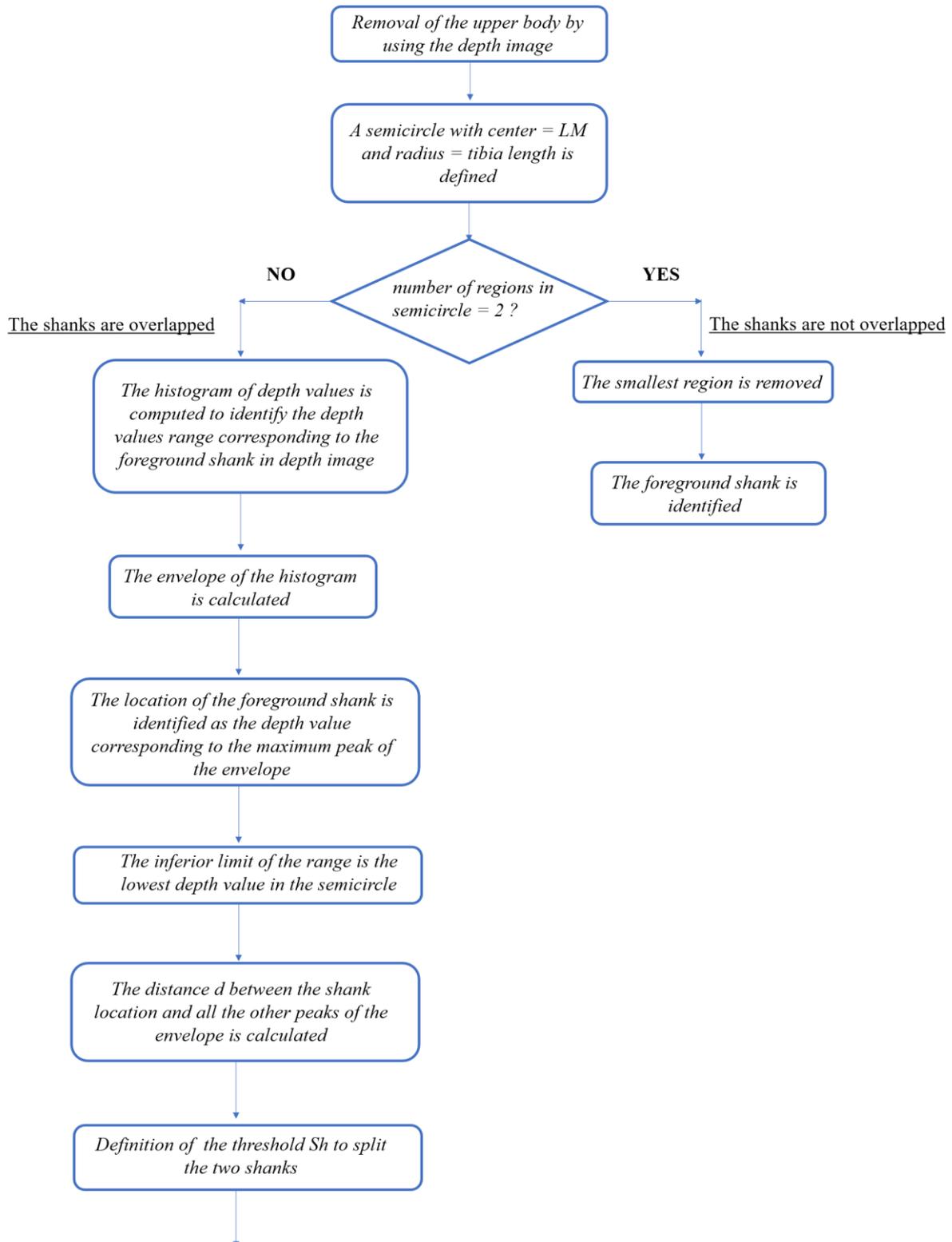


Figure 4.29: Foreground foot identification: the red line represent the edge of the extracted foot after the removal of the distal region

To remove the distal region of the foot, the sole of the foot is computed as the line best fitting the point between heel and metatarsus identified as portions of the foreground foot. The threshold for cutting the foot is chosen by analysing the images. The distal region of the foot is removed because it represents a non-rigid part so unrepresentative of the lateral malleolus (LM) position and orientation and for this reason it could cause errors in the matching of the foot in all the other frames (it will explain in 4.3.3).

4.3.2.2 Shank identification

To identify and to extract the foreground shank, the depth image is considered. The block diagram used to the 'shank identification' is shown in Figure 4.30. Firstly, the segmented image is overlapped on the depth image (Figure 4.31) to isolate the subject and to remove the other parts of the depth image.



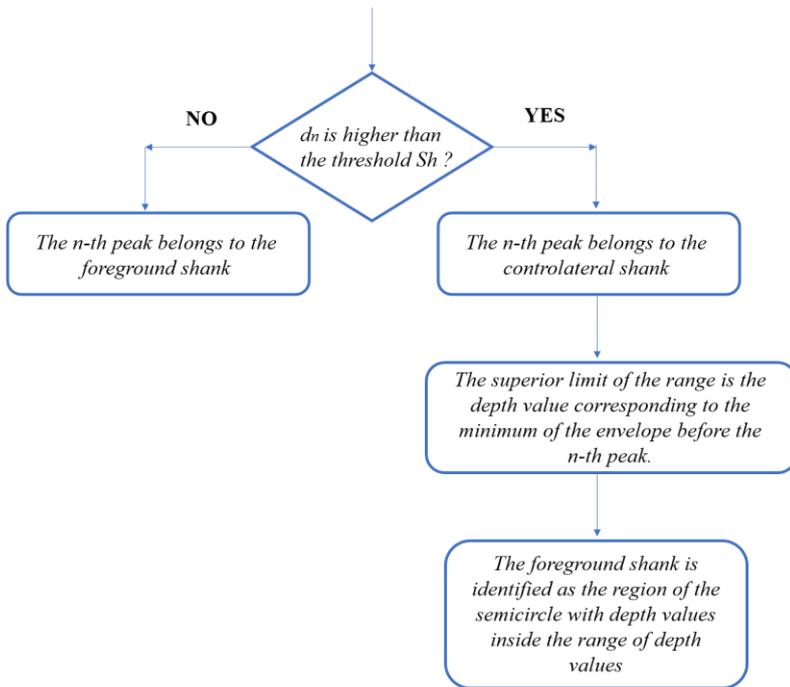


Figure 4.30: Block diagram of 'Foreground shank identification'

Firstly, the segmented image is overlapped on the depth image (Figure 4.31) to select only the subject and to remove the other parts of this image.



Figure 4.31: Segmented image (yellow surface) is overlapped on the depth image in order to selected only the subject

In other words, as shown in Figure 4.32 the upper body is removed and only the portion of the mask below the coordinate of the GT is considered to be analysed subsequently.

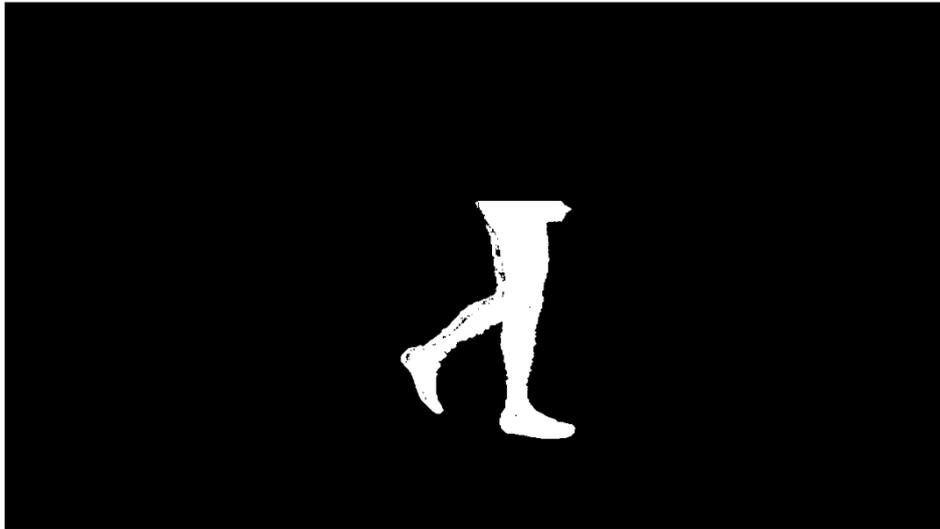


Figure 4.32: Lower limb segmentation. The segmented image above the GT is removed

A semicircle with centre in the LM and radius equal to the length of the tibia, defined as Euclidean distance between the LM and the LE manually identified during the calibration phase, is traced to remove the body regions outside this circle.

After this step, two situations are possible:

1. The two shanks are not overlapped (Figure 4.33 - a);
2. The two shanks are overlapped (Figure 4.33 - b).

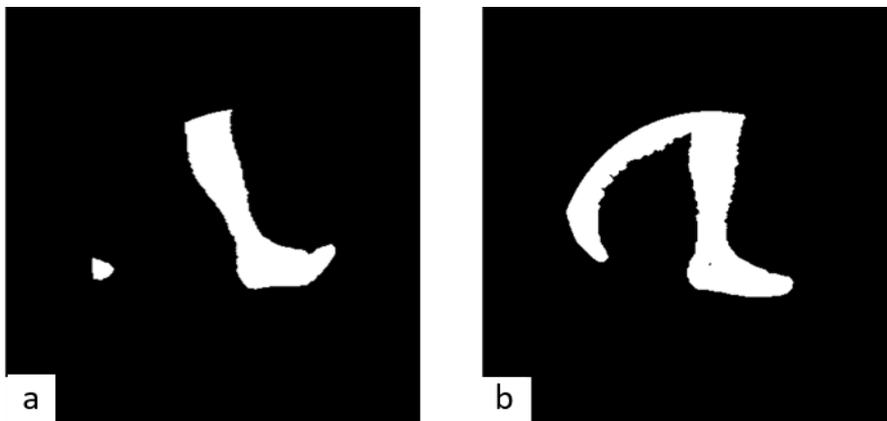


Figure 4.33: Region of the segmented image inside the semicircle with center in the LM and radius the length of the tibia. In (a) there is only the foreground shank because the smallest region is the contralateral foot. In (b) the two shanks are overlapped.

These two conditions are identified by counting how many areas there are in the ROI. In particular:

- 1) If the number of areas is equal to 2, the two shanks are not overlapped because the biggest area represents the foreground shank while the other one is a small region of the foot as shown in Figure 4.33 - a.
- 2) If the number of areas is equal to 1, the two shanks are overlapped, as shown in Figure 4.33 - b.

In the first situation (1), the small region is removed, and the remaining part is the foreground shank (Figure 4.34).

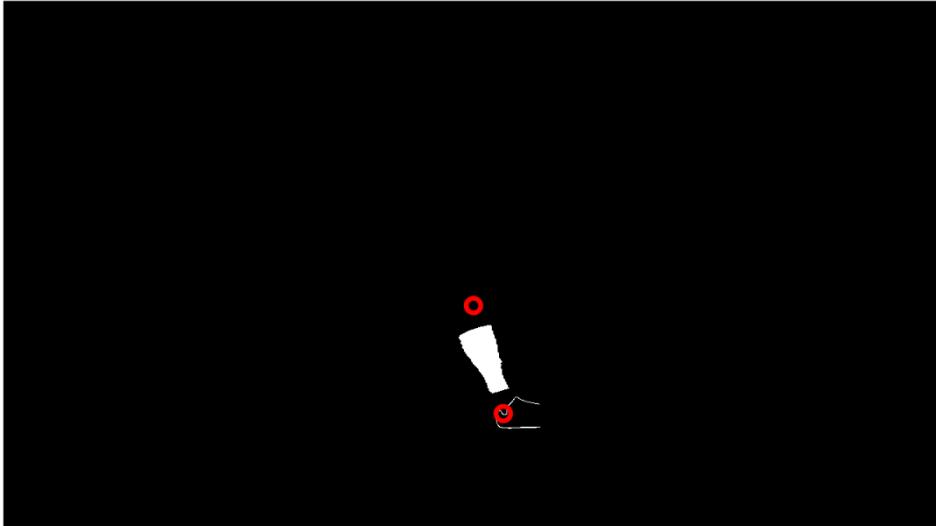


Figure 4.34: Extraction of foreground shank. The red dots are the LE and LM and they identify the knee and the ankle joint, respectively.

The second situation (2) is more critical, because the shanks are partially or entirely overlapped. For this reason, the histogram of the depth values in the considered region is computed to choose a threshold to split the two shanks. In fact, it is possible to isolate the foreground shank by considering two constrains:

- The foreground shank is represented by the lowest depth values in the depth image because it is closer to the depth camera;
- The foreground shank has the highest area in the depth image (i.e. the highest frequency) because the contralateral shank is covered by the foreground one.

The envelope of the histogram is calculated by implementing a moving average filter. To satisfy these two constraints, ideally the position of the shank in the depth image would be represented by a depth value corresponding to the highest peak of the histogram (vertical dashed red line in Figure 4.35)

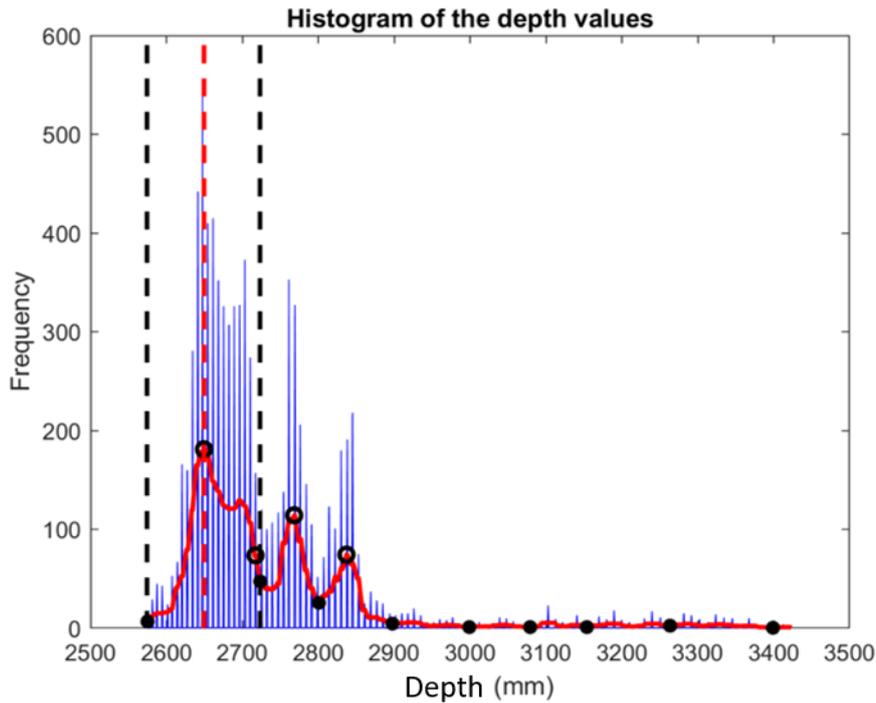


Figure 4.35: Histogram of the depth values. The red line represents the envelope of the histogram obtained by implementing a moving average filter. The red dashed line is the location of the foreground shank. The black dashed lines are the inferior and the superior limit of the depth values range, respectively.

However, the shank being a three-dimensional body it is not represented by a single value but by a range of values around the highest peak of the histogram previously identified. By analysing the trend of the depth histogram, it is possible to identify the limits of this range which represents the foreground shank in the depth image, (vertical dashed black lines in Figure 4.35). When the values outside this range are removed, the foreground shank is isolated.

To select the inferior limit of the depth values range (first black circle on the left of Figure 4.35), the first constraint is considered because it is identified as the lowest value of the depth image since it is the closest point to the camera.

Then, the maxima and the minima of the envelope are computed.

To select the superior limit of the depth values range, a suitable threshold Sh is defined by considering that the area of the foreground shank cannot be smaller than an anatomical value. Moreover, the distance between the two shanks cannot be smaller than a threshold value of distance which includes the step width and the extension of each shank.

The location of the maxima corresponding to the deeper maxima represent candidate positions of the contralateral shank, while the depth values corresponding to the minima are candidate threshold to split the two shanks.

The implementing criterion to achieve this aim is explained in Figure 4.36.

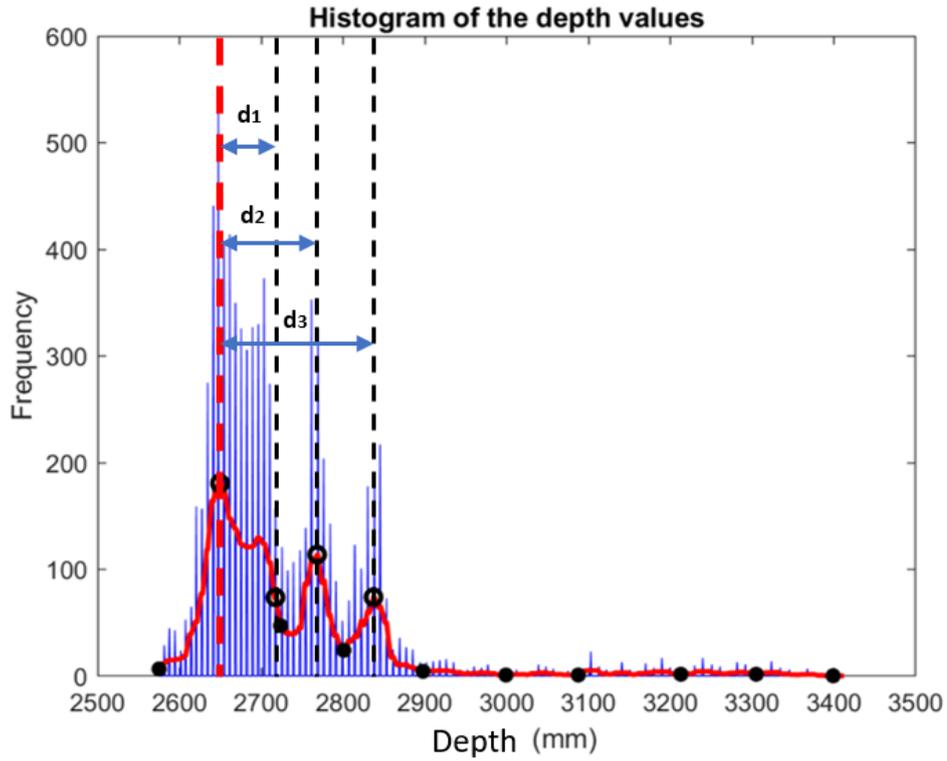


Figure 4.36: The implementing criterion to split the two shanks

As shown in Figure 4.36, the distances between the peak representing the location of the foreground shank and all the other candidate peaks are calculated (4.15). In this case the number of the candidate peaks is equal to three, but this algorithm can be extended to an indefinite number N of peaks. The difference d between the location of the shank and other peaks is computed as:

$$d_n = \text{candidate peak}_n - \text{foreground shank location} \quad (4.15)$$

In particular, if d_n is higher than Sh , the corresponding n -th peak belongs to the contralateral shank. As shown in Figure 4.36, d_2 and d_3 are higher than the threshold Sh . The upper limit is identified in the first minimum before the peak representing the position of the contralateral shank. The depth values outside the defined range are removed and the foreground shank is extracted. Subsequently, the middle portion of the shank segment (between the 25% to 75% of the segment length, defined as the distance between LM and LE) including both LM and LE positions, as shown in Figure 4.37, is extracted.



Figure 4.37: Extracted foreground shank. The blue surface represents the shank, the blue line represents the foot, the green line represents the LM-LE segment. The lowest empty red dot is the LM and the highest empty red dot is the LE.

In addition to the above-mentioned criteria chosen to set the threshold Sh , an additional consideration must be introduced. In fact, Sh is set by considering also the relative distance between the shank region and the position of the camera. In fact, due to the lens distortion, when the subject is at the right or left limits of the camera field of view, the two shanks appear closer than they appear in a central frame in which the parallax is absent. For this reason, Sh is adjusted depending on the relative distance and, in particular, Sh decreases if the subject is far from the camera and it increases if the subject is closer to the camera.

4.3.2.3 Thigh identification

Similar to the processing described for the shank identification, to identify and to extract the foreground thigh, the depth image is considered.

Firstly, the segmented image is overlapped on the depth image to isolate the subject and to remove the other parts of the depth image. In other words, the upper body is removed, and only the portion of the mask below the coordinate of the GT is considered to be analysed subsequently.

A semicircle with center in the LE and radius equal to the length of the thigh, defined as Euclidean distance between the LE and the GT manually identified during the calibration phase, is traced to remove the body regions outside this circle.

After this step, there may be two situations:

1. in the semicircle obtained there is also the foreground hand (Figure 4.38)
2. in the semicircle obtained there is not the foreground hand (Figure 4.39)

These two conditions are identified by analysing the histogram of the depth values. The histogram of the depth values in the considered region is computed in order to choose a threshold which splits the two thighs. The envelope of the histogram is calculated by implementing a moving average filter. Then, the maxima and the minima of the envelope are computed.

These two conditions (with the hand and without the hand) are identified by considering the height of the first peak of the histogram, by setting an experimental threshold which in this case is equal to 80% of the maximum peak of the histogram envelope.

In particular:

1. If the first peak is lower than 80% of the maximum peak of the histogram envelope, then this peak is assumed to represent the hand and the maximum peak represents the position of the foreground thigh;
2. If the first peak represents the maximum peak of the histogram envelope, this represents the position of the foreground thigh and the foreground hand is not in the semicircle;

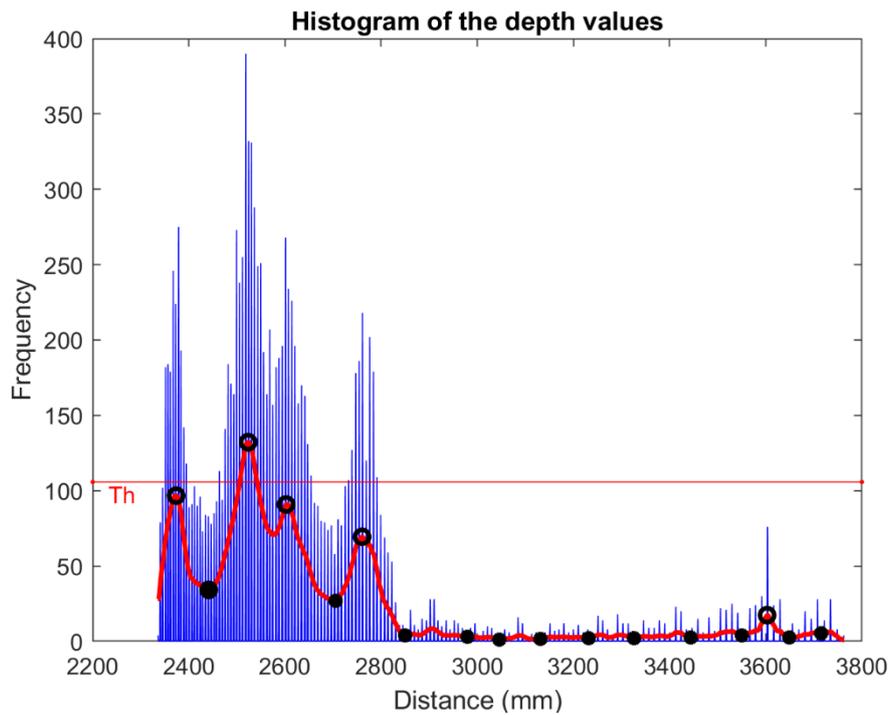


Figure 4.38: Histogram of the depth values. The red line is the envelope of the histogram. The horizontal red line is the 80% of maximum of the envelope. The first peak represents the position of the hand because lower than the threshold.

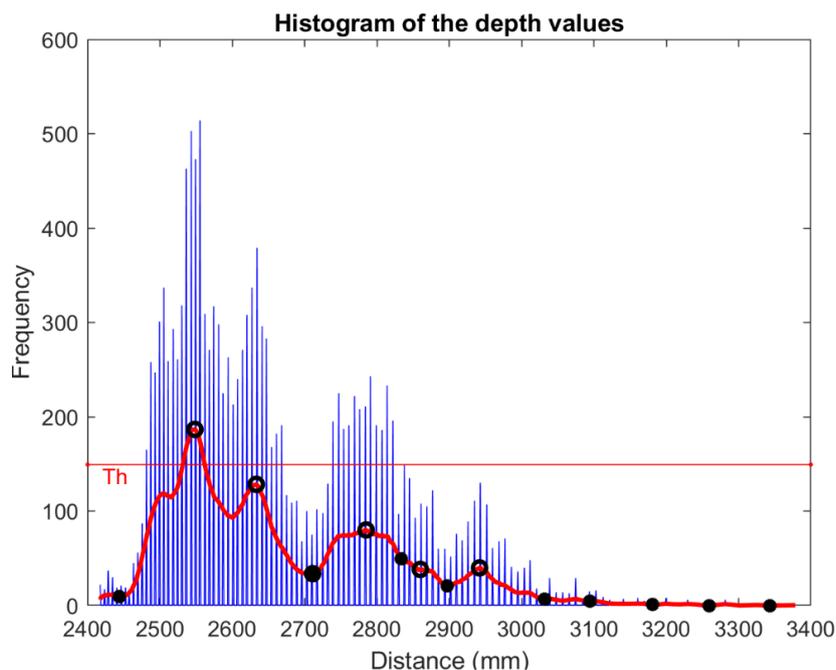


Figure 4.39: Histogram of the depth values. The red line is the envelope of the histogram. The horizontal red line is the 80% of maximum of the envelope. The first peak represents the position of the foreground thigh because is higher than the threshold.

Ideally the position of the thigh in the depth image would be represented by a depth value corresponding to the highest peak of the histogram. However, the thigh being a three-dimensional body it is not represented by a single value but by a range of values around the highest peak of the histogram previously identified. By analysing the trend of the depth histogram, it is possible to identify the limits of this range which represents the foreground thigh in the depth image. When the values outside this range are removed, the foreground thigh is isolated. The inferior limit of the depth values range is:

- Case with hand: The minimum depth value between the depth values of the foreground thigh and the hand;
- Case without hand: The lowest depth value in the histogram.

To select the superior limit of the depth values range, a suitable threshold Th is defined by considering that the area of the foreground thigh cannot be smaller than an anatomical value. Moreover, the distance between the two thighs cannot be smaller than a threshold value of distance which includes the step width and the extension of each thigh.

The location of the maxima corresponding to the deeper maxima represent candidate positions of the contralateral thigh, while the depth values corresponding to the minima are candidate threshold to split the two thighs.

The implementing criterion to achieve this aim is explained in Figure 4.40.

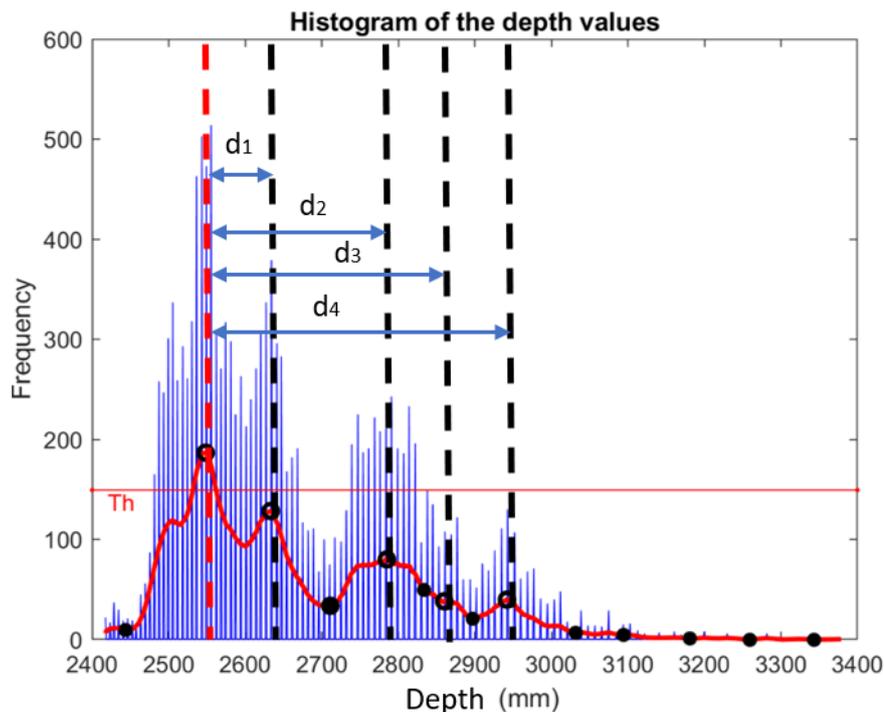


Figure 4.40: The implementing criterion to split the two thighs

As shown in Figure 4.40, the distances between the peak representing the location of the foreground shank and all the other candidate peaks are calculated (4.15). In this case the number of the candidate peaks is equal to four, but this algorithm can be extended to an indefinite number N of peaks.

The difference d between the location of the thigh and other peaks is computed as:

$$d_n = \text{candidate peak}_n - \text{foreground thigh location} \quad (4.16)$$

In particular, if d_n is higher than Th , the corresponding n -th peak belongs to the contralateral thigh. As shown in Figure 4.40, d_2 , d_3 , d_4 are higher than the threshold Th . The upper limit is identified in the first minimum before the peak representing the position of the thigh. The depth values outside the defined range are removed and the foreground thigh is extracted. The identified depth range in all the two cases (case with hand and case without hand) are shown in Figure 4.41.

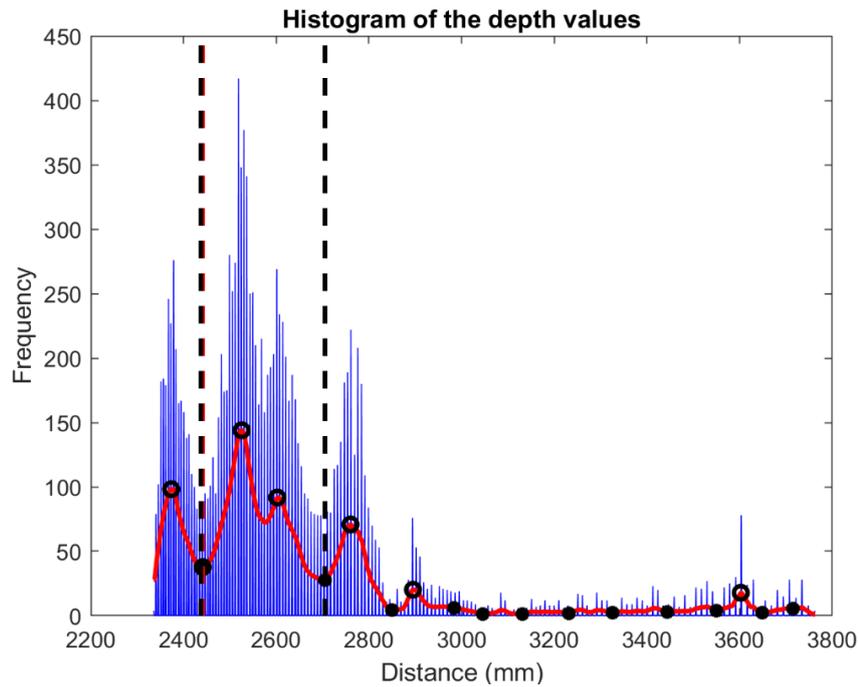


Figure 4.41: Case with hand. The red dashed line is the position of the foreground thigh. The black dashed lines are the inferior and the superior limit of the range of depth values of the foreground thigh, respectively.

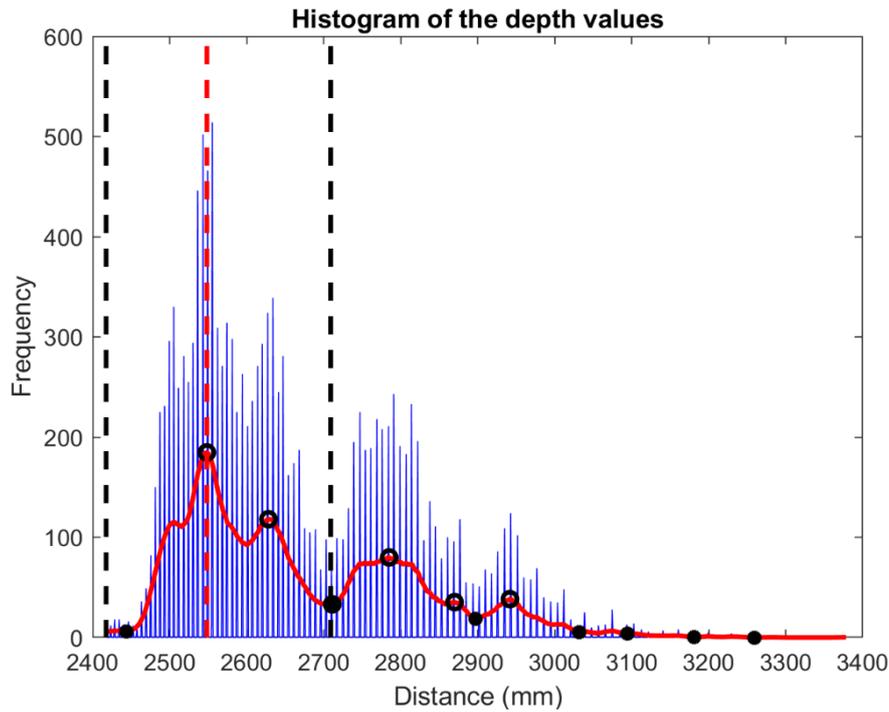


Figure 4.42: Case without the hand. The red dashed line is the position of the foreground thigh. The black dashed lines are the inferior and the superior limit of the range of the depth values of the foreground thigh, respectively.

The threshold Th is set by considering also the relative distance from the camera and the position with respect to it.

In fact, due to the lens distortion, when the subject is at the right or left end of the camera field of view (FOV), the two thighs appear closer than they appear in a central frame in which the distortion is absent. For this reason, the threshold Th is not the same for each frame of the selected gait cycle, but it depends on the position of the subject. In particular, Th decreases if the subject is far from the camera and it increases if the subject is closer to the camera. In conclusion the threshold Sh adapts to the position of the subject with respect to the camera.

Subsequently, the middle portion of the thigh segment (between the 25% to 75% of the segment length, defined as the distance between LE and GT) including both LE and GT positions, as shown in Figure 4.44 and Figure 4.43.

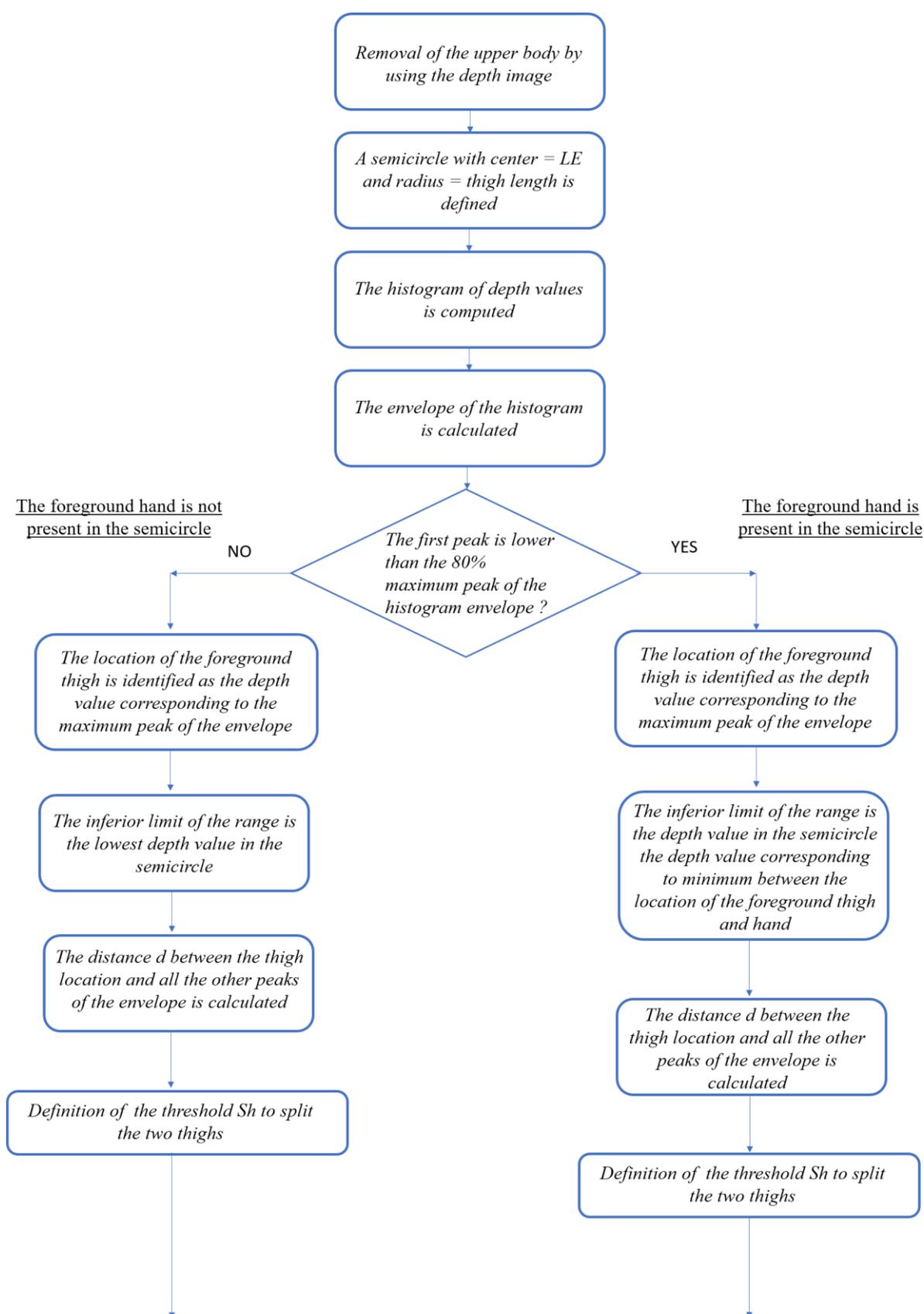
The block diagram of the algorithm is shown in Figure 4.45.



Figure 4.44: Extracted foreground thigh in situation 1. The blue surfaces represent the shank and the thigh, the blue line represents the foot, the green lines represent the LM-LE segment and the LE-GT segment. The first empty red dot is the LM the second empty red dot is the LE, and the higher empty red dot is the GT.



Figure 4.43: Extracted foreground thigh in situation 2. The blue surfaces represent the shank and the thigh, the blue line represents the foot, the green lines represent the LM-LE segment and the LE-GT segment. The first empty red dot is the LM the second empty red dot is the LE, and the higher empty red dot is the GT.



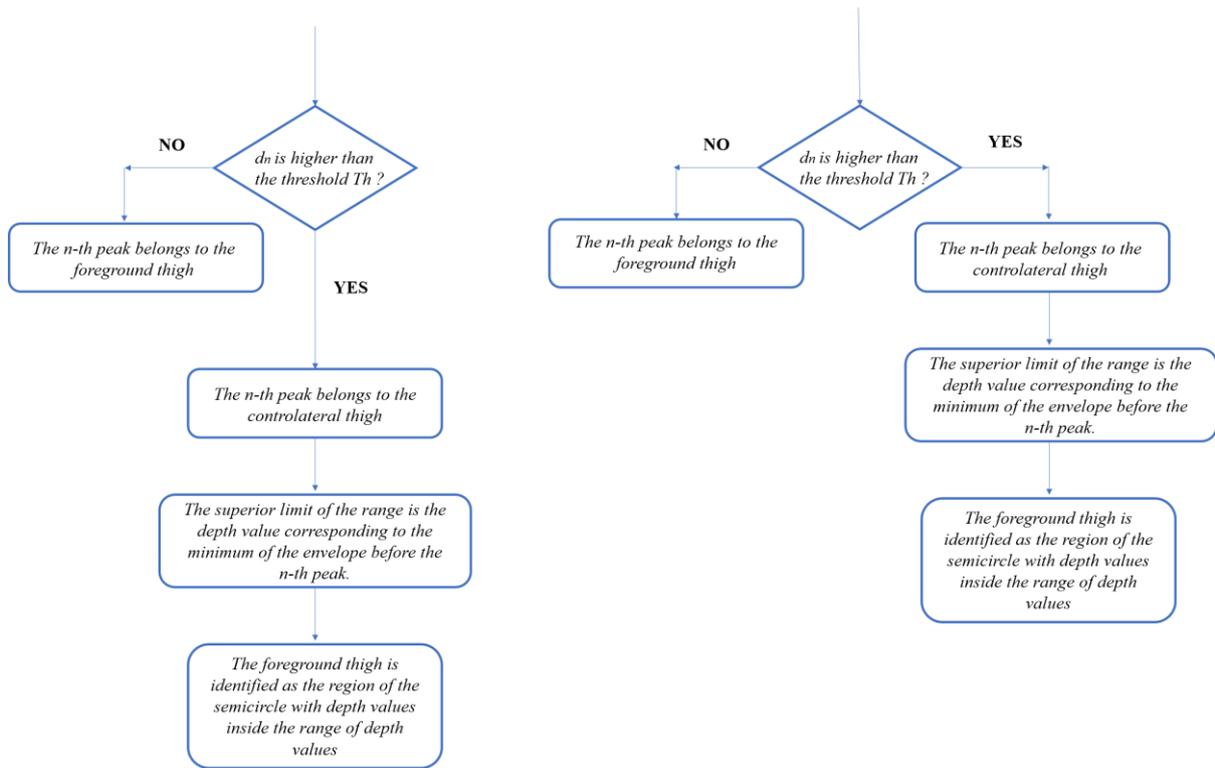


Figure 4.45: Block diagram of 'Foreground thigh identification'

4.3.3 Joint center tracking trajectories during the gait cycle though the ICP algorithm

The multi-segmental model was tracked using a bottom-up tracking approach, starting from the foot and moving up to the pelvis. For each frame, the position of the LM was reconstructed by minimizing the distance between the point cloud of the portion of interest of the segmented foot and the foot model as shown in Figure 4.46, using an iterative closest point (ICP) technique (described in 4.3.3.1) which included scaling.

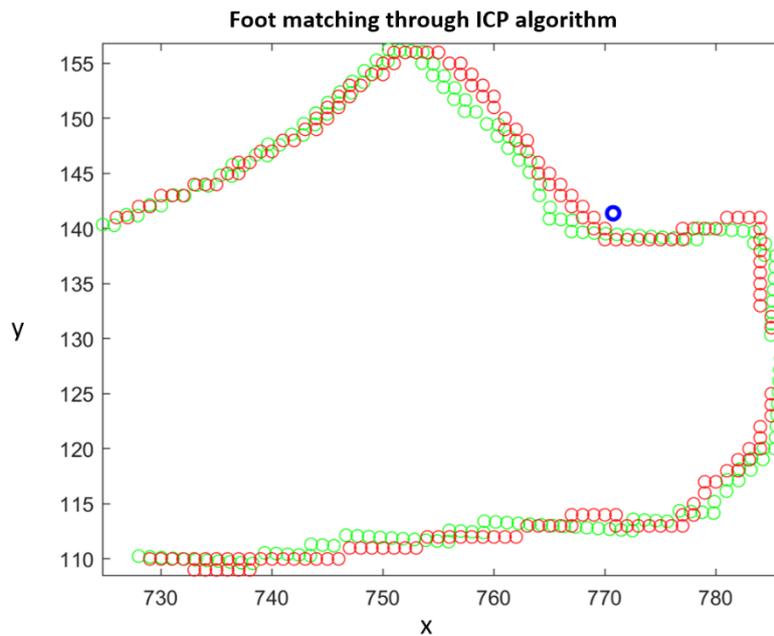


Figure 4.46: Foot matching: the red dots represent the coordinates of the foot model and the green dots represent the coordinates of the segmented foot to align. The blue dot represents the LM.

To reconstruct the position of the LE, first the portion of interest of the shank was extracted from the segmented image, exploiting the position of the LM, the ankle joint kinematic constraint, and information relative to the shank length; then we minimized the distance between the point cloud of the segmented shank and the shank model using the ICP technique as shown in Figure 4.48.

A similar procedure was employed to reconstruct the position of the GT by exploiting the position of the LE, the knee joint angle constraint, and information relative to the thigh length and implementing an ICP technique. In order to manage with legs overlapping during the gait cycles, when necessary we used both RGB and depth data to isolate the foreground and background shank and thigh portions of interest as shown in Figure 4.47.

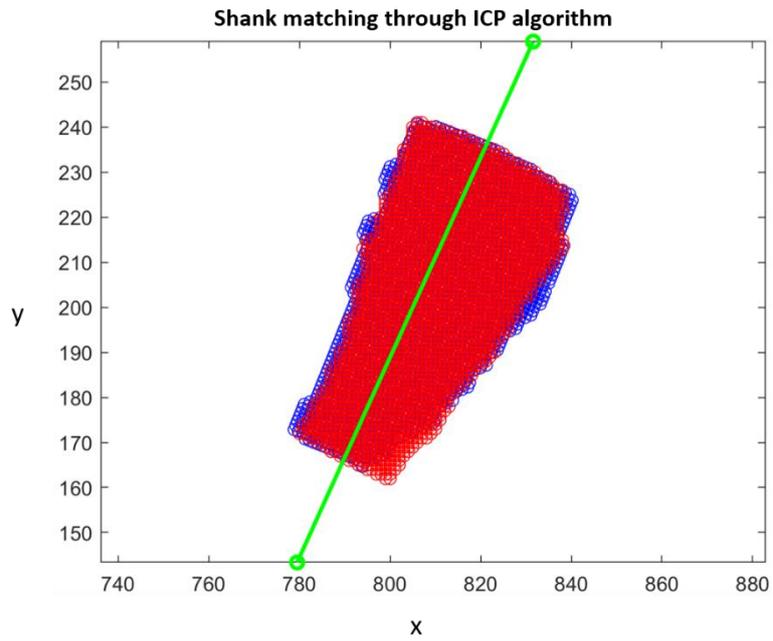


Figure 4.48: Shank matching: the blue dots represent the shank model and the red dots represent the segmented shank to align. The green line represents the segment LM-LE.

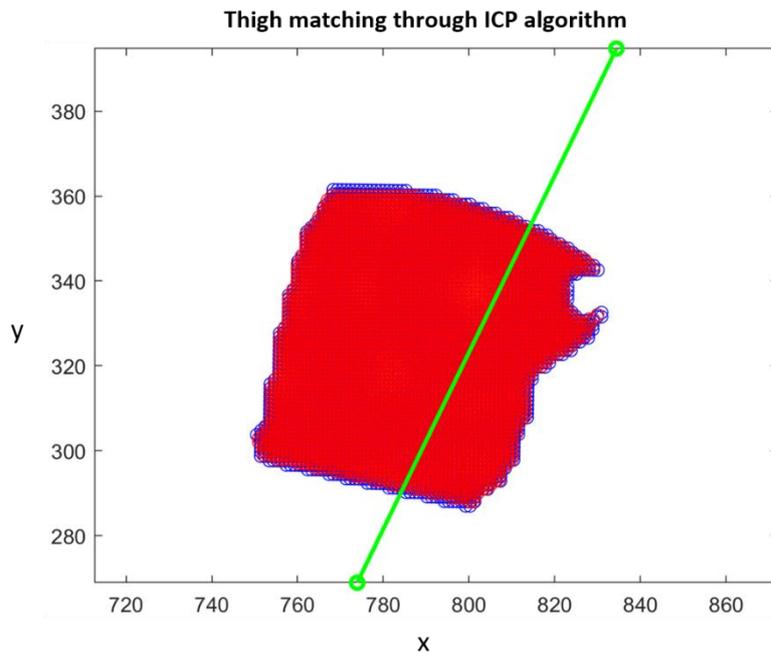


Figure 4.47: Thigh matching: the blue dots represent the thigh model and the red dots represent the segmented thigh to align. The green line represents the segment LE-GT.

4.3.3.1 Iterative closest point (ICP) algorithm

The Iterative closest point (ICP) algorithm is an independent method for the computationally efficient and accurate registration of three- dimensional and two-dimensional shapes including free-form curves and surfaces.

This method presents full six-degrees of freedom and the procedure aims to find the closest point on a geometric entity to a given point. The ICP algorithm always converges monotonically to the nearest local minimum of a mean-square distance metric.

In particular, the model shape is X and the data shape which is moved in order to reach the best alignment with the model is P .

The ICP algorithm presents the following steps:

- Compute the closest points: the distance metric d between a data point p and a model shape X is computed as:

$$d(p, X) = \min_{x \in X} \|x - p\| \quad (4.17)$$

y is the closest point in X that yields the minimum distance $d(p, y) = d(p, X)$ where $y \in X$. The closest point operator is C in order to denote Y as the resulting set of closest points:

$$Y = C(P, X) \quad (4.18)$$

- Computer the registration: the least squares registration is computed by minimizing the objective function:

$$f(q) = \frac{1}{N_p} \sum_{i=1}^{N_p} \|x_i - \mathbf{R}(q_R)p_i - q_T\|^2 \quad (4.19)$$

where:

- N_p is the number of point belonging to X ;
- q_R is the rotation vector;
- q_T is the translation vector.

The least squares registration is computed as described above:

$$(q, d) = Q(P, Y) \quad (4.20)$$

where Q is the registration operator and q is the complete registration state vector. Then, the position of data shape point are update via $P = q(P)$.

After obtaining P, the iterative process starts by initializing $P_0 = P$, $q_0 = [1 \ 0 \ 0 \ 0 \ 0 \ 0]^t$ and $k=0$.

1. Computation of the closest points:

$$Y_k = C(P_k, X) \quad (4.21)$$

2. Computation of the registration:

$$(q_k, d_k) = Q(P_0, Y_k) \quad (4.22)$$

3. Application of the registration:

$$P_{k+1} = q_k(P_0) \quad (4.23)$$

4. Computation of the least mean square distance:

$$d_k = \frac{1}{N_p} \sum_{i=1}^{N_p} \|y_{i k} - \mathbf{R}(q_k R)p_{i 0} - q_k T\|^2 \quad (4.24)$$

5. Terminate the iteration when the mean square error is below a present threshold τ :

$$d_k - d_{k+1} < \tau \quad (4.25)$$

4.3.4 Computation of lower limb joint angles

Once the trajectories of the LM, LE, and GT were reconstructed during the entire gait cycle:

- The ankle joint angle could be easily determined as the angle between the foot and the LM-LE segment
- The knee joint angle could be easily determined as the angle between the LM-LE and LE-GT segments.
- The hip joint angle could be easily determined as the angle between the LE-GT and the y-axis.

The above-mentioned procedure was implemented for each of the three participant-specific lower limb models, thus producing three different knee angles. These curves were then combined into a single curve by linearly weighting each contribution according to the percentage of the gait cycle and the knee joint angle value. This operation reduced the errors associated with unavoidable changes in shape and size of the foot, shank, and thigh models in the image frames recorded during the gait cycles (Cappello *et al.*, 1997 [26], Cereatti *et al.*, 2015 [27]).

5 Experimental acquisitions

The purpose of the experiments described in this chapter is to investigate the accuracy of the proposed 2D markerless gait analysis protocol for estimating the sagittal lower limb joint kinematics in different experimental set up conditions i.e. with and without a green background and with two or one additional leds to improve the quality of light conditions.

Before conducting the experimental session, a preliminary investigation was conducted to select the hardware, between Kinect II and Intel RealSense D435, which presents the highest performances in terms of resolution and sharpness of the image.

After this step, all the gait analysis trials are acquired with both the selected RGB-D camera and the stereophotogrammetry (SP) system which represents the gold standard for the estimation of sagittal lower limb kinematics. The joint angles obtained by RGB-D camera are compared to those of the SP system (Vicon system).

5.1 Preliminary investigations: hardware selection

The hardware selection is a fundamental step of the project thesis because the performances of the camera influences the quality of the results. In [2] the Kinect II is used to acquire the video for the gait analysis. Nowadays, the Kinect II is out of production, so this hardware may be replaced by another hardware with similar characteristics and performances. For this reason, the Intel RealSense Depth camera is analysed in order to decide if it could replace the Kinect II.

In this subchapter 5.1.1, the description of the two hardwares is conducted and subsequently a comparison between the two cameras in order to choose the best camera.

5.1.1 Hardware description

5.1.1.1 Kinect II

The Kinect II (Figure 5.1: Kinect 2) sensor is an RGB-D camera released by Microsoft. This sensor is a laser scanner which integrates depth measurements with traditional two-dimensional images [28].



Figure 5.1: Kinect 2

This RGB-D camera is composed by an RGB camera (CMOS image sensor), an IR sensor and an IR illuminator as shown in Figure 5.2.



Figure 5.2: Sensors belonging Kinect 2 camera: RGB camera, IR sensor and IR illuminator

The features of Kinect II are shown in Table 1.

Table 1: Characteristics of Kinect II

Kinect II	Features
Field of view (H×V)	70° × 60°
Camera resolution (H ×V)	1920 × 1080 @ 30 fps
Depth resolution (H ×V)	512 × 424
Maximum depth range	4.5 m
Minimum depth range	50 cm
Depth technology	Indirect time of flight

This sensor uses a novel image system with indirect time of flight (ToF) that replaces the structure light pattern of Kinect 1.

The ToF system uses a square wave to modulate the camera light source. The phase detection is used to calculate the time it takes light to travel from the IR light source to the object and back to the sensor and calculates distance from the results as shown in Figure 5.3 .

The timing generator produces a modulation square wave. The system uses this signal to modulate both the local light source (transmitter) and the pixel (receiver). The light travels to reach the object and back in time Δt [29].

“The system calculates Δt by estimating the received light phase at each pixel with knowledge of the modulation frequency. The system calculates depth from the speed of light in air: 1 cm in 33 picoseconds”.

In particular, Kinect II uses multiple measurements performed by three pulse frequencies and three amplitudes to obtain for each pixel:

- The amount of reflected light;
- The depth of scene from the phases shifts for the multiple measurement;
- The amount of ambient light.

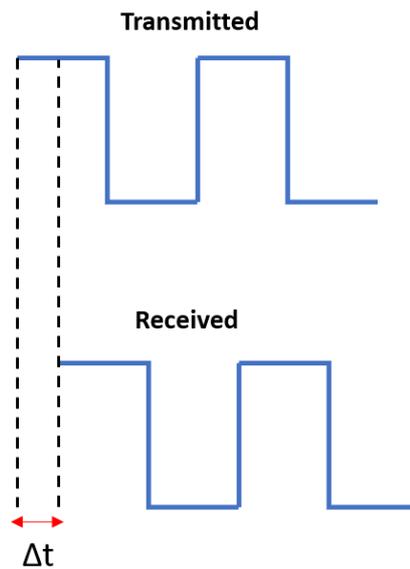


Figure 5.3: Time of flight (ToF)

RGB and IR images acquired with the Kinect II are partially overlapped, since the RGB camera has a wider horizontal Field of View (FOV), while the new IR camera has a larger vertical FOV as shown in Figure 5.4.

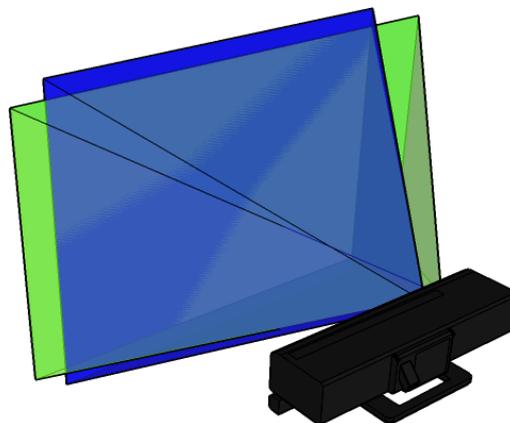


Figure 5.4: Kinect II RGB e IR camera FOV. In blue is the FOV of the IR camera and in green is FOV of the RGB camera

The first step of the data processing consists in correcting the optical distortion of both the RGB and depth images captured by the Kinect.

A stereo calibration is performed to create new depth images that can be overlapped [30] with the RGB images, so that the depth information can be associated to each point in the RGB image (i.e. depth information can be used to identify different part of the body in the RGB images, such as the forward and the backward leg).

Both lens distortion and stereo calibration are carried out by capturing several calibration images of a chessboard, recorded with both the Kinect sensors (RGB and IR/depth) as shown in Figure 5.5

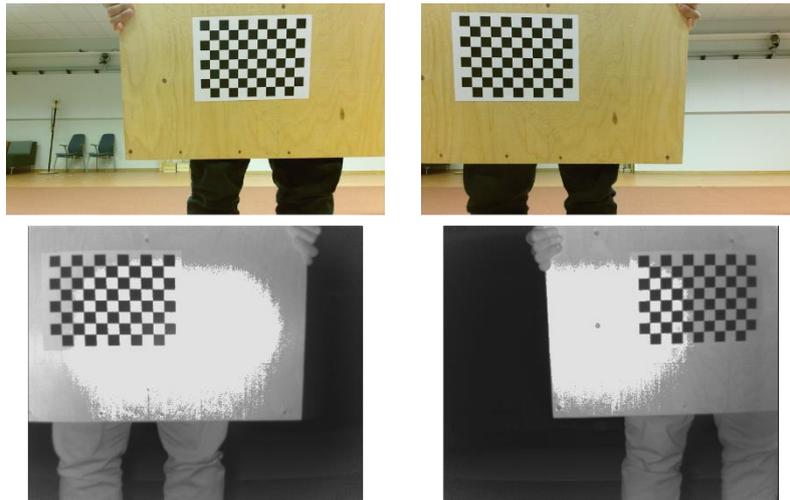


Figure 5.5: several calibration images of a chessboard recorded with both the Kinect sensors

By using RGB and IR images (10-20 images) containing the chessboard positioned in different position within the shared FOV, the calibration of Kinect II is performed, and it provides:

- Elimination of the mirror effect introduced by the Kinect II;
- Extraction of intrinsic parameters i.e. focal length, main point) of the two sensors (RGB camera and depth sensor);
- Image rectification;
- Identification of a common ROI, as shown in, through the cropping of RGB images (1920×1080 becomes 1536*1080) in width and of IR one in height (512×424 becomes 512×360). The RGB images are three times larger the IR images.



Figure 5.6: Identification of a common ROI in both RGB and depth images.

- Scaling of a factor equal to $\frac{1}{3}$ of RGB images to obtain an RGB image with a size of 512×360;
- Extraction of intrinsic parameters (focal length, main point) and extrinsic (rotation and translation matrices) related to modified images.

Once the system calibration is completed, the mapping of the RGB / depth pair of a frame is carried out using the following steps:

- Projection of the pixel coordinates (Figure 5.7) of the depth image (px_d, py_d) in the world reference system centred in the depth sensor (wx_d, wy_d, wz_d) using the intrinsic parameters obtained through the system calibration of the depth sensor;

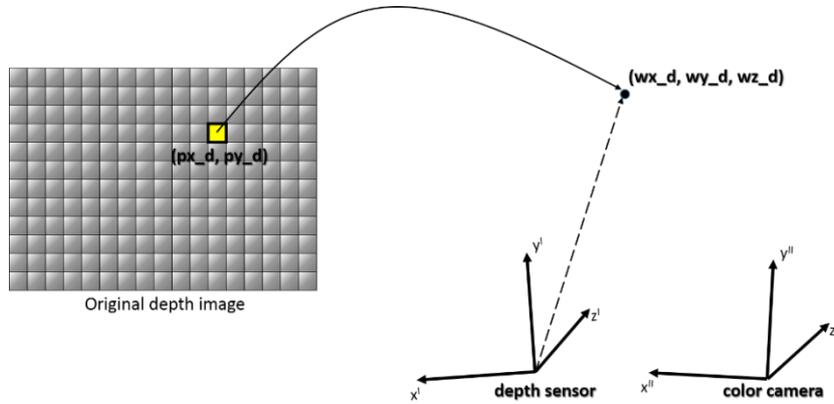


Figure 5.7: Projection of the pixel coordinates of the depth image

- Transformation of the world coordinates in the reference system centred in the depth sensor (wx_d, wy_d, wz_d) in world coordinates in the reference system centred in the video camera (wx_c, wy_c, wz_c), using the rototranslation matrix (Figure 5.8);

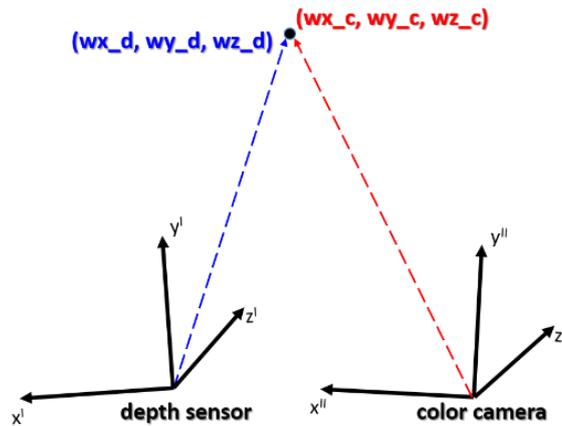


Figure 5.8: Transformation of the world coordinates in the reference system centered in the depth sensor in world coordinates in the reference system centered in the video camera

- Back-projection (Figure 5.9) of world coordinates in the reference system centred in the RGB sensor (wx_c, wy_c, wz_c) on the RGB image plane (px_c, py_c) using the intrinsic parameters obtained through the system calibration of the RGB sensor.

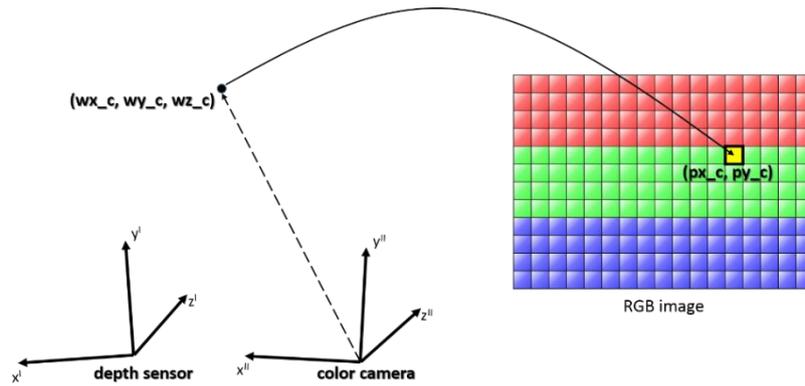


Figure 5.9: Back-projection of world coordinates in the reference system centred in the RGB sensor on the RGB image plane

After this procedure the correspondence between the pixels of the initial depth image (px_d, py_d) and the pixels of the RGB image (px_c, py_c) is obtained, to which it is therefore possible to associate one's own depth value.

Starting from the depth data in the original depth image, following the correspondence just obtained, a new depth image ($512 * 360$) is created which will represent the depth map of the RGB image.

In fact, each pixel of the new image will contain the depth value of the corresponding pixel in the RGB image as shown in Figure 5.10.

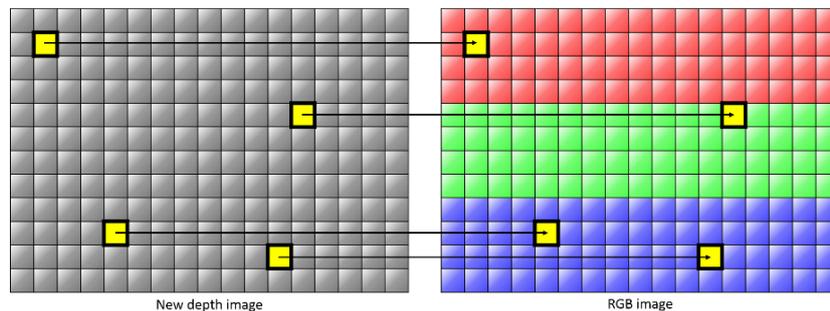


Figure 5.10: mapping of RGB and depth images

5.1.1.2 Intel real sense depth camera D435

The D435 is a USB-powered depth camera and consists of a pair of depth sensors, RGB sensor and infrared projector as shown in Figure 5.11.



Figure 5.11: Intel Real Sense Depth camera D435

The depth perception based on stereo vision is implemented by two image sensors and an infrared projector. The infrared projector projects non-visible structured IR pattern to improve depth accuracy in scenes in low light conditions [31] as shown in Figure 5.12.

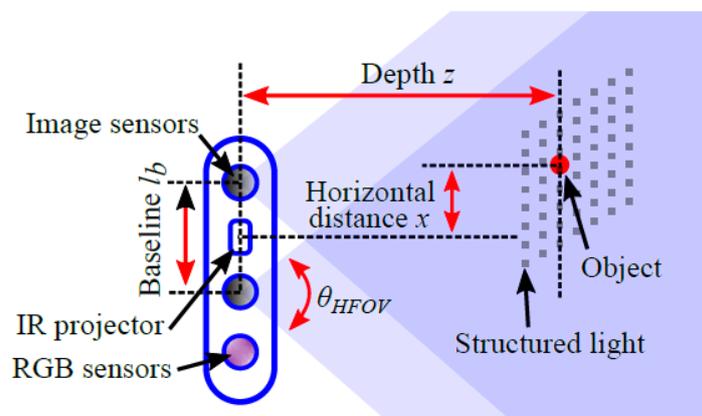


Figure 5.12: Depth perception based on active infrared stereo vision technology

To understand how this camera works, it is important to introduce the terminology used for this hardware:

- Imagers: Depth camera system uses a pair of cameras referred as imagers to calculate depth. They are identical cameras with identical configuration settings;
- Stereo Depth Baseline: the distance between the center of left and right imagers in a stereo camera;
- Field of View (FOV): angular extent of a given scene that is imaged by a camera. It can be measured horizontally, vertically and diagonally ()
- IR projector: source of infrared (IR) light used to illuminate a scene, object or person to collect depth data (in low light condition);

The specifications of the Intel RealSense D435 are shown in Table 2 [32].

Table 2: Specification of Intel RealSense D435

Intel RealSense D435	Features
Depth Field of view (H×V)	85.2° × 58°
RGB sensor Field of view (H×V)	69.4° × 42.5°
RGB sensor resolution (H ×V)	Up to 1920 × 1080 @ 30 fps
Depth resolution (H ×V)	Up to 1280 × 720 @ 90 fps
Maximum depth range	10 m
Minimum depth range	0.11 m
Depth technology	Active IR stereo

The depth image processor obtains the scene data by the two image sensors, and the depth values for each pixel can be calculated by correlating the points on the left image to the right image. A block-diagram to illustrate the algorithm used to obtain depth measurements is shown in Figure 5.13.

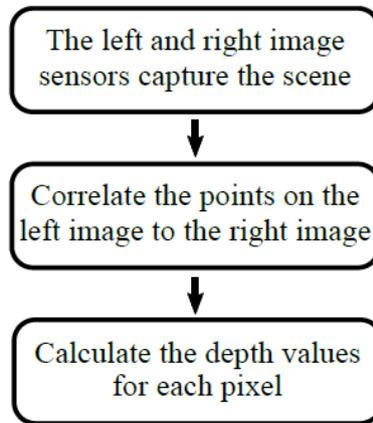


Figure 5.13: Block-diagram to explain the algorithm used to obtain depth perception

The active stereo vision is a vision algorithm inspired by the human binocular vision system. It is based on two parallel view-ports and the depth is calculated by estimating differences between matching key-points in the left and right images (Figure 5.14)

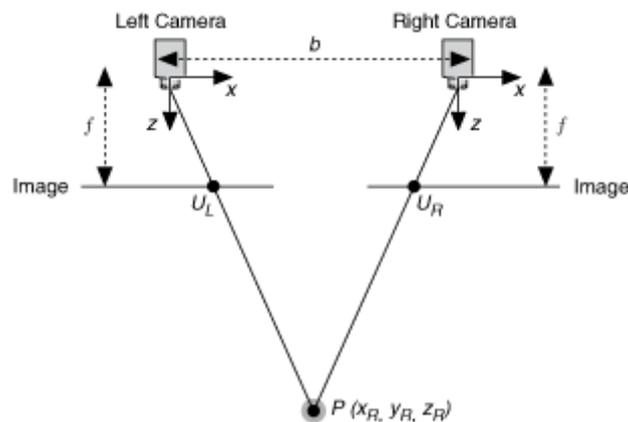


Figure 5.14: Active stereo vision

To obtain this matching a Sum of Squared Differences (SSD) algorithm is implemented as:

$$SD = (Img_1 - Img_2)^2 \quad (5.1)$$

$$SSD = \sum_{i=1}^N \sum_{j=1}^M SD(i, j) \quad (5.2)$$

where:

- Img_1 is the left image;
- Img_2 is the right image;
- N is the number of the rows of the image;
- M is the number of the columns of the image.

This matching between the two images introduces a depth error that can be quantified by calculating the RMS error with this equation:

$$e = \frac{z^2 p_s}{l_f l_b} \quad (5.3)$$

where:

- z is the depth between the camera and the analysed object;
- l_f is the focal length;
- l_b is the baseline of the two sensors;
- p_s is the size of the subpixel.

Another error is called ‘Invalid depth band’. This band is generated because the depth data generated with active stereo vision uses the left imager as the reference for stereo matching resulting in a non-overlap region in the FOV of the left and right imager. For this reason, there will not be depth data at the left edge of the frame.

As shown in Figure 5.15, closer scenes have a wider invalid depth band than scenes at further distances.

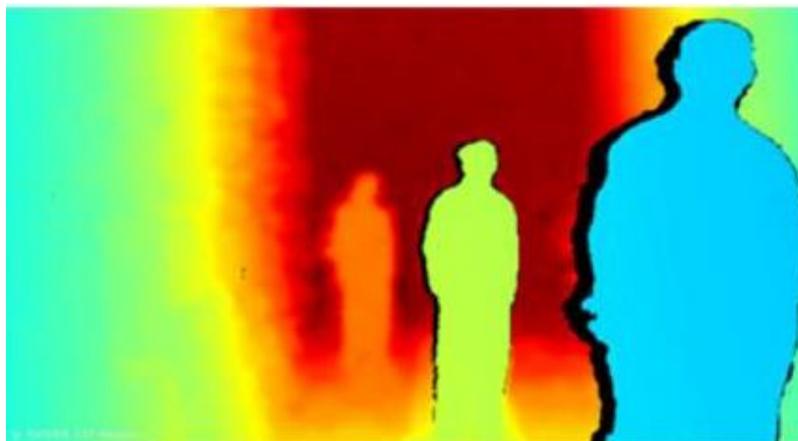


Figure 5.15: Left invalid depth area

The great advantage of this camera is the presence of options that can be modified, i.e. the possibility to adjust and set numerous parameters. For example, regarding the depth camera, it is possible to adjust the exposure, the gain and the power. As regards, instead, the RGB camera it is possible to decide for example the exposure, the brightness, the contrast, the gain, the power, the saturation and the white

balance. Of all these characteristics that are variable by the user, the most interesting variable is the exposure that can be made fixed or variable.

The exposure is the total amount of light that passes through the optical system during a defined time period called exposure time.

This variable is important because an improper exposure carries the possibility to have blurred images.

5.1.2 Hardware selection

All the video acquired by the Kinect II and by the Intel RealSense D435 are analysed and compared in order to select the hardware with higher performances.

In particular, in [2] one of the most important aspect which has a negative impact on the estimation of lower limb kinematics is the presence of blurred images obtained by the use of Kinect II due to an improper variable exposure time which cannot be changed as it is a parameter automatically set by the camera.

By comparing all the video in different experimental set up conditions as shown in Figure 5.16 i.e. two or one additional leds and with/without the green background, it is possible to notice the higher image quality of the images acquired with Intel RealSense D435 because, although the images are darker that the ones acquired by the Kinect II, these are not blurred because the exposure time was set based on the environment in which the experiments were conducted.

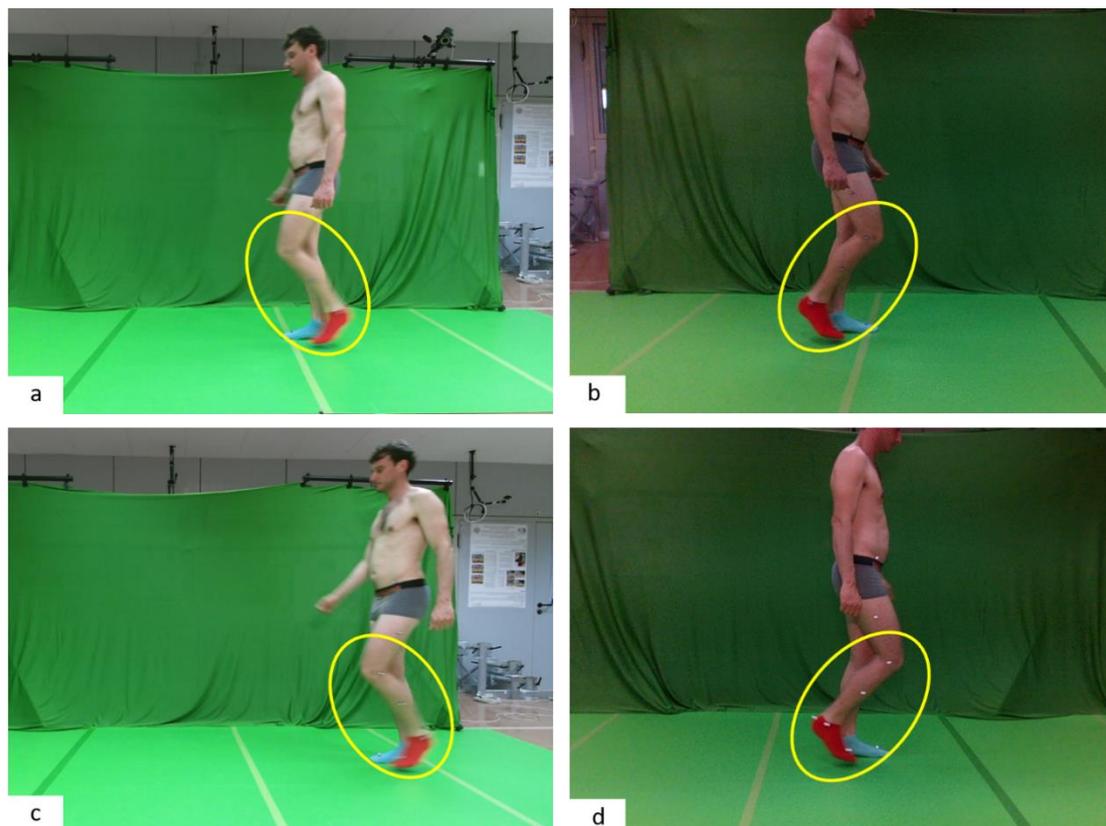


Figure 5.16: Image acquired by Kinect II (on the left side) and by Intel RealSense D435 (on the right side) with the green background and with two additional leds. The yellow lines represent the shank and foot which are the body regions more blurred.

These blurred images cause the wrong identification of shank and foot which leads errors in the estimation of lower limb joint kinematics because the extracted segment with connects the knee and ankle joint has a wrong orientation.

This problem increase during the acquisitions without the green background, as shown in Figure 5.17. For this reason, you choose to elaborate only the images acquired by the Intel RealSense D435.

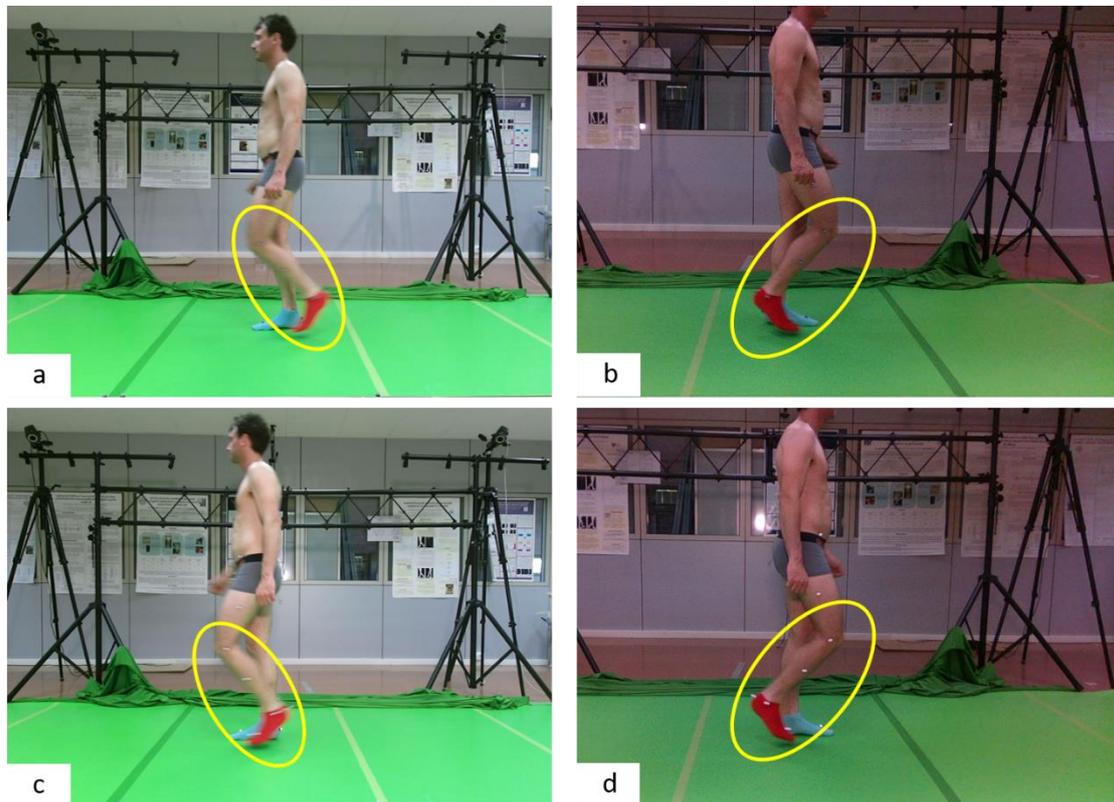


Figure 5.17: Image acquired by Kinect II (on the left side) and by Intel RealSense D435 (on the right side) with the green background and with two additional leds. The yellow lines represent the shank and foot which are the body regions more blurred.

5.2 Experimental set up

The experiments were conducted in one healthy subject. The environment chosen was a gait analysis laboratory because there was a SP system (Vicon system shown in Figure 5.18) composed by:

- Nine infrared cameras with $f_s=120$ Hz;
- One RGB camera with $f_s=120$ Hz.

The RGB-D camera was positioned laterally to the walkway and the image coordinate system of the video camera was aligned to the sagittal plane.

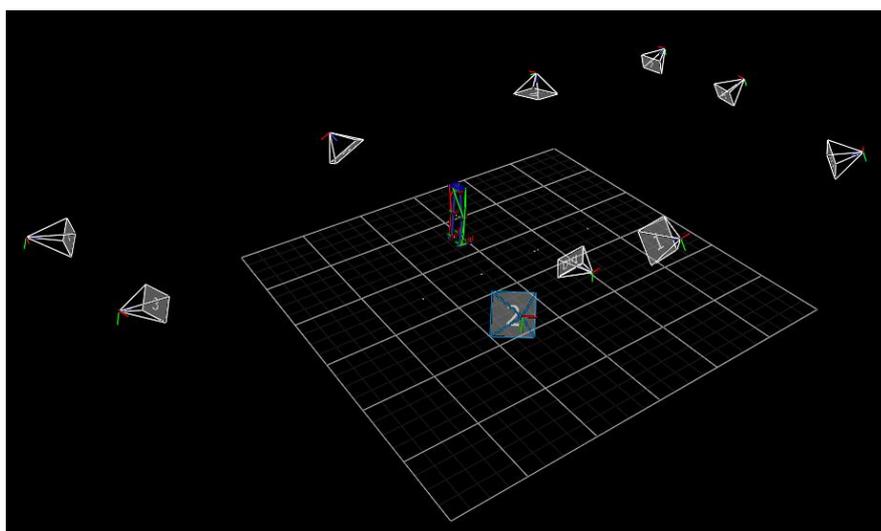


Figure 5.18: 3D perspective view of the gait analysis laboratory with SP system. There are 9 infrared cameras and one RGB camera.

5.2.1 Two different setup conditions: with and without green background

The experiments was conducted in two different setup conditions as shown in Figure 5.19 and Figure 5.20 :

- With a green background and with a green walkway;
- Without a green background and with a green walkway.

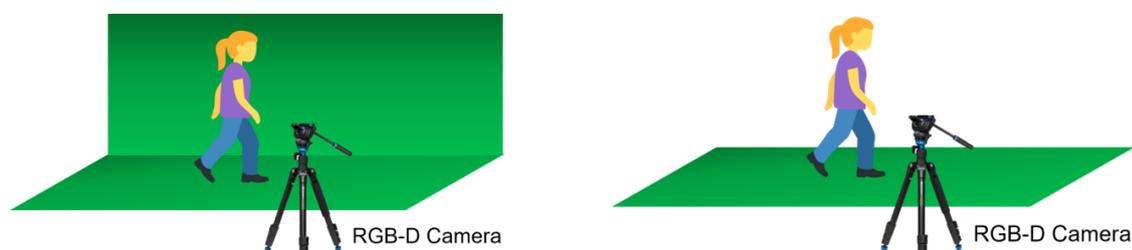


Figure 5.19: Two different experimental setup conditions: with green background (left) and without the green background (on the right). In both situations there is the green carpet.



Figure 5.20: Two different experimental setup conditions: with green background (left) and without the green background (on the right). In both situations there is the green carpet.

The aim is to investigate the accuracy of the proposed 2D markerless gait analysis protocol by comparing the accuracy of the estimated lower limb kinematics in this two setup conditions.

5.2.2 Two different light conditions: two LED lamps vs one LED lamp

The experiments was conducted in two different light conditions as shown in Figure 5.21 and Figure 5.22 :

- One additional LED lamp (40 W);
- Two additional LED lamps (40 W each lamp).



Figure 5.22: Experimental setup with one additional LED lamp: the LED lamp is behind the Intel RealSense D435



Figure 5.21: Experimental setup with two additional LED lamps: the LED lamps are on the left and right of the Intel RealSense D435.

The aim is to investigate the accuracy of the proposed 2D markerless gait analysis protocol by comparing the accuracy of the estimated lower limb kinematics in this two light conditions.

5.3 Subject preparation

The subject wears a T-shirt, underwear and coloured ankle socks (red for right foot and blue for left foot) as shown in Figure 5.23. In order to compare the results of the 2D markerless gait analysis with the gold standard, represented by the SP system, the markers are fixed on the subject to compose the Plug-in gait lower limb model as shown in Figure 5.24.



Figure 5.23: Ankle socks (red for right foot and blue for left foot)

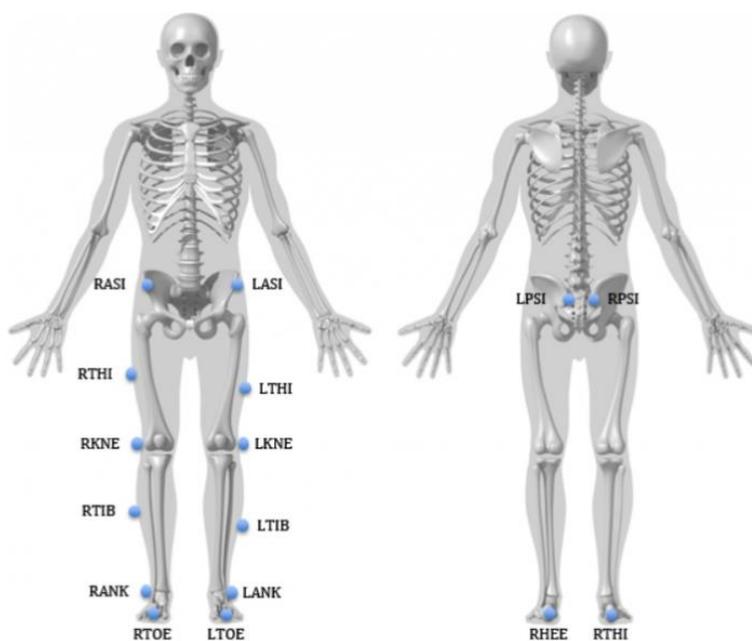


Figure 5.24: Plug-in gait lower limb model

5.4 Calibration of the SP and ML systems

During this phase, the SP system must be calibrated in order to obtain both intrinsic (focal length, lens distortion) and extrinsic (position and orientation) parameters of multiple cameras [33]. Full metric calibration is obtained from an active calibration wand (shown in Figure 5.25) with five active markers at known fixed distance.

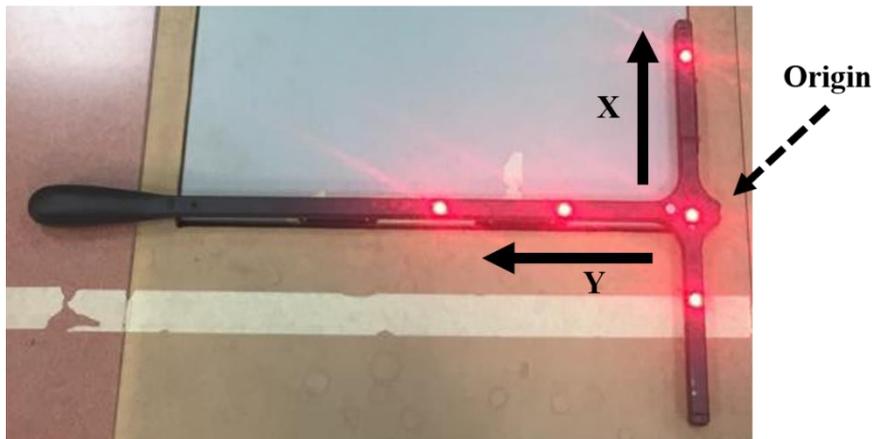


Figure 5.25: Active calibration wand with five active markers.

An expert operator swings the wand around in the FOV of the multiple cameras because all the markers must be seen by all the cameras in order to calibrate each of them. The software of the SP system (Nexus) helps the operator by showing the interface (Figure 5.26) which displays the movements of the calibration wand seen by each camera.

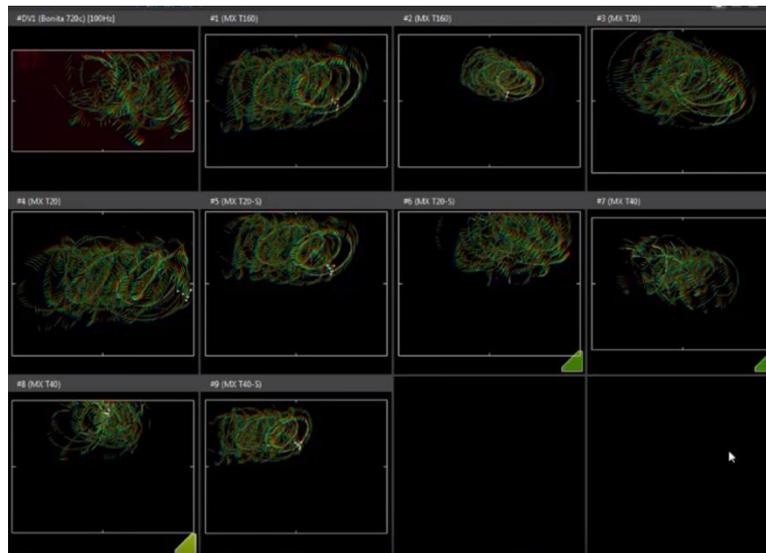


Figure 5.26: Nexus interface: the green lines are the trajectories of the calibration wand seen by each camera: 9 infrared camera and one RGB camera.

After the calibration of the system, the subject (with passive markers on his body) enters in the FOV of the multiple cameras in order to create a stick model.

To create this, all the anthropometric measurements are registered i.e. thigh length, shank length, distance between two malleolus, knee diameter and pelvis width.

The fundamental step to obtain the model is to perform the labelling of the anatomical landmarks by an expert operator. In particular, each marker is associated with its anatomical landmark and the result of this labelling is shown Figure 5.27.

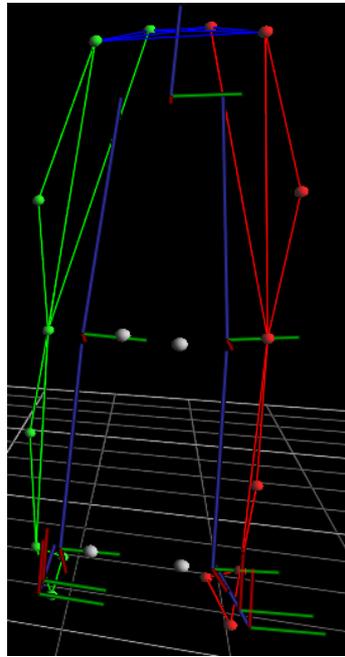


Figure 5.27: Stick model: all the segments (foot, shank, thigh and pelvis) of the body are defined.

After the calibration of SP system, there is a parameter setting phase, in which an expert operator sets the value of the time of exposure, the contrast, the white balance, the brightness in order to obtain a good quality image by avoiding blurred image. It is important to consider all light conditions (two additional LED lamps and one additional LED lamp) to set these parameters.

5.5 Measurement protocol

At the beginning of each trial discussed in this section, the subject was asked to stay in the neutral standing posture for few seconds for both the right and the left side. After the static acquisitions, the subject was asked to walk at a self-selected normal speed in one direction for three times both for the right and the left side. To obtain frames without the subject, each acquisition starts when the subject is not in the FOV of the camera.

The experimental acquisition for each subject were:

	1 LED lamp	2 LED lamps
With background	3 trials (R and L)	3 trials (R and L)
Without background	3 trials (R and L)	3 trials (R and L)

6 Results and discussions

6.1 Comparison between manual and automatic masks in the different setup conditions

To validate the human segmentation, the automatic masks are compared to the manual ones elaborated by an expert operator. For this reason, the Jaccard Index (JI) is calculated for each segmentation. JI is a statistic measurement for attesting the similarity of two samples and it is defined as follows:

$$JI = \frac{|M \cap A|}{|M \cup A|} \quad (6.1)$$

where:

- A is the automatic mask;
- M is the manual mask.



Figure 6.1: Automated and manual mask. The white surface is the automatic mask while the red line represents the edge of the manual mask drawn by an expert operator.

The entire data set is validated by calculating the JI and the results are shown in Table 1.

Table 1: Mean and standard deviation of the JI calculated for the entire data set

JI (3 trials)	Right Leg		Left Leg	
	1 LED lamp	2 Led lamps	1 LED lamp	2 Led lamps
With background	0.925 ± 0.013	0.929 ± 0.010	0.919 ± 0.011	0.925 ± 0.011
Without background	0.864 ± 0.030	0.909 ± 0.019	0.833 ± 0.040	0.885 ± 0.026

As shown in Table 1, the first evidence is represented by the worsening of the JI values when the segmentation is computed without the background in all the considered conditions. In fact, by considering the right leg it decreases of 6.6 % and 2.2% in one and two leds light conditions, respectively, while for the left leg it decreases of 9.4 % and 4.5% in one and two leds light conditions. The highest inaccuracies are in the body regions which have the same color of the background behind. As example, in the situation illustrated in Figure 6.2, the legs and floor behind have the same pixel intensity and they cannot be distinguished after the background subtraction. For this reason, the threshold cannot be able to effectively separate these regions from the background (as shown in Figure 6.2)

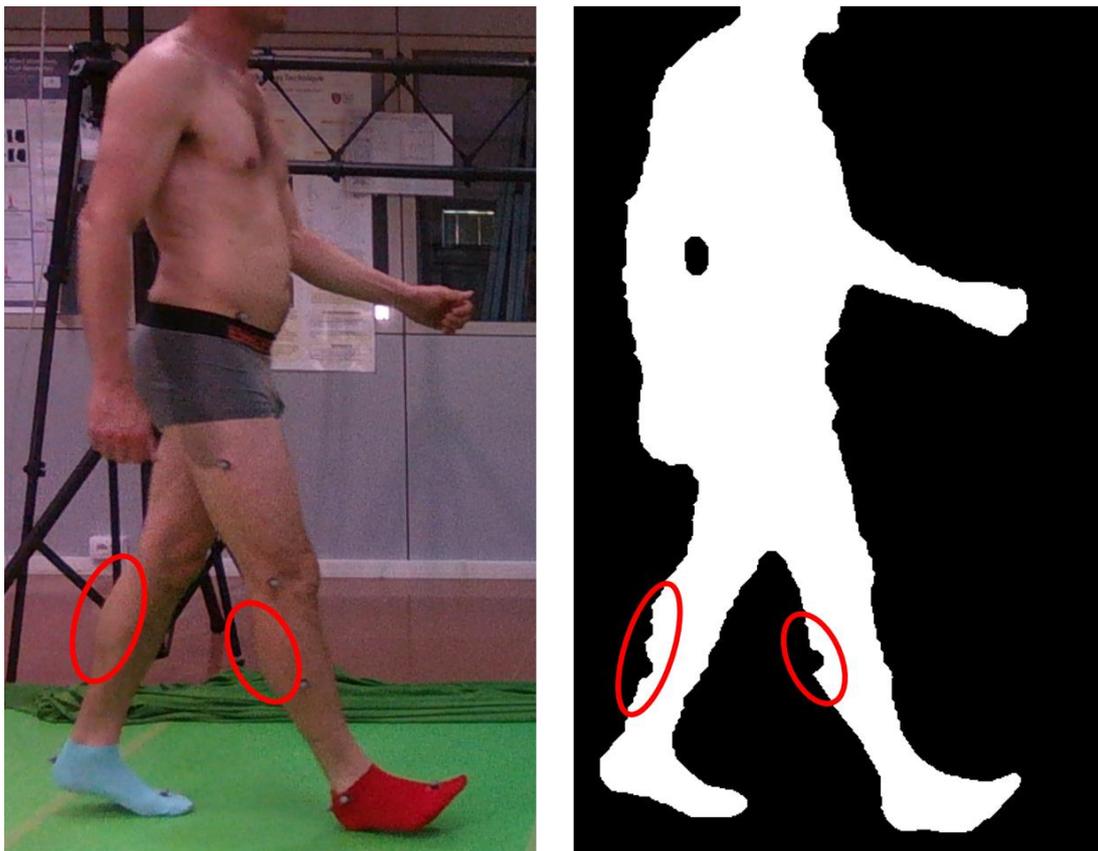


Figure 6.2: Human segmentation. The red lines represent the segmentation inaccuracies because these regions have the same colours of the background behind.

Furthermore, if the background is not homogeneous the separation of the subject from the former is corrupted by what is behind the subject body. As example, in Figure 6.4 and Figure 6.3, it is shown a comparison between the images D obtained with and without the green background. It is possible to assess that, in the case without the green background, the subject region in the region D is affected by the portion of the background which lies behind the subject. For this reason, the threshold identification for the modified version of Salvi and Molinari (2018) is more difficult without the green background.



Figure 6.4: The difference image D obtained with green background. The green background is homogeneous and for this reason, the difference image D is not affected by the background behind the subject.

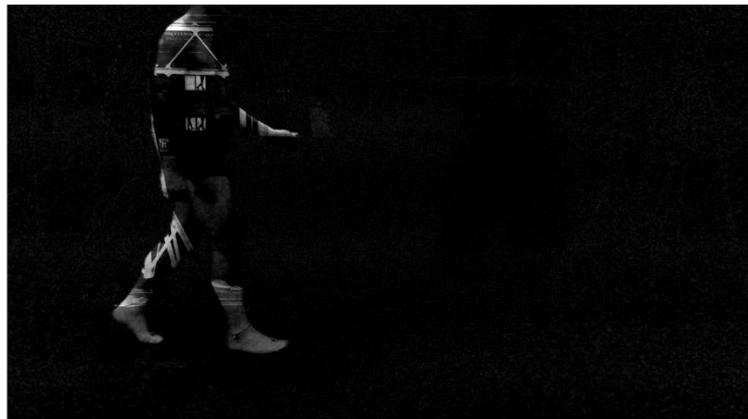


Figure 6.3: The difference image D obtained without the green background. The background is not homogeneous and for this reason, the difference image D is affected by the portion of the background which lies behind the subject.

By analysing the performances of the segmentation algorithm obtained in the two different light conditions without the green background, the JI decreases when one LED lamp is removed (5 %). The LED lamp is positioned behind the camera and it raises the local brightness, which means more inhomogeneity, due to the concentration of the collimated light beam. In particular, this effect is emphasized in presence of a high reflective background, like the glass window in Figure 6.5.



Figure 6.5: This image presents the experimental setup without the green background and with one additional LED lamp. It is possible to notice the light reflections on the background which increase the homogeneity of the background.

The light condition affects the homogeneity of the background, in fact in presence of two additional leds, that are positioned on the right and left sides of the camera (as described in Chapter 5.2) the resulting light beam is more distributed over the background surface as depicted in Figure 6.6

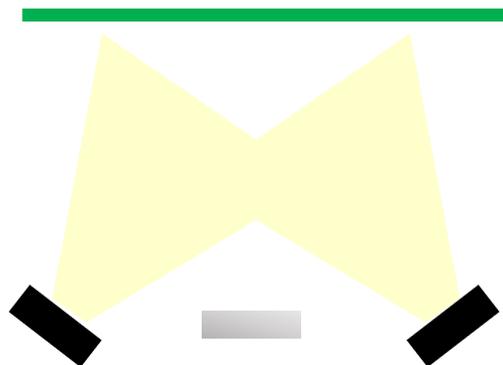


Figure 6.6: Experimental set up with two additional led lamps. The grey rectangle represents the camera and the black rectangles are the two additional led lamps.

When the green background is present, there are no appreciable differences in terms of JI between two and one leds conditions. In fact, the percentage difference amounts to 0.4% for the right leg and 0.6% for the left side. This could be explained by calculating the luminance (mean value of the image intensity) of the green background region in the two different light conditions which amounts to 64.13 and 63.76 for one and two LED lamps, respectively. Consequently, when the green background is

present, it is possible to conduct the markerless gait analysis by using only a one additional LED lamp which simplifies the experimental set up.

In general, the segmentation accuracy is very repeatable in fact the standard deviation of the JI values is lower than 0.04 which it is obtained without background and with one LED lamp. This combination represents the worst experimental condition in terms of segmentation performances.

6.2 Comparison between marker-based and 2D markerless kinematics

To validate the estimate lower limb joint kinematics, the joint angles obtained from the 2D GA are compared with the ones from the 3D GA through the SP system. For this reason, the Root Mean Square Deviation (RMSD) between the two extracted kinematics of the lower limb is computed.

To compare this two protocols, the offset between the two kinematics should be removed because the 3D GA and the 2D GA consider different anatomical points and consequently different anatomical axes around which the three joint angles are decomposed.

During the experimental acquisitions, three trials (for each side, Right and Left) are acquired for each experimental conditions and for this reason the mean curve of the three trials is computed both for the ML protocol and SP system to estimate the offset between them. All the obtained curves are shown in Figure 6.7 and Figure 6.8.

Right Leg, With background vs without background, 2 LED

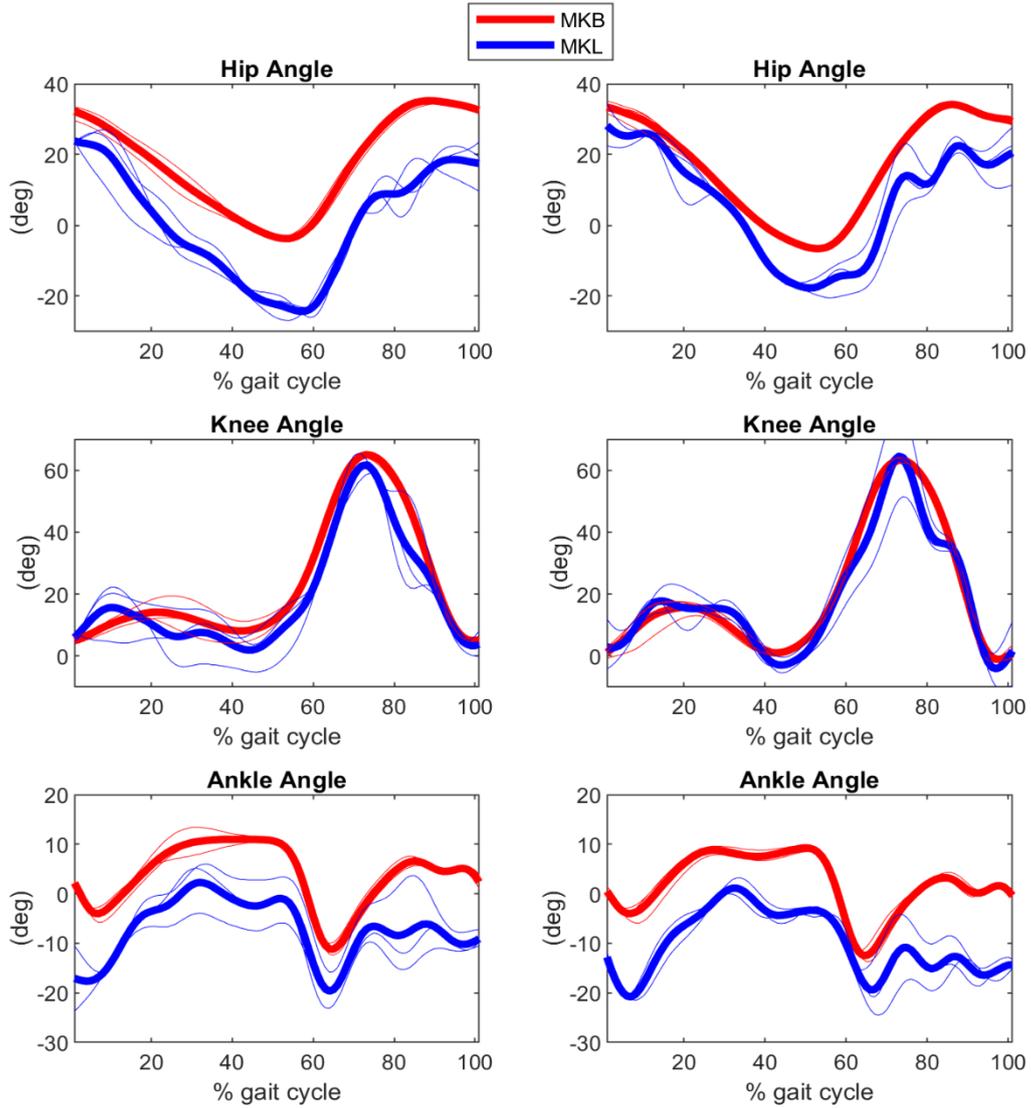


Figure 6.7. Lower limb kinematics obtained from the 2D GA (red lines) and from 3D GA (blue lines) for 3 trials with two leds, with background (left) vs without background (right). The red line in bold is the mean curve of the 3 trials obtained from 2D GA and the blue line in bold is the mean curve of the 3 trials obtained from the 3D GA.

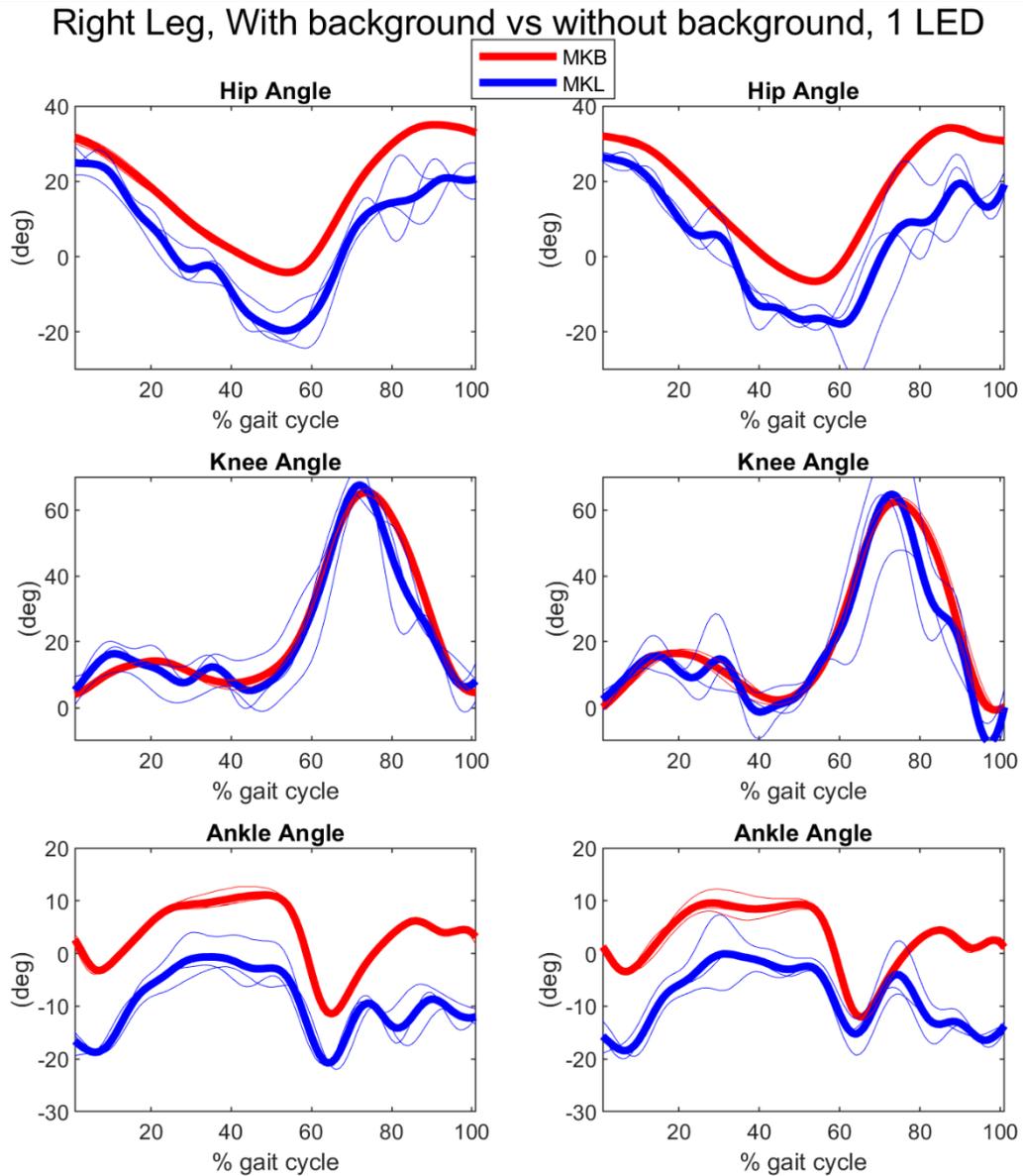


Figure 6.8: Lower limb kinematics obtained from the 2D GA (red lines) and from 3D GA (blue lines) for 3 trials with one LED lamp, with background (left) vs without background (right). The red line in bold is the mean curve of the 3 trials obtained from 2D GA

The offset between the mean kinematics of hip, knee and ankle joint are evaluated and shown in Table 2, Table 3, Table 4, respectively.

Table 2: The mean and standard deviation of the offset between the mean curves of the hip kinematics obtained from 3D GA and 2D GA.

Offset Hip (3 trials) (deg)	Right Leg		Left Leg	
	1 LED lamp	2 Led lamps	1 LED lamp	2 Led lamps
With background	12.2 ± 3.3	16.9 ± 4.5	7.81 ± 4.5	11.4 ± 3.5
Without background	11.0 ± 3.3	12.4 ± 3.5	12.8 ± 4.6	9.9 ± 4.7

Table 3: The mean and standard deviation of the offset between the mean curves of the knee kinematics obtained from 3D GA and 2D GA.

Offset Knee (3 trials) (deg)	Right Leg		Left Leg	
	1 LED lamp	2 Led lamps	1 LED lamp	2 Led lamps
With background	3.5 ± 3.2	5.4 ± 3.3	4.4 ± 2.8	4.9 ± 3.3
Without background	4.6 ± 3.6	2.7 ± 2.4	4.0 ± 4.0	3.4 ± 3.2

Table 4: The mean and standard deviation of the offset between the mean curves of the ankle kinematics obtained from 3D GA and 2D GA.

Offset Ankle (3 trials) (deg)	Right Leg		Left Leg	
	1 LED lamp	2 Led lamps	1 LED lamp	2 Led lamps
With background	12.9 ± 2.9	10.9 ± 2.7	9.9 ± 3.2	11.7 ± 3.1
Without background	14.0 ± 3.9	13.5 ± 2.9	11.2 ± 4.7	12.0 ± 3.6

By analysing the values listed in the tables, it is possible to notice that the offset of the sagittal hip and the ankle angles are higher than one of the knee angle because the anatomical axes in are defined differently by the 3D GA and 2D ML for the hip and ankle. This can be justified since the 3D GA uses the markers on the pelvis and on the foot to obtain the anatomical axes for these segments to compute the corresponding joint angles. These information are not considered by 2D ML since the positions of the pelvis markers are not visible to the RGB-D camera during gait, while the markers located on the feet do not lie on the sagittal plane. Instead, for the knee joint the offset is lower because the anatomical axes are defined similar to those used by the SP system.

To evaluate the accuracy of the estimated lower limb kinematics, the RMSD is computed for all the experimental conditions after the removal of the offset. In this way, it is possible to evaluate only the relative variations of each joint angle.

Table 5: Mean and standard deviation of the RMSD values after removing the offsets of 3 trials between the hip joint kinematics obtained from 3D GA and 2D GA.

RMSD Hip (3 trials) (deg)	Right Leg		Left Leg	
	1 LED lamp	2 Led lamps	1 LED lamp	2 Led lamps
With background	5.7 ± 0.3	5.5 ± 1.5	5.5 ± 0.9	4.4 ± 1.4
Without background	6.1 ± 0.8	4.5 ± 1.3	7.3 ± 2.8	5.5 ± 1.2

Table 6: Mean and standard deviation of the RMSD values after removing the offsets of 3 trials between the knee joint kinematics obtained from 3D GA and 2D GA.

RMSD Knee (3 trials) (deg)	Right Leg		Left Leg	
	1 LED lamp	2 Led lamps	1 LED lamp	2 Led lamps
With background	6.9 ± 0.7	6.8 ± 1.0	6.9 ± 1.6	5.5 ± 0.2
Without background	7.2 ± 1.9	5.1 ± 2.1	7.4 ± 0.5	5.5 ± 1.2

Table 7: Mean and standard deviation of the RMSD values after removing the offsets of 3 trials between the ankle joint kinematics obtained from 3D GA and 2D GA.

RMSD Ankle (3 trials) (deg)	Right Leg		Left Leg	
	1 LED lamp	2 Led lamps	1 LED lamp	2 Led lamps
With background	3.5 ± 0.2	3.3 ± 0.3	3.8 ± 0.5	3.4 ± 0.4
Without background	4.2 ± 0.8	3.0 ± 0.3	4.9 ± 1.1	4.1 ± 0.2

By analysing the RMSD values, it is possible to notice that the mean errors for the three joints considered are lower when the acquisitions are conducted with 2 leds, both with and without green background. In particular, when the green background is present, the RMSD values obtained with one and two leds are similar. Consequently, when the green background is present, it is possible to conduct the markerless gait analysis by using only a one additional LED lamp which simplifies the experimental set up.

On the contrary, without the background the results obtained with two leds are more accurate. This aspect could be explained by considering that the mean JI index is lower than 0.9 without the green background and with one LED lamp. These inaccuracies, which arise in the first step of the protocol (human segmentation), are strongly reflected in the accuracy of the extracted lower limb kinematics. In fact, in this condition all the errors for each joint are higher than the error obtained in all the other experimental conditions because the inaccuracies in segmentation are the highest. This finding about the influence of light conditions is in line with those obtained for the JI accuracy and offset estimation. In general, the RMSD values in ankle kinematics (3.77° on average) are lower than the other joints (5.56° for the hip joint and 6.41° for the knee joint, on average) because the foot is directly extracted by implementing a color filter in the RGB domain without relying on the depth image in which is not possible to distinguish the foot, as shown Figure 6.9.

For this reason, the orientation of the foreground foot is well estimated and the errors in all the experimental conditions are, on average, lower than the other joints.

The higher RMSD values are obtained in knee joint kinematics because these errors include the orientation inaccuracy of both the thigh and the shank.

In the following paragraph, all the factors affecting the accuracy of the lower limb joint kinematics estimation are analysed.

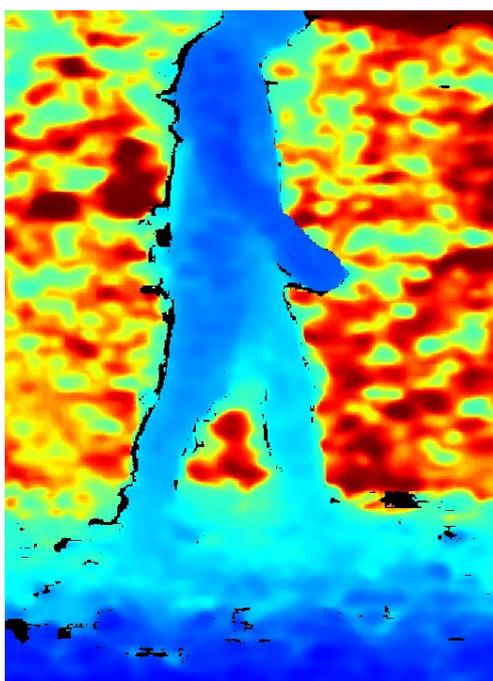


Figure 6.9: Depth image: it is not possible to distinguish the feet.

6.2.1 Factors affecting the accuracy of the joint kinematics estimation

In this paragraph, all the factors affecting the accuracy of the joint kinematics estimation are analysed to understand the reason of the obtained errors in terms of RMSD values.

6.2.1.1 Misalignment between RGB and depth images

The most important problem which causes the highest errors in terms of RMSD value is the misalignment between RGB and depth images. Even though the camera is calibrated by providing images (RGB and Depth images) which are overlapped, there are some misalignments between the two and these problems causes errors in orientations of the lower limb segments.

An example is reported in Figure 6.10 to show the effects of the misalignment between RGB and depth images on the extraction of the lower limb segments.

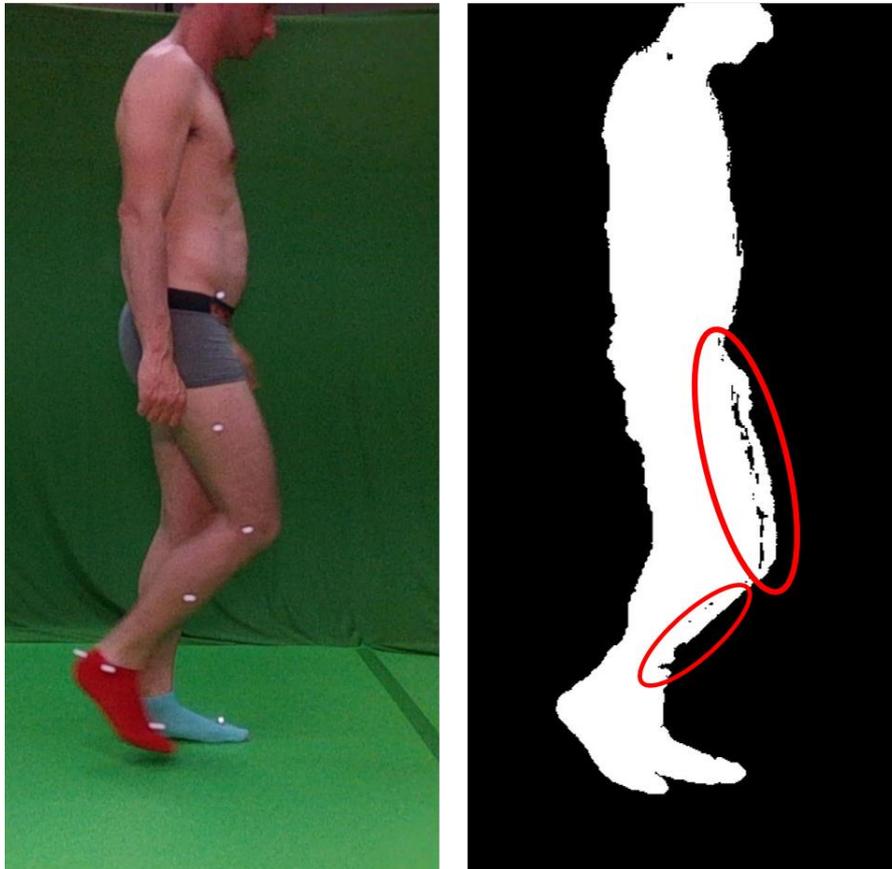


Figure 6.10: Misalignment between RGB and depth images. On the left the RGB image is shown, on the right the segmentation (obtained from RGB image) is overlapped with the depth image to obtain only the depth region inside the segmentation mask. The red lines represent the shift of the depth image compared to the segmentation obtained from the RGB images.

To assess the misalignment between the RGB and depth images, the segmentation mask obtained by the elaboration of RGB image is overlapped with the depth image to select only the region of the depth image inside the segmentation mask. After this step, as shown on the right side of the Figure 6.10, the resulting image, which should contain only the extracted subject, includes also some regions of the background represented by higher depth values. The presence of some parts of the background in the resulting image is caused by a shift of the depth image compared to the RGB image. This image is taken as input in the algorithm for extracting the foreground lower limb and this misalignment influences the inclination of the thigh and shank as shown in Figure 6.11.

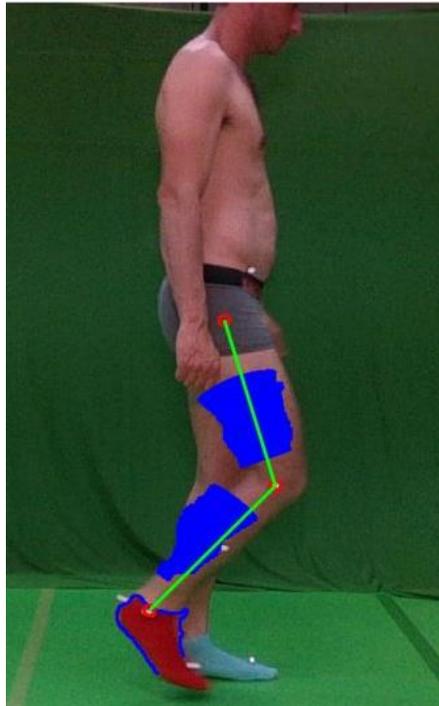


Figure 6.11: The blue surfaces represent the extracted thigh and shank of the lower limb. It is possible to see the effect of the misalignment between the RGB and depth images as the extracted thigh and shank are shift compared to the real position of the two segments (thigh and shank). The RMSD value is about 7 degrees because there is an evident shift between the extracted and real position.

These misalignments cause the highest errors in terms of RMSD values (about 6-7 degrees). The misalignments between the two images acquired by the camera are not present in all trials. In fact, when the misalignment is absent as shown in Figure 6.12, the RMSD values are low (about 3-4 degrees).

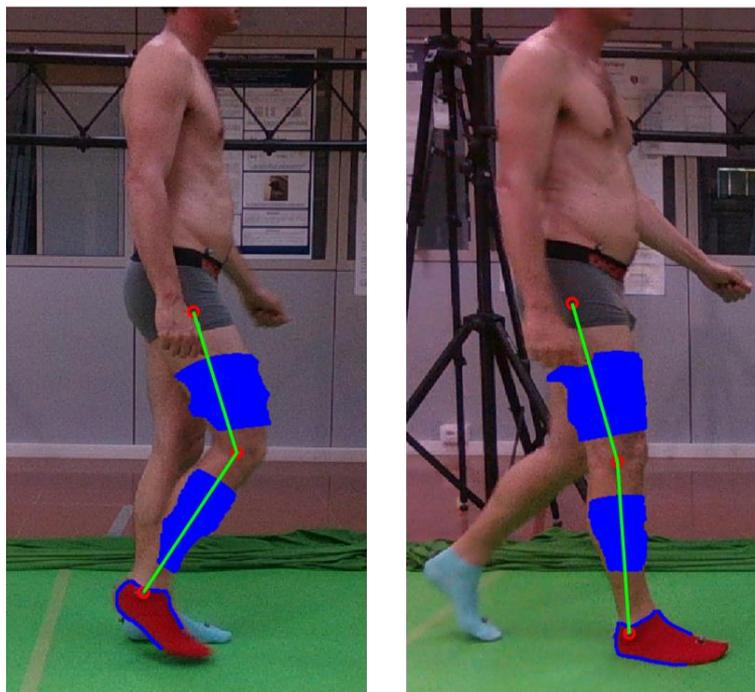


Figure 6.12: The blue surfaces represent the extracted thigh and shank of the lower limb. In this trial the RGB and depth images are aligned and for this reason the RMSD value is about 3 degrees.

The example shown in Figure 6.12 represents the extracted lower limb segments obtained in trial without the green background which is characterised by a segmentation accuracy lower than the acquisitions conducted with the green background (shown in Figure 6.11). This means that one of the biggest is the misalignment between the two images and the segmentation inaccuracy could be accepted on condition that the majority of the body is segmented properly (this aspect will be discussed in the following paragraph '*Accuracy of the subject segmentation*').

Another consideration regards the fact that the ankle kinematics presents the lowest RMSD value because the foot is not extracted from the depth image and for this reason this is not affected by the problem of the misalignment.

In conclusion, the calibration of the camera may be customized in order to obtain the best alignment between the two images acquired by the camera (this aspect will be analysed in the chapter '*Future works*').

6.2.1.2 Accuracy of the subject segmentation: how errors in segmenting the subject affect joint angles estimation

As shown in Table 1, the JI values obtained during the acquisitions without the green background are lower than the ones obtained with the green background. In particular, with two leds the performances are higher than ones with only one LED lamp because, as explained in paragraph '*Comparison between manual and automatic masks in the different setup conditions*', this condition decreases the homogeneity of the background which is a fundamental requirement for a meaningful implementation of the modified version of Salvi and Molinari (2018). In presence of the green background the choice of an optimal threshold very difficult.

The background of the considered images is very difficult to treat because, as shown in Figure 6.13 :

- The floor has a similar color of the skin of the subject;
- It is composed by different coloured regions;
- There are high reflective regions i.e. glass window.



Figure 6.13: It is possible to notice several different coloured regions which increase the inhomogeneity of the background.

The most relevant problem is the presence of a floor which presents the same color of the some body regions of the subject. This aspect involves into two difference consequences:

- After the background subtraction the regions belonging to the subject which present the same color of the floor behind will have pixel intensities about equal to zero. For this reason, the threshold is not able to distinguish these regions of the subject from the background because the pixel values are the same;
- It is not possible to obtain the edge of the subject by implementing an image gradient because these regions have also a similar color of the floor around them as shown in Figure 6.14;

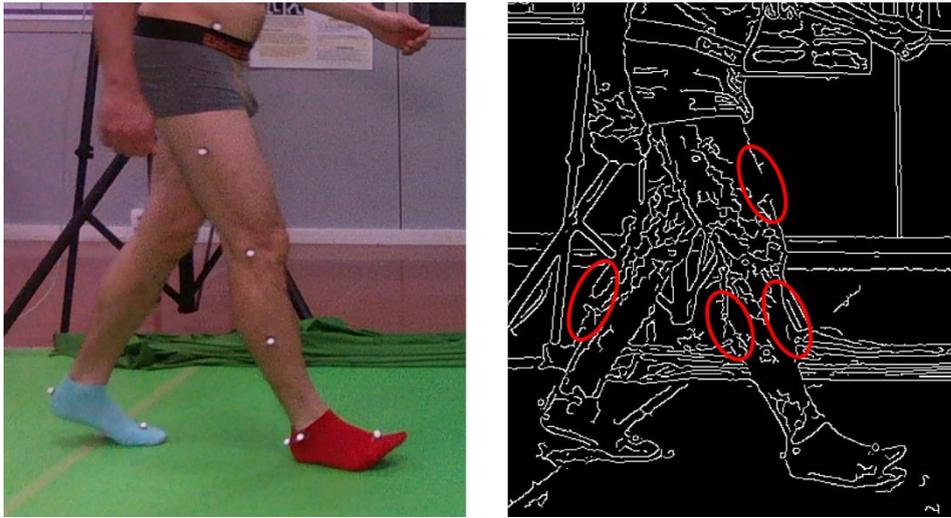


Figure 6.14: Image gradient obtained by using canny detector. The red lines surround the discontinuities of the gradient because these regions have similar colours to the floor.

The result of these two consequences is an under segmentation because these regions are not included in the mask as shown Figure 6.15.

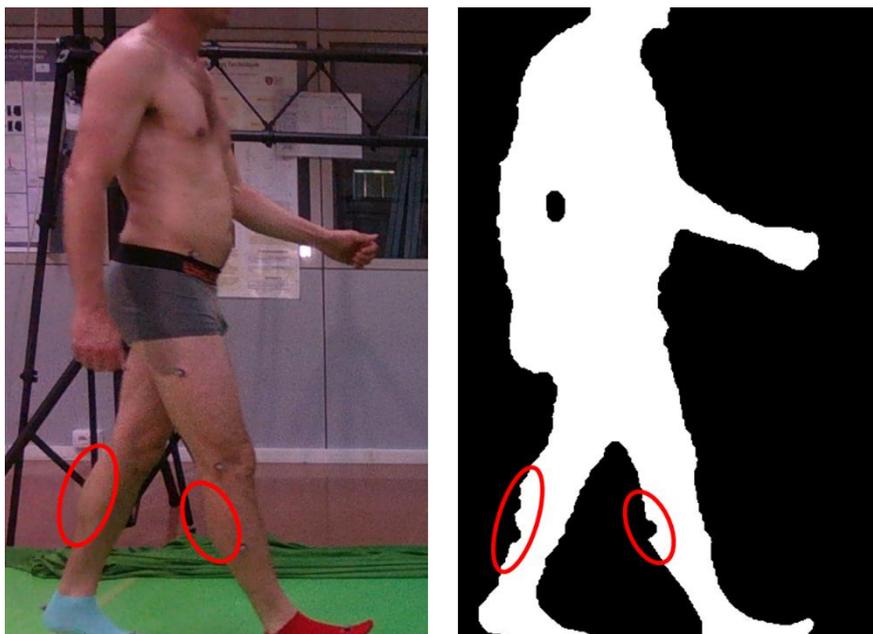


Figure 6.15: Human segmentation. The red lines represent the under segmentations because these regions have the same colours of the background behind.

7 General conclusions

The feasibility of estimating the lower limb joint kinematics by means of a markerless protocol without a homogeneous background has been dealt in this thesis project. In fact, if on one hand the optoelectronic stereophotogrammetry gait analysis represents the gold-standard in clinical practice, on the other there is still a need for simpler and faster assessment protocols.

Nowadays, markerless methods are not widely available for biomechanical and clinical applications because the accuracy of the joint kinematics estimates is not sufficient to assess pathologies or gait alterations. However, recent advancements in depth camera (RGB-D) technology and video processing have paved the way to the development of a new generation of low-cost movement analysis systems because the use of a single camera reduces the number of tools required for gait analysis.

The proposed two-dimensional markerless gait analysis protocol uses a single camera which provides the RGB images together with the depth distance of each pixel which could be useful to identify different regions of the human body to improve the quality of the segmentation. In fact, the depth information is used to distinguish the foreground lower limb from the rest of the image.

The proposed protocol consists of two parts:

- Image segmentation to separate the subject from the background by implementing a modified version of Salvi *and Molinari (2018)*, adapted to human segmentation by introducing anatomical and biomechanical hypothesis. The human segmentation algorithm used in this thesis is a fully automatic thresholding method which is evaluated on the image histogram resulting from the subtraction between a frame containing the subject and a frame with only the background captured at the beginning of the trial. The choice of an automatic setting of the threshold value only on the basis of the image histogram makes this algorithm robust to variable backgrounds without the need of a single coloured towel as in the previous experimental set-up conditions. The robustness is limited only by the level of heterogeneity of the environment behind the subject.
- Two-dimensional joint kinematics estimation: the implementation of a modified version of Pantzar-Castilla *et al (2018)* to extract the foreground lower limb and to estimate the lower limb joint sagittal kinematics through a participant-specific multi-segmental model which allows to generalize the proposed method also for subjects with gait alterations. The anatomical landmarks are manually identified in three different reference frames which represent static upright standing, load and swing phases and the trajectories of the joint centers are automatically estimated in all the other frames by using the Iterative Closest Point (ICP) algorithm.

An experimental session is conducted to validate the entire markerless protocol. Prior to performing the experiments, the quality of the images provided by Kinect II and Intel RealSense D435 is performed. The second camera is selected since the Kinect II produces blurred images which make it difficult to correctly identify the foreground foot and shank.

The experimental session is executed on one subject who is asked to walk straight at a self-selected speed on a carpet. The experiments are conducted with and without a green background and in two different light conditions (two and one additional LEDs) to quantify how these factors affect the accuracy of both the segmentations and the joint angle estimates.

The segmentation algorithm relative performances are validated by calculating the Jaccard index between the obtained segmentations and the corresponding manual masks for each segmentation.

The computed joint angles are finally compared with those obtained from a SP system (only the sagittal plane is considered) for the different setup conditions by considering the mean waveforms obtained from three trials. In particular, the offset between the mean waveforms and the RMS value of the waveforms difference without their offset are computed for each condition under analysis.

The first evidence, by analysing the mean JI values, is their worsening when the segmentation is computed without the background in all the considered conditions. In particular, by considering the right leg the performances decrease of 6.6 % and 2.2% in one and two LEDs light conditions, respectively, while for the left leg they decrease of 9.4% and 4.5% in one and two LEDs light conditions.

The most relevant problems during the segmentation without the green background are:

- the presence of a floor which presents a similar color of some body regions of the subject, which is reflected in an under segmentation because these regions are not included in the mask;
- the reduction of the homogeneity of the background due to the presence of different coloured and also high reflective regions (i.e. glass window).

The percentage decrease of the JI values, obtained without the green background, decrease of 5% passing from two additional LEDs to only one because the effects of the two above-mentioned problem are emphasized since the light-beam is more focused at the expense of the brightness homogeneity.

By analysing the offset of the mean joint kinematics waveforms, the hip and the ankle exhibit higher values than one of the knee. This could be explained by considering that the anatomical axes of hip and ankle defined within the stereophotogrammetric protocols are different from those defined for 2D video-based gait analysis. Instead in the case of knee joint, the reference axes are defined starting from the same anatomical landmarks in both protocols and this explains the lower offset values. For the sake of completeness, it has also to be said, however, that the stereophotogrammetry protocols assume the centre of knee joint as the middle point between the right and left epicondyles of each knee. On the contrary, the RGB-D camera, being positioned orthogonal to the subject's sagittal plane, can obviously record the movement of the closest epicondyle only. This difference could lead to some disagreements when comparing the results obtained from the two methods.

In general, by considering the RMSD values, the average errors for all the three joints considered are lower when the acquisitions are conducted with two LEDs both with and without the green background. When the green background is present, the RMSD values two LED lamps are slightly better with respect to one LED lamp only, while without the green background, the results obtained with two LED lamps are more accurate (about 2° on average). Consequently, when the green background is present, it is possible to conduct the markerless gait analysis by using only one additional led which simplifies the experimental set up.

Without the green background, the results obtained with two LEDs are more accurate. This aspect could be attributed to the inaccuracies which arise in the first step of the protocol (human segmentation). In fact, in this condition the mean JI index is lower than 0.9. In general, it is expected that the errors of the segmentation algorithm, will be strongly reflected in the accuracy of the extracted lower limb kinematics.

The RMSD values in ankle kinematics (3.8° on average) are lower than the other joints (5.6° for the hip joint and 6.4° for the knee joint, on average) because the foot is directly extracted by implementing a color filter in the RGB domain without considering the depth information.

The most important problem which causes the highest errors in terms of RMSD value is the misalignment between RGB and depth images since the accuracy of the factory calibration is not completely effective in aligning the two images. These misalignments are reflected in the estimated orientations of the lower limb segments. Furthermore, the misalignment is not constant, but it varies depending on the position within the field of view of the camera.

The problem can be reduced by a custom calibration of the camera to obtain the best alignment between the two images.

In conclusion, to sum up, the main findings of this study are that it is possible to conduct a markerless gait analysis acquisition without the need of a homogeneous background since with two LED lamps the decrease of the performances is neglectable (only 0.2° on average) with respect to the acquisitions with the green background. This represents an excellent result since the experimental set-up is simplified in favour of both system transportability and installation.

7.1 *Limitations and future work*

For what concerns the two-dimensional markerless gait analysis protocol for estimating the lower limb joint kinematics some limitations must be underlined. First of all, this technique requires the need of coloured socks (in this case red for the right foot and blue for the left) used to easily identify the feet.

Secondly, it is possible to analyse only few gait cycles because the study is limited to the field of view of the camera. Furthermore, the experimental set-up to guarantee good performances requires the use of two additional LEDs which increases the number of tools needed. Finally, an expert operator is required to identify the anatomical landmarks. For this reason, the proposed protocol is not fully automated, but this step takes only few seconds in a clinical field where it is not required to process a large quantity of images.

The following improvements can be implemented to overcome the discussed limitations:

- custom calibration of the RGB-D camera to be sure that RGB and depth images are well-aligned. To calibrate the camera both the intrinsic and extrinsic parameters must be calculated i.e. focal length, principal point, distortion, rotations and translations. The tool required including a calibration target which must be seen from both the RGB and depth sensor. The minimum number of images required is six to see the target from different viewpoint. the choice of the viewpoint is critical to the accuracy of the calibration results and in general the six images should cover as much as possible the entire FOV of the camera;
- the reduction of manual approach by identifying the anatomical landmarks only in the first frame and by implementing a method which compensate for the parallax effect and its impacts on size and shape distortion of the lower limb;
- the improvement of the human segmentation by introducing a morphological model of the lower limb to fill the missing regions which are not included in the segmentation mask because of the similar color of the background;
- to automatically identify the gait cycle to analyse by adapting the method presented in Zeni Jr *et al* (2007) to ML technique. In particular, this method is based on the position of a marker placed on the heel. In particular, the heel strike event is identified when the longitudinal coordinate of the heel marker reaches the maximum distance from the marker placed on the pelvis. This method may be adapted to ML protocol by identifying the centroid of feet and pelvis automatically for each frame.
- to extend the proposed protocol to other sports activities i.e. running to study the related biomechanical parameters.

Appendix

This thesis project is a part of a European project with a Swedish hospital center. In particular, in Sweden there is a preventive follow-up program for children with cerebral palsy to increase the chances of identifying children at risk of crouch gait Figure 0.1.

For this reason, the 2D markerless gait analysis protocol proposed in Castilla et al [] was validated on 18 children with cerebral palsy by using a single RGB-D camera. In particular, firstly I have been collaborating with this Swedish hospital center to study the advantages and limitations of the adopted 2D gait analysis markerless protocol. For this reason, I conducted the preliminary study described in this thesis on healthy subjects to evaluate how the experimental conditions affect the quality of the estimated lower limb kinematics. The results will simplify and improve the experimental scenario in the Swedish hospital center. In particular, Kinect II camera, which provides blurred images, will be replaced with the Intel Real Sense D435. Moreover, since the quality of the estimated joint angles is satisfactory the number of additional LEDs will be decreases from four LEDs to only two. Finally, the best advantage is the removal of the green background which complicates the experimental set up and increase the time preparation before each acquisitions.

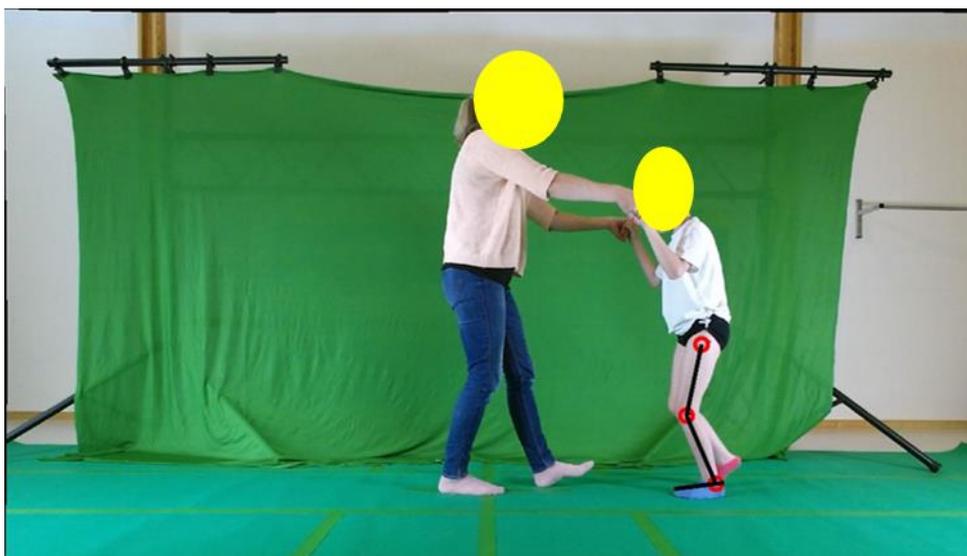


Figure 0.1: 2D GA protocol applied on a child with cerebral palsy who is helped to walk by the clinician. For this reason, also this situation should be considered in the implementation of the algorithm.

The first part of this thesis project was also aimed at conducting a comparative study of the performances of the most common segmentation algorithms in terms of Jaccard index. This works has been accepted on 'Gait and posture' as a conference proceedings of the Italian Society of the Clinical Movement Analysis (SIAMOC) held in Bologna in October 2019.

Bibliography

- [1] M. Roberts, D. Mongeon, and F. Prince, "Biomechanical parameters for gait analysis: a systematic review of healthy human gait," *Phys. Ther. Rehabil.*, vol. 4, no. 1, p. 6, 2017.
- [2] E. Pantzar-Castilla *et al.*, "Knee joint sagittal plane movement in cerebral palsy: a comparative study of 2-dimensional markerless video and 3-dimensional gait analysis," *Acta Orthop.*, vol. 89, no. 6, pp. 656–661, 2018.
- [3] E. E. Butler, K. M. Steele, L. Torburn, J. G. Gamble, and J. Rose, "Clinical motion analyses over eight consecutive years in a child with crouch gait: A case report," *J. Med. Case Rep.*, vol. 10, no. 1, pp. 1–10, 2016.
- [4] K. J. Bell, S. Ounpuu, P. A. DeLuca, and M. J. Romness, "Natural progression of gait in children with cerebral palsy," *J. Pediatr. Orthop.*, vol. 22, no. 5, pp. 677–682, Sep. 2002.
- [5] A. Castelli, G. Paolini, A. Cereatti, and U. Della Croce, "A 2D markerless gait analysis methodology: Validation on healthy subjects," *Comput. Math. Methods Med.*, vol. 2015, 2015.
- [6] V. Agostini, G. Balestra, and M. Knaflitz, "Segmentation and classification of gait cycles," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 5, pp. 946–952, 2014.
- [7] S. Mohammed, A. Samé, L. Oukhellou, K. Kong, W. Huo, and Y. Amirat, "Recognition of gait cycle phases using wearable sensors," *Rob. Auton. Syst.*, vol. 75, pp. 50–59, 2016.
- [8] D. Bookwater, G. Williams, and B. Blidy, "Letter to the Editor," *Art Educ.*, vol. 55, no. 3, p. 5, 2002.
- [9] P. Komdeur, F. E. Pollo, and R. W. Jackson, "Dynamic Knee Motion in Anterior Cruciate Impairment: A Report and Case Study," *Baylor Univ. Med. Cent. Proc.*, vol. 15, no. 3, pp. 257–259, 2002.
- [10] G. Wu *et al.*, "ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion - Part II: Shoulder, elbow, wrist and hand," *J. Biomech.*, vol. 38, no. 5, pp. 981–992, 2005.
- [11] C. P. Charalambous, "Measurement of lower extremity kinematics during level walking," *Class. Pap. Orthop.*, vol. 8, no. 3, pp. 397–398, May 2014.
- [12] R. Baker, "The history of gait analysis before the advent of modern computers," *Gait and Posture*, vol. 26, no. 3, pp. 331–342, Sep-2007.
- [13] T. P. Andriacchi and E. J. Alexander, "Studies of human locomotion: Past, present and future," *J. Biomech.*, vol. 33, no. 10, pp. 1217–1224, 2000.
- [14] L. Mündermann, S. Corazza, and T. P. Andriacchi, "The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications," *J. Neuroeng. Rehabil.*, vol. 3, 2006.
- [15] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," *Proc. Int. Conf. 3-D Digit. Imaging Model. 3DIM*, pp. 145–152, 2001.
- [16] P. J. Besl and N. D. McKay, "A Method for Registration of 3-D Shapes," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, vol. 14, no. 2, pp. 239–256.

-
- [17] A. Leu, D. Ristic-Durrant, and A. Graser, "A robust markerless vision-based human gait analysis system," *SACI 2011 - 6th IEEE Int. Symp. Appl. Comput. Intell. Informatics, Proc.*, pp. 415–420, 2011.
- [18] G. L. Alexander, T. C. Havens, M. Skubic, M. Rantz, J. M. Keller, and C. Abbott, "Markerless human motion capture-based exercise feedback system to increase efficacy and safety of elder exercise routines," *Gerontechnology*, vol. 7, no. 2, pp. 1–6, 2008.
- [19] D. A. Winter, *Biomechanics and Motor Control of Human Movement: Fourth Edition*. John Wiley & Sons, 2009.
- [20] M. Goffredo, J. N. Carter, and M. S. Nixon, "2D markerless gait analysis," *IFMBE Proc.*, vol. 22, no. 1, pp. 67–71, 2008.
- [21] W. T. Dempster and G. R. L. Gaughran, "Properties of Body Segments Based on Size and Weight'," vol. 18, no. 7414, pp. 33–54, 1889.
- [22] D. Mehta *et al.*, "VNect: Real-time 3D human pose estimation with a single RGB camera," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–13, 2017.
- [23] R. Girshick, J. Shotton, P. Kohli, A. Criminisi, and A. Fitzgibbon, "Efficient regression of general-activity human poses from depth images," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 415–422, 2011.
- [24] M. Salvi and F. Molinari, "Multi-tissue and multi-scale approach for nuclei segmentation in H&E stained images," *Biomed. Eng. Online*, vol. 17, no. 1, pp. 1–13, 2018.
- [25] A. Cereatti *et al.*, "Standardization proposal of soft tissue artefact description for data sharing in human motion measurements," *J. Biomech.*, vol. 62, pp. 5–13, 2017.
- [26] A. Cappello, A. Cappozzo, P. F. La Palombara, L. Lucchetti, and A. Leardini, "Multiple anatomical landmark calibration for optimal bone pose estimation," *Hum. Mov. Sci.*, vol. 16, no. 2–3, pp. 259–274, 1997.
- [27] A. Cereatti, D. Trojaniello, and U. Della Croce, "Accurately measuring human movement using magneto-inertial sensors: Techniques and challenges," *2nd IEEE Int. Symp. Inert. Sensors Syst. IEEE ISSS 2015 - Proc.*, pp. 1–4, 2015.
- [28] R. A. Clark, B. F. Mentiplay, E. Hough, and Y. H. Pua, "Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and Kinect alternatives," *Gait Posture*, vol. 68, pp. 193–200, 2019.
- [29] J. Sell and P. O'Connor, "The xbox one system on a chip and kinect sensor," *IEEE Micro*, vol. 34, no. 2, pp. 44–53, 2014.
- [30] D. Pagliari and L. Pinto, "Calibration of Kinect for Xbox One and comparison between the two generations of microsoft sensors," *Sensors (Switzerland)*, vol. 15, no. 11, pp. 27569–27589, 2015.
- [31] S. Liu, D. Gao, P. Wang, X. Guo, J. Xu, and D. X. Liu, "A depth-based weighted point cloud registration for indoor scene," *Sensors (Switzerland)*, vol. 18, no. 11, pp. 1–11, 2018.
- [32] M. Carfagni *et al.*, "Metrological and critical characterization of the intel D415 stereo depth camera," *Sensors (Switzerland)*, vol. 19, no. 3, 2019.
- [33] J. Mitchelson and A. Hilton, "Wand-based multiple camera studio calibration," *Cent. Vision, Speech Signal Process*, pp. 1–30, 2003.