Politecnico di Torino

POLITECNICO DI TORINO

MASTER'S DEGREE IN BIOMEDICAL ENGINEERING

Master's Degree Thesis

A preliminary validation of a 3D markerless method for estimating the kinematics of a two-segment foot model using a single RGB-D camera

Supervisors:

Prof. Ing. Andrea Cereatti

Ing. Diletta Balta

Author:

Francesco Richetto

Abstract

Nowadays, the measure of the gait and the lower limb joint kinematics is essential in clinical practice. The marker-based optoelectronic stereophotogrammetry gait analysis represents the gold standard to evaluate human motion since the measure of markers position is highly accurate. Despite its accuracy, the marker-based system has several limitations, such as the proper marker placement on the skin of the patient which implies considerable preparing time, the necessity of a prepared personal and the cost of dedicated spaces.
The video-based markerless system presents a promising and cost-effective alternative to the marker-based system, offering the advantage of speeding up the acquisitions by removing the need of markers placement on the subject's body. In the recent years, several companies have developed affordable RGB camera integrated with infrared depth sensor (RGB-Depth). The accessibility and affordability of those systems have made systems opened to wider range of applications, leading advancements in various fields that require movement analysis. These alternatives, based on a single RGB-Depth camera (e.g. Azure Kinect body tracking SDK, OpenPose), model the foot as a single segment without articulating the metatarso-phalangeal joint kinematics, which is crucial to guarantee an affective load of the foot and correct progression (Allan J. et al., 2020). Van den Herrewegen I et al., 2014, proposes a 3D multi-segmental foot model through a D3DScan4D (5 scanner units), which relies on surface shape to identify foot anatomy. However, the high cost associated with this approach limits its clinical applicability.
This thesis aimed to design a markerless method based on a single RGB-Depth camera to estimate sagittal ankle and metatarso-phalangeal kinematics using a two-segment 3D foot model and explore its clinical applicability on children with foot deformities. This approach suggests an extension of a 2D Markerless protocol proposed by Balta et al 2020 by including a 3D two segments foot model composed by two segments (Mid-Rear-foot and Forefoot segments).

The gait data of ten subject affected by Clubfoot (a congenital condition that affects the development of foot and ankle) were collected in the Gait Analysis Laboratory of Skaraborgs Hospital in Skövde, mean age 13 (7-17), 5 females and 5 males. The subjects were asked to walk at self-selected speed for six trials (three for left and three for right) in front of an RGB-Depth camera (Azure Kinect, fs = 30 fps) placed 2.5 m laterally to the walkway Two static lateral views (right and left side) of the subject were captured along with four static views of the foot (Frontal, Lateral, Medial and Posterior) placing the camera on the floor at a 0.6 m distance. External

anatomical landmarks (*LE*: lateral epicondyle, *LM*: lateral malleolus; MTP5: 5th metatarsophalangeal joint and *TOE:* the most distal point on the foot surface) were identified by palpation and marked with small labels.

This thesis work proposes to develop two parts. The first aims to develop an algorithm to create a 2-segment 3D foot model by merging the four static views by aligning three common points on the foot sole of each view (ZF: the most anterior point of Frontal view, XL: the most lateral point of Lateral view and ZP: the most distal point of the Posterior point). Then the foot template was calibrated by manually selecting on the image the three external anatomical landmarks *LM, MTP5* and *TOE*. Subsequently, based on the position of *MTP5*, the model was divided into two 3D segments: Mid-Rear Foot and Forefoot.
The second part aims to develop an algorithm able to estimate the joint kinematics during a gait cycle. The method proposed by Balta et al., 2020 was employed to identify the most central gait cycle based on the foot positions. Then a depth completion technique, based on a low pass filter, was implemented to reconstruct, during the gait trials, missing depth information by exploiting RGB information. The positions of *LM* and *MTP5* were reconstructed by matching the 3D foot template to the dynamic point clouds applying ICP algorithm (Besl et al., 1992), the *TOE* was identified as the most distal foot point. *LE* position was extracted implementing the 2D method proposed by Balta et al., 2020, in order to obtain the sagittal ankle angle.

The acquisitions of the markerless system were not performed synchronously to the marker-based system due to Infrared interferences from the MB system in the depth images acquired from the Azure Kinect caused by similar wavelengths (850 nm). Kinematic curves were validated against manually labelled anatomical landmarks on the RGB images. The accuracy of the proposed MS method was assessed in terms of offset between the two curves and waveform similarity by estimating the root mean square errors (RMSE) after removing their mean values.

The acquisition protocol aims to achieve a good balance between the accuracy of the modelling and the comfort of the patient, minimizing the possible discomfort, which the subject may feel in upright static position.
Despite the limited number of views, the proposed method showed acceptable performance in terms of model reconstruction (error on foot length (%): 5.2 (R), 5.7 (L)). Errors are mostly associated with the technological limitations of the RGB-D device (RMSE (deg): MTP: 4.8 (R), 5.3 (L); ankle: 3.5 (R), 3.9 (L) and offset (deg): MTP: 3.5 (R), 6.5 (L); ankle: 3.1 (R), 2.9 (L)). Results highlight the influence of depth completion on joint center estimation and the challenges it presents in high

speed due to motion blurs and inaccurate reconstruction by the device. The presence of missing points in foot reconstruction affects accuracy, especially in the Forefoot (Forefoot to Mid-Rear-Foot points ratio = 0.26), impacting joint position estimation. The non-rigid nature of the human foot and soft tissue artifacts pose limitations to the rigid assumption of the ICP algorithm and a pre-alignment improves its efficiency and accuracy. Future studies will be focused on applying more accurate depth completion techniques (e.g. inpainting-based or deep learning based models) to improve the reconstruction of missing depth values. In conclusion, considering the rapid technological advancement in depth sensing, the proposed approach seems to be a very promising solution for evaluating gait of subjects with foot deformities.

A summary of the work achieved in this thesis project named: *"A preliminary validation of a 3D markerless method for estimating the kinematics of a two-segment foot model using a single RGB-D camera"* has been submitted for the annual conference of the Italian Society of clinical Movement Analysis (SIAMOC) that will be held in Roma in October 2023.

Acknowledgments

At the end of this work, it is fundamental for me to express my sincere gratitude to some important people.
First of all, I need to thank Professor Cereatti, my primary supervisor, for making this work possible and for giving me the opportunity to be an active part of this project: without you, this entire work couldn't have been achieved.

Moreover, I need to demonstrate my immense gratitude to Ing. Balta, my other supervisor: your support was fundamental from the beginning to the end of this thesis project. Thanks for your great professionalism and availability, your advice have always been of great significance to me and to the achievement of my goal. Even if we have often been far from each other, you have always been there to guide me: I'll never forget it.

Last but not least, I would like to thank Dr. Riad, who was guide during the acquisition period in Sweden. Thank you for your great hospitality and for making me feel home from the first day. The time I spent in Sweden has been very important to me, because I had the opportunity to challenge myself in my work field and to discover a new, very different but amazing country.

Contents

1 Clinical relevance and aim of work

Walking is the most common activity in daily life and plays a crucial role in a person's overall health. The evaluation of gait, using instrumented measurements and the estimation of gait parameters, holds significant importance in clinical practice, biomechanical research, and sports-related applications.
Gait analysis is a valuable tool for analysing lower limb functionality and evaluating the effectiveness of clinical therapies and training sessions. It is particularly beneficial in assessing individuals with gait alterations such as cerebral palsy [1], Clubfoot, musculoskeletal disorders or neurological conditions. Additionally, patients who have undergone lower limb orthopedic surgery or have experienced related injuries that affect their normal locomotion can benefit from gait analysis. By examining gait parameters and identifying any deviations or abnormalities, clinicians can tailor treatment plans, rehabilitation programs, and interventions to improve overall mobility and functional outcomes.

Clubfoot is a congenital condition that affects the development of foot and ankle, with a worldwide incidence of 1 on 1000 live births. It is more frequent on males than on females and it occurs unilateral in 50% of the cases and the right side is more often affected [2]. Also called talipes equinovarus, this pathology is recognizable at birth: its rigidness makes it distinguishable from other positional foot anomalies. This pathology is not passively correctable and if it is left untreated, it could provoke infections, foot and leg deformities, pain, and limits mobility [3].
There are two different types of clubfoot:

- Congenital (CTEV)
- Idiopathic (ITEV)

The first one occurs at birth and it is normally caused by many environmental or genetic factors. Usually, this type of clubfoot is associated with other anomalies (chromosomal abnormalities or neural tube defects) and can be classed as isolated or syndromic. If the patient is affected by isolated clubfoot (the most common case), that means that it is the only anomality he suffers from. On the other hand, if the patient suffers from syndromic clubfoot, that means that he has other congenital abnormalities.
The second case, idiopathic talipes equinovarus, is a type of clubfoot that appears without any known cause and it occurs sporadically [4].
The diagnosis of clubfoot is detected with ultrasonography after the 16^{th} week of gestation and like hip dysplasia and idiopathic scoliosis, clubfoot is a developmental

deformation. The ligaments located at the posterior and medial aspects of the ankle and tarsal joints are significantly thick and tight. As a result, they restrict the foot from achieving proper dorsiflexion (*equinus position*) and cause the adduction and inversion of the navicular and calcaneus bones.

Figure 1.1: *Representation of normal foot (left) and foot affected by clubfoot* [5].

The treatment for CTEV and ITEV is similar and can involve a non-surgical method such as Ponseti technique or surgical methods, depending on the severity of the condition. Following this technique, the foot is gradually rotated, around the head of talus, followed by cast application during the entire childhood with the first treatment in the first five weeks after birth.

Figure 1.2*: Ponseti correction treatment* [6].

After the treatment, in the case of unilateral clubfoot, the treated foot is slightly shorter (mean 1.3 cm) and narrower (mean 0.4 cm) than the healthy foot. Also the limbs are different in terms of the circumference of the leg, that results smaller (mean 2.3 cm) than the normal one [6].

The common gait features of patient affected by clubfoot are:

- **Toe walking**: also known as equinus gait, it refers to a way of walking predominantly on the toes of affected foot with minimal or no contact between heel and ground.
- **Foot inversion**: the foot may turn inward or tilting during the swing phase. The sole faces inwards, towards the midline of the body producing an abnormal foot placement and reducing the stability.
- **Limping**: the subject affected by significant clubfoot deformity may show an important limp, due to a restricted range of motion and an altered position.
- **Shortened Stride Length**: the stride length may be reduced because of the mobility and the reduced flexibility of the affected foot.

***Figure 1.3**: Representation of Toe walking* [7] *and illustration of foot inversion* [8].

To date, marker-based (MB) systems are considered the gold standard for evaluating lower limb joint kinematics due to their high accuracy [9]. These systems utilize multiple cameras and advanced image processing techniques to track and determine the instantaneous 3D position of markers placed on specific anatomical landmarks of interest. By tracking the movement of these markers throughout the

gait cycle, MB systems can provide detailed information about joint angles, ranges of motion, and overall lower limb biomechanics. Despite its accuracy, the marker-based system has several limitations, such as the proper marker placement on the skin of the patient which implies considerable preparing time, the necessity of a prepared personal and the cost of dedicated spaces.

The Video-based markerless system presents a promising and cost-effective alternative to the marker-based system, offering the advantage of speeding up the acquisitions by removing the need of markers placement on the subject's body [10]. In recent years, the development of affordable RGB cameras integrated with infrared depth sensors (RGB-Depth) has opened up new possibilities for movement analysis in various fields. These systems, such as the Azure Kinect body tracking SDK and OpenPose, offer accessibility and affordability, allowing for wider applications. However, these alternatives typically model the foot as a single segment without articulating the metatarso-phalangeal joint kinematics, which is essential for proper foot load and progression [11]. Another approach proposed by Van den Herrewegen I et al., 2014 [12] involves a 3D multi-segmental foot model using D3DScan4D (5 scanner units), which relies on surface shape to identify foot anatomy. However, the high cost associated with this method limits its clinical applicability.

Following the idea of 3D multi-segmental foot of Van den Herrewegen I et al., 2014, the aim of the thesis project is to design a Markerless method based on a single RGB-Depth camera to estimate sagittal ankle and metatarso-phalangeal kinematics using a two-segment 3D foot model and explore its clinical applicability on children with foot deformities. The method proposes to expand the 2D subject-specific foot model developed by Balta et al., 2020 [13] to 3D subject-specific foot model composed by two segments: Mid-Rear-foot with ankle joint and Forefoot with fifth metatarso-phalangeal joint.

The first two section of this thesis provide a comprehensive introduction to gait analysis, with a specific focus on lower limb joint kinematics and optical methods used for assessing gait. Furthermore, a thorough literature review is conducted to explore the markerless approach in gait analysis, with a particular focus on the RGB-Depth sensor technology.

The second section provides to explain the methodology of the proposed thesis project, which is structured in four parts:

1) The first one aims to explain the algorithm proposed by Balta et al., 2020 [13] to identify the most central gait captured by the Azure Kinect based on the foot trajectories.
2) The second aims firstly to develop an algorithm to create a 2-segement 3D foot model by merging the four static views by aligning three common points on the foot sole of each view, and then the multi-segmental model proposed by Balta et al., 2020 in order create the 2D shank template was employed. Through a manual selection of the external landmarks on the image, the templates were calibrated.
3) The third part aims to develop an algorithm able to estimate the joint kinematics during a gait cycle. Through a depth completion technique, was implemented to reconstruct, during the gait trials, missing depth information by exploiting RGB information. The positions of the ankle and fifth metatarso-phalangeal joint were reconstructed by matching the 3D foot template to the dynamic point clouds applying ICP algorithm [14] while the position of knee joint centre was extracted implementing the 2D method proposed by Balta et al., 2020.
4) The last part is centred on the computation on the sagittal kinematics.

In the subsequent section the acquisition protocol and the main hardware component of this work which is the Microsoft Azure Kinect are described. The gait data of ten subject affected by Clubfoot were collected in the Gait Analysis Laboratory of Skaraborgs Hospital in Skövde, where asked to them to walk at self-selected speed in front of an RGB-Depth camera placed 2.5 m laterally to the walkway. The acquisitions of the markerless system were not performed synchronously to the marker-based system due to Infrared interferences from the MB system in the depth images acquired from the Azure Kinect caused by similar wavelengths (850 nm). For this reason, the kinematic curves of sagittal angles were validated against manually labelled anatomical landmarks on the RGB images and root mean square errors (RMSE) were estimated.

In the last section of work, the results are commented, taking into account the strengths and weaknesses. The limitations of the method are analysed, and suggestions for future development in ML-based gait analysis are proposed, aiming to enhance the estimation of sagittal foot kinematics.

2 Fundamentals of gait analysis

2.1 Gait cycle

Gait cycle can be defined as the time interval between two successive occurrences of one of the repetitive events of walking [15]. The cycle can be divided in two big phases: stance and swing. During the first one, the foot is on contact with the ground, while during the second it is moving forward.

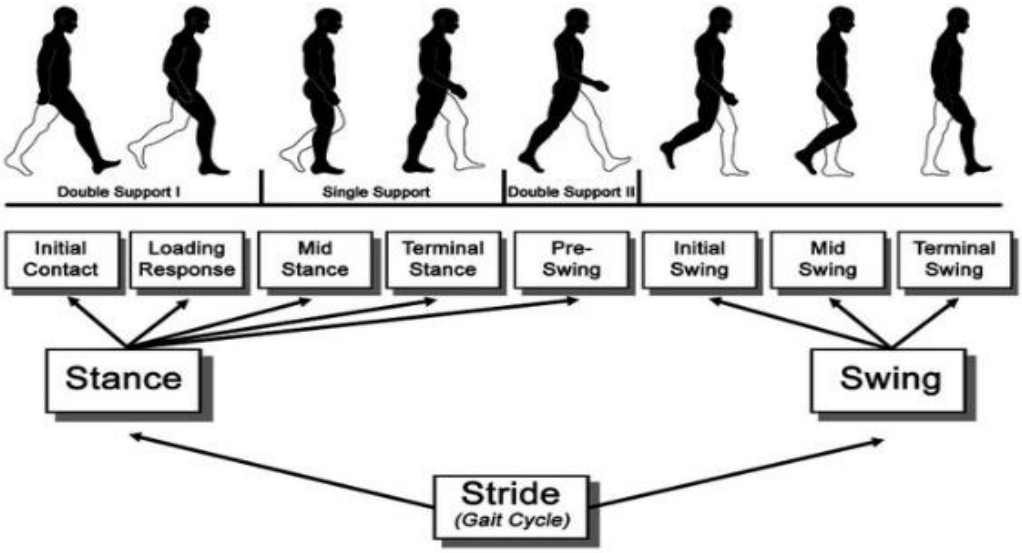

Figure 2.1: *Gait Cycle phases.* [16]

The two phases also can be divided in eight periods (*Figure 2.1*): the first five are in the stance phase (double and single support), while the last three are in the swing phase.

2.1.1 Stance

- *Initial contact*
  The ground contact is made by the heel (*heel strike*): the ankle is dorsiflexed, the knee is extended, and the hip is flexed. Body weight is moved into the foot.

- *Loading response*
  It is represented by the first 10 % of the gait cycle. The ankle plantar flexion limits the heel rocker by forefoot contact [15] and thanks to the flexion of knee the shock is absorbed. In this phase both feet are on contact with the ground and it ends when the opposite limb starts the swing period.

- *Mid stance*
  The period occupies between 10% and 30% of the entire cycle. Here the foot is fully planted on the ground and the body weight is directly on the foot since it is in the first half phase of single support interval. Ankle is dorsiflexed while knee and hip are extended.

- *Terminal stance*
  The body begins to move forward and the heel starts to rise (*heel off*). This phase lasts up to 50% of cycle and it completes the single support interval ending with the *heel strike* of the other foot. Ankle begins to be in plantar-flex, knee is slightly flexed and the extension of the hip puts the limb in a trailing position.

- *Pre-swing*
  It is the last period of the stance phase, and it begins with the *heel strike* by the other foot. Characterized by an increasing knee and ankle flexion, this subphase is also known as *weight release* where the limb is promptly unloaded by the sudden transfer of body weight. It ends when the foot leaves the ground (*Toe off*).

2.1.2 Swing

- *Initial swing*
  Limb is advanced by hip and knee flexion, while the ankle is in part dorsiflexed and it takes place from 60% to 75% of the entire cycle. The other limb is starting the mid stance and when it reaches the flat foot, this phase ends.

- *Mid swing*
  The flexion of the hip allows the limb to advance and the knee to extend in response to the gravity, while the ankle ends to dorsiflex. Opposite to the stance limb, this period ends when the opposite tibia is vertical and the swinging limb is forward.

- *Terminal swing*
  It is the last period of cycle (87%-100%) and the limb advancement is completed with a deceleration and stabilization by the muscles to prepare a new *heel strike.* Hip is in early flexion, ankle is in position of dorsiflexion and knee is extended.

2.2 Joint kinematics

Joint kinematics is the study of relative motion between two consecutive segments of the human body [17]. It involves analysing the various components of joint motion, including direction, range, and quality of movement. In the gait analysis it permits to value how joints work during different activities, so it can be useful to understand normal joint mechanics as well identifying abnormalities or dysfunctions.

Kinematics considers three anatomical planes to describe the joint range of motion: sagittal frontal and transverse planes (*Figure 2.2*) and reporting the joints motion, it is recommended by Standardization and Terminology Committee (STC) of the International Society of Biomechanics (ISB) to adopt the Joint Coordinates System (JCS) provided by Grood and Suntay in 1983 for the knee [18] and for the ankle and the hip joints [19], [20]. For the metatarso-phalangeal joint is introduced a new joint coordinate system.

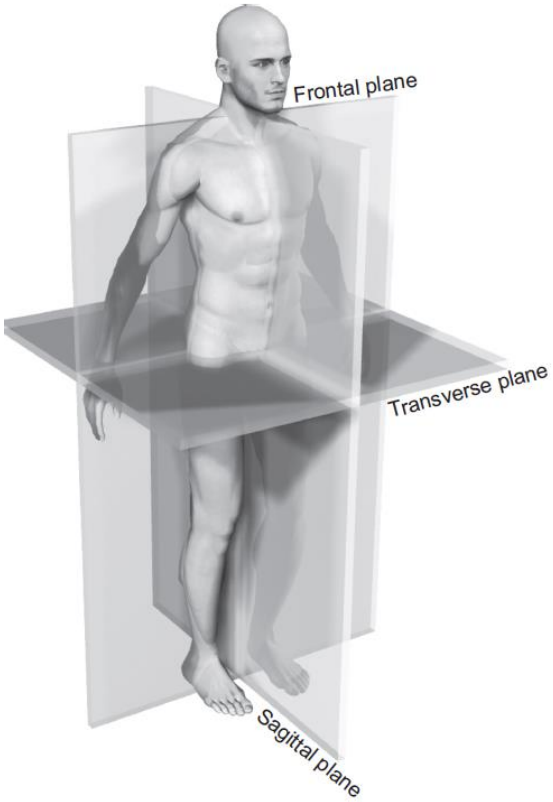

Figure 2.2: *Anatomical planes* [17].

The foot and the ankle constitute a complex and intricate system composed of 28 bones, 33 joints, and 112 ligaments. This system is regulated by 13 extrinsic muscles and 21 intrinsic muscles.
The foot is divided into three main regions: rearfoot, midfoot, and forefoot. Each of these subdivisions performs various functions such as providing balance, supporting body weight, shock absorption and transferring ground reaction forces. In particular:

- Hindfoot is the most posterior part of the foot and it comprises talus and calcaneus (two of seven tarsal bones) combined by subtalar joint.
- Midfoot consists of five tarsal bones and the joint linked with the Hindfoot, is called Chopart's joint.
- Forefooot is the most anterior part of the foot and it includes metatarsal and phalangeal bones and it is linked with Midfoot thanks of the mid-metatarsal joints also known as Lisfranc joint [21].

2.2.1 Ankle Joint

The ankle joint complex consists of two main joints: the talocrural joint and the subtalar joint. The first one is formed between the talus and the tibia/fibula, while the second links the talus with calcaneus. According to Grood and Suntay [18] , the Joint Coordinate System (JCS) is defined starting from two coordinate systems (CCS) consists of two body-fixed axes, e_1 and e_3, and one "floating" axis, e_2, which is mutually perpendicular to them.

- e_1 is fixed to the tibia/fibula, and it coincides with the *z*-axis of tibia/fibula CCS. The rotational movements are the plantarflexion (negative) and dorsiflexion (positive).
- e_3 is coincident with the *y*-axis and fixed to the calcaneus. Rotational movements are the external rotation (negative) and the internal rotation (positive).
- e_2 is the common axis perpendicular to e_1 and e_3. Rotational movements are eversion (negative) and inversion (positive).

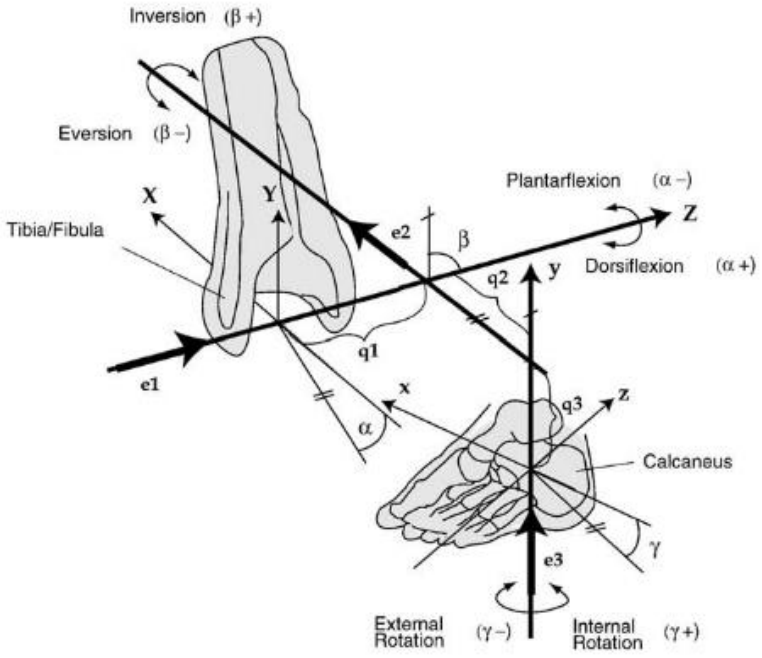

Figure 2.3: *Illustration of the JCS for the right ankle joint* [19], [20].

2.2.2 Metatarsal-phalangeal joint

The metatarsal-phalangeal joint (*MTP*) is the joint formed between the metatarsal bones and the proximal phalanges of the toes. There are five metatarsal joints of in each foot, corresponding to the five toes. These joints allow for flexion and extension movements, facilitating the activities which involve foot movement. The presence of adequate *MTP* joint dorsi-flexion, is essential during the terminal stance and pre swing phases to enable smooth forward progression of the foot [11].

- The rotational movements on the z-axis are the plantarflexion (negative) and dorsiflexion (positive).
- The rotational movements on the y-axis are the external rotation of abduction (negative) and the internal rotation of adduction (positive).
- The rotational movements on the x-axis are eversion (negative) and inversion (positive).

  The centre of coordinates is positioned on the first MTP joint.

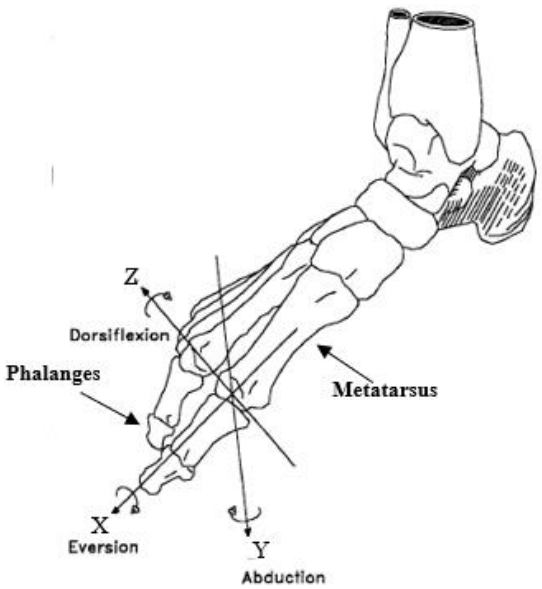

Figure 2.4: *Coordinate system centred on the first MTP joint, anterior-medial view* [22].

3 Motion capture

Human motion capture also known as “human mocap”, is a specialized application of motion capture technology that focuses on capturing and analysing the movements of human body using specialized sensors or markers.

Mocap encompass various approaches based on electro-magnetical, inertial, electro-mechanical, ultrasound or optical systems to acquire the movement. In particular, there are two different ways based on optical systems:

1. Marker-based system (MB)

2. Markerless systems (ML)

Both have their own advantages and limitations: the choice depends on the specific requirements and the motion capture application context.

In the clinical gait analysis, the use of mocap plays a crucial role, providing objective and detailed information about the biomechanics of the gait.

3.1 Marker-based system

Currently, optoelectronic systems are considered the most accurate technology for assessing joint kinematics and it is widely acknowledged as a significant tool in the assessment, planning of therapy, and evaluation of disorders associated with walking patterns [9]. However, it is essential to control the accuracy of the marker placement, in order to ensure the precision of estimation in gait analysis [23]. Moreover, according with the literature, it has been identified that marker-based motion capture systems are prone to errors primarily due to instrumental errors [24] and soft tissue artifacts [25].

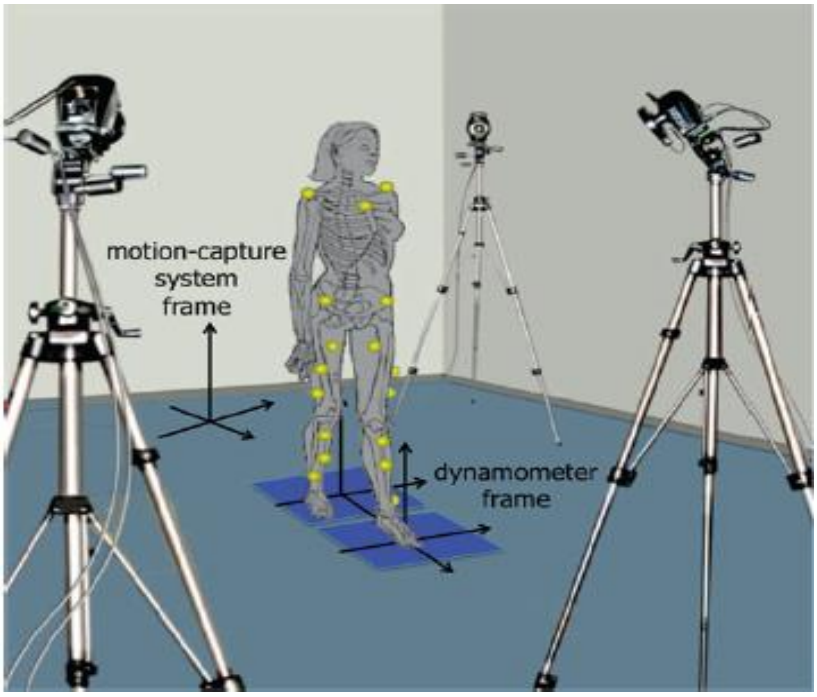

Figure 3.1*: Illustration of a movement analysis laboratory with stereophotogrammetry system, force plates and the associated reference systems*[26].

Three-dimensional (3D) marker-based system uses the principles of stereophotogrammetry and optoelectronic, involving the use of multiple cameras and markers placed on the object or subject of interest. Using image processing techniques, it is possible to detect the instantaneous 3D position of the markers.

Optoelectronic stereophotogrammetry works following these steps:

1) Cameras setup: cameras strategical placement around the capture area.
2) Camera calibration: it is necessary to establish the spatial relationship between the cameras and the subject/object of interest in order to obtain the camera established intrinsic/extrinsic parameters.
3) Markers placement: markers are placed on different landmarks, following specific protocols (*e.g.* Plug in gait) on the subject/object of interest.
4) Data capture: following the subject/object movements, cameras capture images from the markers and consequently from different perspectives.
5) Marker tracking: by the triangulation of the marker positions, thanks to the use of specialized software, it is possible to reconstruct the 3D trajectory of each marker. Moreover, every bone and rotational joints of interest are modelled as different segments[27].
6) Data analysis: the collected data are then used to extract kinematics and temporal-spatial parameters.

In particular, two types of markers are used: active and passive. The active ones (*e.g*. Optotrack) emit their own infrared light (IR) and they are sequentially powered in order to be detected one at time by the cameras.
The passive ones (*e.g*. Vicon) reflect the light emitted by the mocap systems cameras and they receive the light at the same time in order to have all the different perspectives, allowing the triangulation of the markers positions in 3D space. Differently by the actives one, the passive markers have to be labelled in post-processing phase, because of the impossibility to distinguish one from another, but on the other hand they don't need cables nor power supplies and they let the subject have a high range of motion.
Optoelectronic stereophotogrammetry has some issues, like having specialized laboratories (which are generally very expensive), the possibility of limited number of gait cycles, the long patient preparation and the long calibration times. However, this motion capture system, nowadays, is recognized as gold standard in gait analysis.

3.2 Markerless systems

The demand for motion capture methods that are more efficient in terms of time, do not rely on specialized personnel and are less susceptible to errors such as tissue artefact, currently is always increasing.
The markerless motion capture uses a standard video recording to capture movement without the presence of markers to identify body positions and orientation. However, the widespread adoption of markerless motion capture has been hindered by the need to the requirement of advanced coding skills and in-depth computer science knowledge [10].
This technique not only not require specialized environments and operators, but its possibility to reprocess old markerless videos with new pose estimation algorithms, demonstrates that it has the potential to perform movement analysis by reducing data collection and processing time compared to traditional MB systems [10].

Currently, two main ML systems are commonly used in mocap. The first one utilizes standard video cameras, while the second system relies on depth cameras.

Both systems could be implemented using either single camera or multi-cameras for data capture.

Markerless motion capture with a standard camera have same some limitations comparable to MB system, such as the capture volume is restricted by the number of the cameras, but differently from MB approach, this system doesn't suffer the sunlight or the interference of other systems that work simultaneously. However, the extraction of considerable information form recorded images using software can be a challenging task that requires a high level of accuracy [10].
One of example of multi-cameras markerless system is developed by Theia Markerless. Thank of the use of multiple cameras synchronized, the system allows to obtain the 3D body segment position and orientation. If on one hand, the system results have more practical advantages than MB systems and it has the potential to facilitate quicker and simpler acquisition of gait kinematics and in wide range of environments [28], in the other hand, to capture a space of large dimensions, this system requires a large number of cameras.
The RGB-Depth cameras have the ability to combine the standard video information and the depth measurements of the scene recorded from the device and in opposite MB systems, they are more economic and accessible.

Three different principles of operation of the depth sensor are implemented in RGB-Depth cameras.

- Stereoscopic vision: the 3D reconstruction is based on the presence of two cameras. Like the binocular vision of human system, it consists in the detection of the common points, in the two images acquired by the respective cameras, that refer to the same point of the object. Knowing the intrinsic and extrinsic parameters of the cameras, it is possible obtain the position of object in the 3D space.
- Fixed structured light: an infra-red light is projected on a known pattern made of grids or other distinctive features. The camera acquires the scene with the pattern included. Comparing the emitted pattern and the captured scene with the pattern, allows to extract, thanks of the triangulation process, the distance of various points in the scene from the camera.
- Time of flight (ToF): the measurement of the time of flight (the time of the IR light takes to travel to the target and back to the sensor), permit to determine the distance of the points from the camera. ToF technology provides a cost-effective a mechanically compact solution for depth imaging. It offers the advantage of being unaffected by change in environment illumination, making it reliable in various lighting conditions and it simplifies the process of figure-ground separation [29]. The most recent RGB-Depth cameras, such as Microsoft Kinect,

Intel Real Sense and Orbecc are equipped with an integrated depth sensor that is able to reconstruct the depth image using this specific principle.

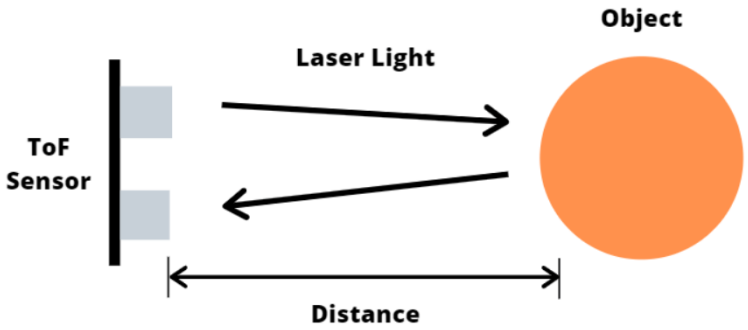

Figure 3.2: *Illustration of ToF principle* [30].

After collecting the video data, the next step of markerless system is the processing the data with a software to detect and extract the joint position. The pose estimation algorithms commonly employ deep learning techniques. These algorithms are trained through a massive dataset that consists of numerous of manually labelled key points [31]. These posed estimation algorithms work with mathematical calculation on each image using Convolutional Neural Network (CNN), that consists of multiple layers, with the input of one layer is the output of the previous one.

This network allows to the algorithm to learn how extract joint centres identifying the features which characterize the body segments.

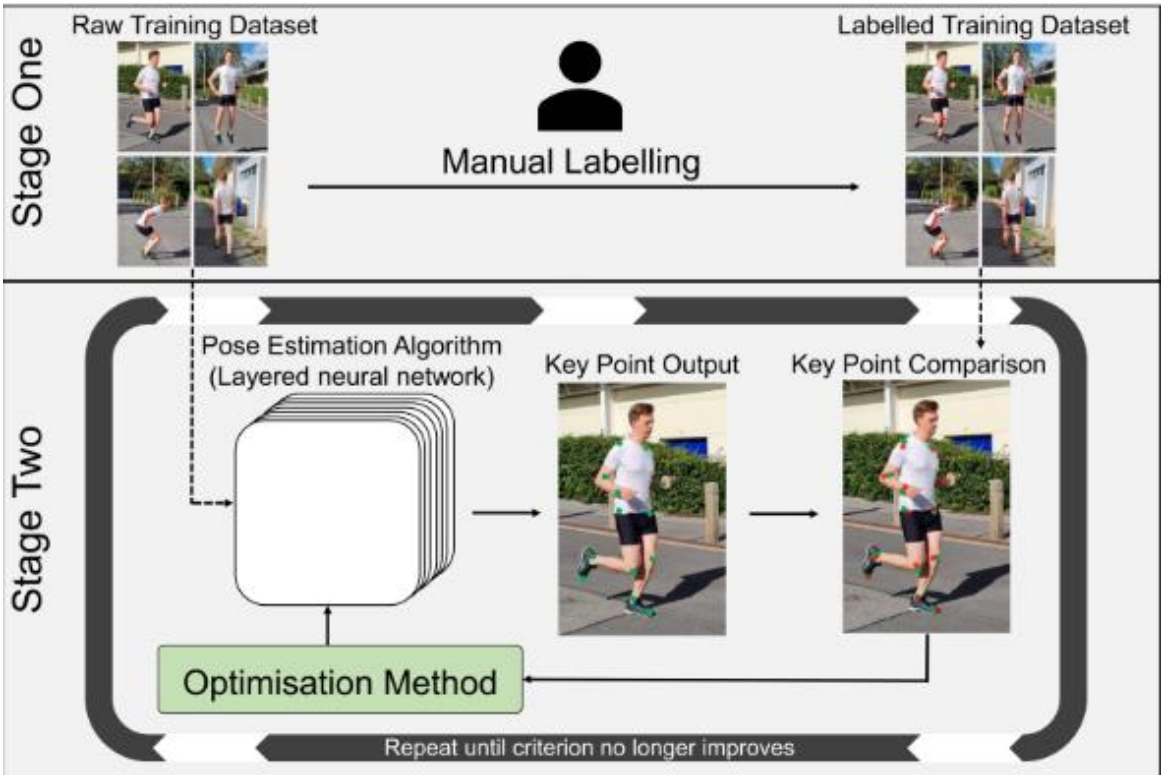

Figure 3.3: *Pose estimation algorithm workflow. Stage One: manually labelled training dataset. Stage Two: pose estimation algorithm* [10].

One of the most successful ML systems employed with deep learning is Openpose, an open-source computer vision library and software that provides real time multi-person key point detection and pose estimation [32]. One of the key features of OpenPose is its real time performance, allowing for live pose estimation in video streams. In additions, it can detect and estimate the 2D positions od body joints for multiple people in image or video, making it a powerful tool for applications in human-computer applications [32]. In addition to obtaining 2D joint coordinates, OpenPose has developed a system of multiple synchronized cameras to obtain 3D joint positions.

Figure 3.4: *Multi-person pose estimation. OpenPose* [32].

Microsoft offers the Azure Kinect body tracking software development kit (SDK), which enables real time tracking of the position of multiple individuals. It is capable of extract e and visualize the segmentation map and the skeleton of each individual within the camera's field of view[33]. The skeleton counts 32 body joints (*Figure 3.5*), but only 9 represents the lower limb.

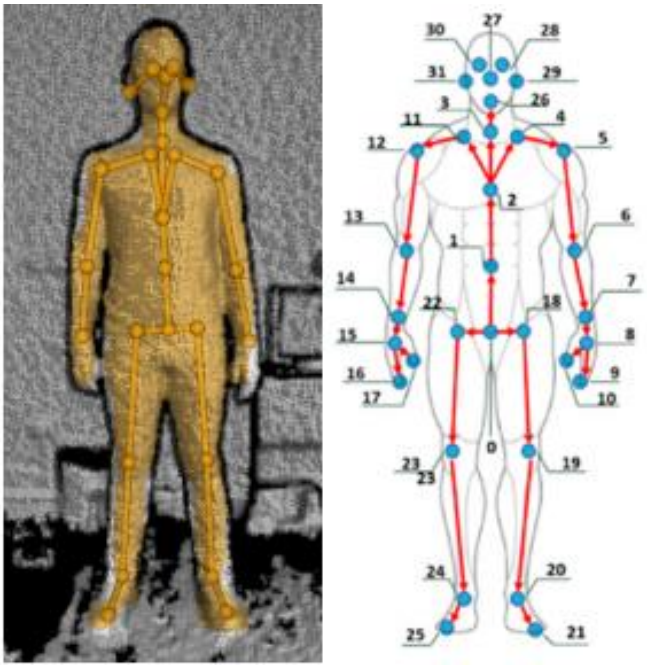

Figure 3.5: *Real time skeleton and segmentation map (left) and the 32 joints identified by Azure Kinect SDK [21].*

Using the IR image acquired, the SDK extract the body segmentation map and the 2D body skeleton through a CNN algorithm. It was trained with on real images and vast dataset of synthetic images generated from videos featuring artificially

simulated human bodies performing different movements. Then, meshing the information obtained by the depth sensor, the 3D joint coordinates and orientation are computed.
As mentioned above, the extraction of the joint coordinates using a Convolutional Neural Network, requires a large training set. However, in the clinical application of gait analysis, obtaining a massive dataset can be challenging, especially in the pathological cases where access to a large amount of image data is limited. So, it become necessary to address the pathological gait analysis with a subject-specific approach.
Castelli A. et al. [9], Pantzar-Castilia E. et al. [1] and [13] have implemented 2D markerless technique to perform the lower limb sagittal plane kinematic analysis using a single video camera. The methods proposed, involved extracting 2D joint positions through multisegmental model of the lower limb, which was calibrated in static image form anatomical landmarks manually identified by operator [9].

In all proposed methods listed with markerless approach, the feet are represented as a single segment with an only joint identified. Van den Herrewegen I. et al. [12] proposed a 4-segment foot model applying a D3DScanning focusing on the kinematics aspect of foot motion. The system consists in 5 scanner units (ViALUX) positioned around and below a glass force platform. The segments are manually selected on a static scan, which allows to acquire the whole foot surface also in the dynamic situations, representing it as point cloud.
The algorithm proposed aligns the dynamic segments, matching them with the static scan through the Iterative Closest Point algorithm (ICP). It allows to define the segment position in the dynamic frame as the one that minimize the root mean square (RMS) between the transformed segment and dynamic frame. Then the calculation of kinematics is obtained with the inverse multiplication of the segment transformation matrices.

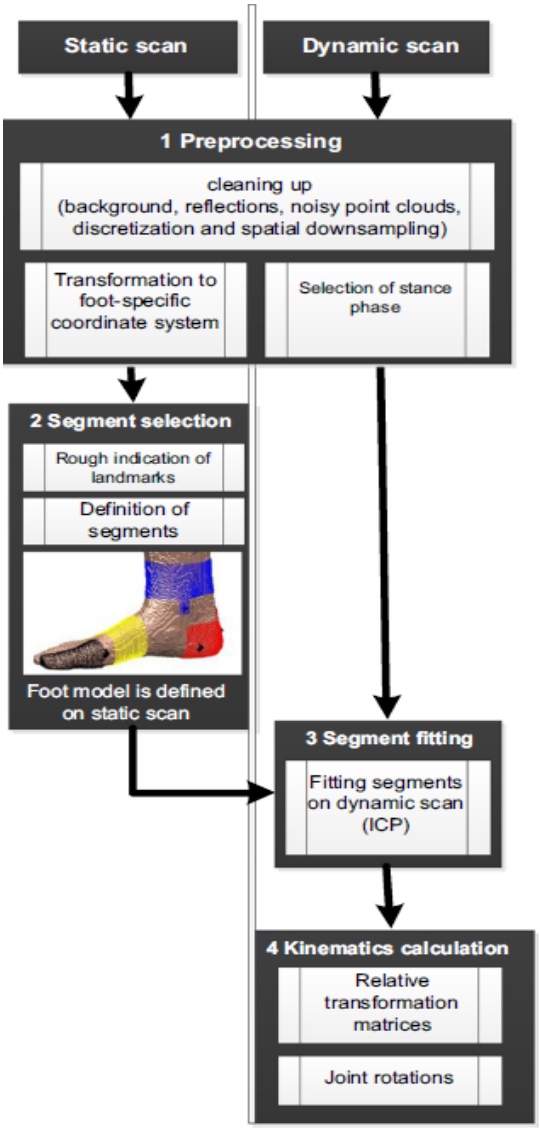

***Figure 3.6**: Workflow to calculate the kinematics form the 3D scan proposed by Van den Herrewegen I. et al* [12].

4 Methods

As mentioned in introduction, the aim of thesis project is to design a markerless method based on a single RGB-Depth camera for estimating sagittal ankle and metatarso-phalangeal kinematics using a new two-segment 3D foot model. In particular, the method proposes to expand the 2D subject-specific foot model developed by Balta et al., 2020 [13] to 3D subject-specific foot model composed by two-segment: Mid-Rear-foot and Forefoot.
The proposed protocol consists in five parts, as illustrated in the block diagram in *Figure 4.1*.

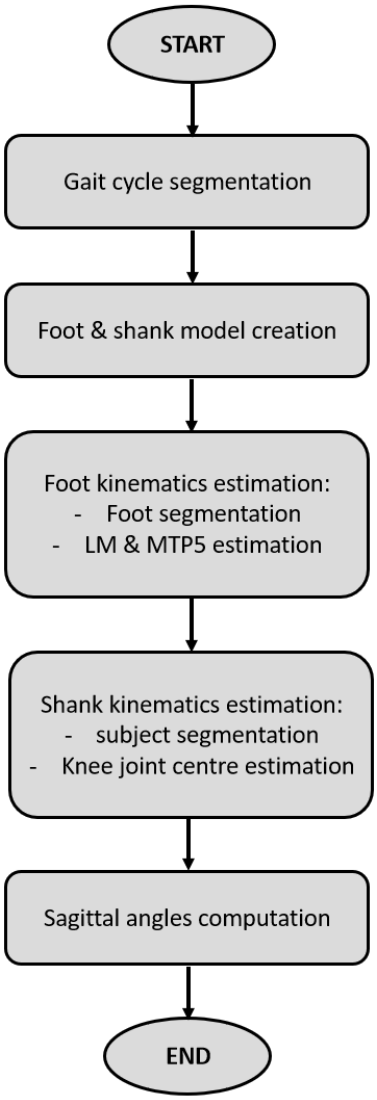

Figure 4.1: *General overview of the of the proposed method for estimating the sagittal shank and foot kinematics.*

4.1 Gait cycle segmentation

The first step of the proposed method is to identify the most central gait cycle between two consecutive initial foot contacts. To identify two consecutive foot contacts with the ground is applied the method based on a 2D clinical gait analysis protocol reported by Balta et al. 2020 [13].

In phase of acquisition, the subject was asked to wear red and blue socks to make the identification of the feet in RGB images easier. The feet are discriminated by applying two coloured filters: red for the right foot and blue for the left one [34].

Figure 4.2: *Color filters applied on the feet.*

Then, heel and toe are automatically identified, by assuming that heel point is determined as the intersection between the segmented foot sole's best-fitting line and the posterior portion of the foot, represented by a vertical segment (*Figure 4.3*). The toe, known a priori, represents the farthest point in the direction of progression.

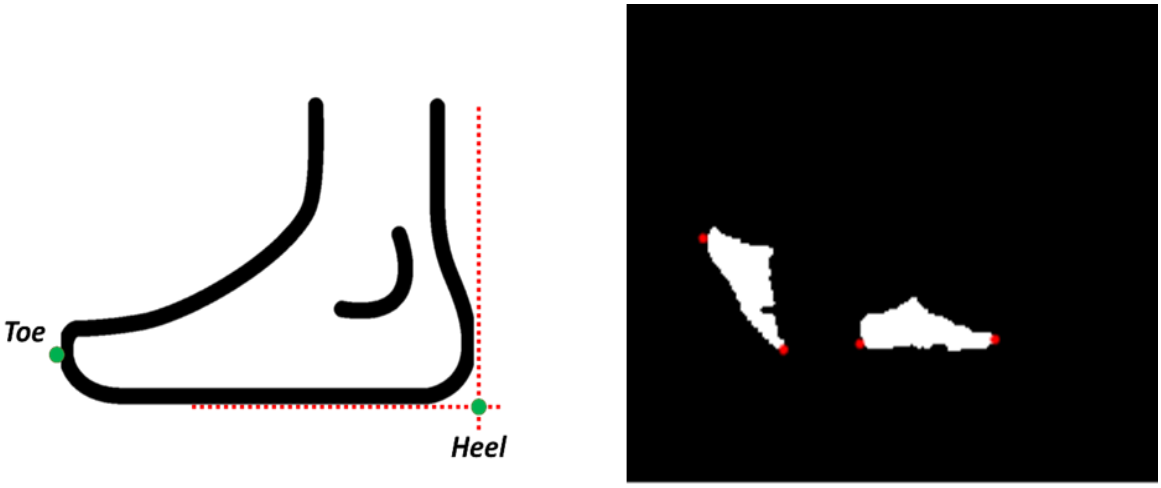

Figure 4.3: *Identification of the heel and toe. On the left there is the sole of the foot and the posterior portion of the foot (dotted red lines). On the right there is an example of the identification of heel and toe on both feet* [13].

After that, the anterior-posterior and vertical coordinates of heel and toe are extracted. Since it is not possible to know a priori how the subject strikes the ground, the point and the time of contact with the ground, is determined as the foot point where both vertical and anterior-posterior coordinates, expressed in pixels, first reach a stationary condition. To evaluate this, the first derivative of heel and toe coordinates (vertical and anterior-posterior) was calculated and the foot extremities were considered as points of impact when the time derivatives variation of both coordinates were lower than a threshold given a priori (± 3 pixels) *Figure 4.4*.

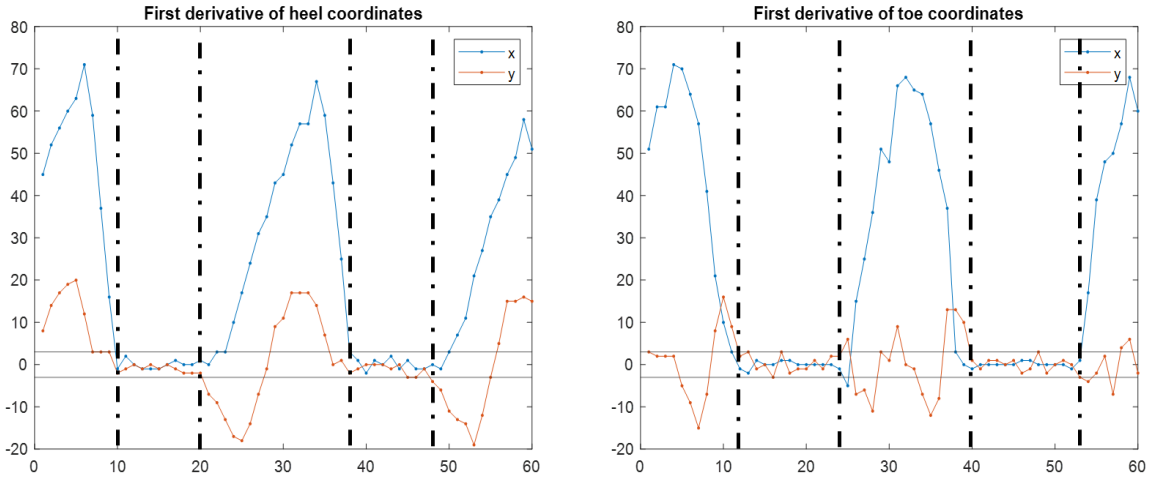

Figure 4.4: *First derivative on heel and toe coordinates. The black dotted lines represent the intervals in which heel and toe are in contact with the ground. In this figure there are two intervals for each extremity. The first interval represents the initial contact* [13].

To avoid the optical distortions and the parallax effect of the camera, the most central gait cycle is selected among the gait cycles inside each dynamic acquisition.

4.2 Foot and shank model creation

4.2.1 3D foot model

The primary objective stated in the thesis is the introduction of a new segment in multi-segment model proposed by Balta et al. 2020 [13]. In particular, the creation of a two-segments foot model can lead to obtain a better information about the foot kinematics. In order to achieve this, the three-dimensional (3D) foot model is created by capturing four views of the foot in a static position: frontal (*FRO*), posterior (*POS*), lateral (*LAT*), and medial (*MED*). Each view captures a different perspective of the foot and, by employing alignment techniques, it becomes possible to create the 3D foot model (*3FM*).

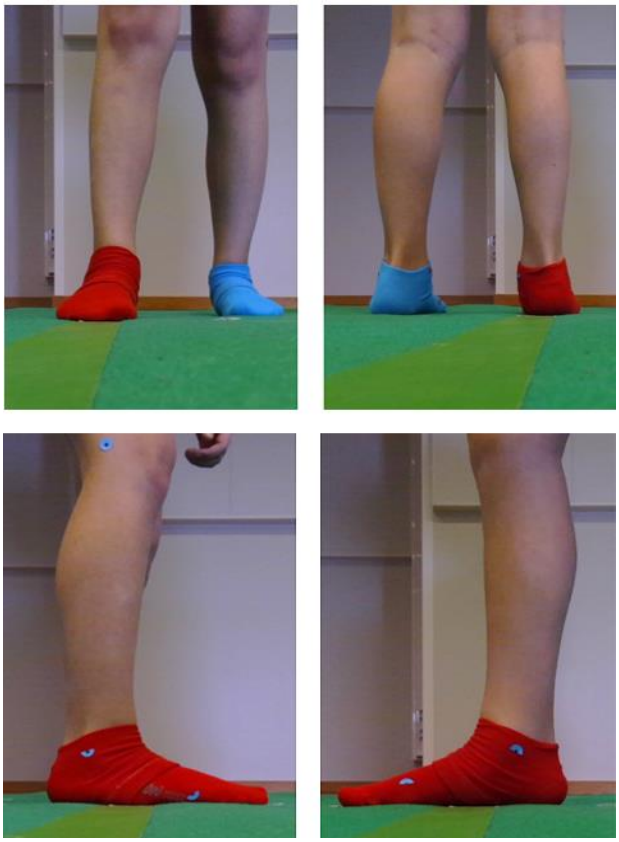

Figure 4.5: *The four views (FRO, POS, LAT, MED)*

The process of creating 3D foot model involves an acquisition protocol, that aims to achieve a good balance between accuracy of the model and the comfort of the patient. This protocol is designed to minimize the discomfort of the patient may feel in the upright static position during the acquisition. The process involves in a step-by-step algorithm divided into 8 stages. The details of each step are provided below:

1) **Point cloud creation**: for each view a 3D point cloud representing a view of the foot is generated using the depth data by exploiting the RGB-Depth information.

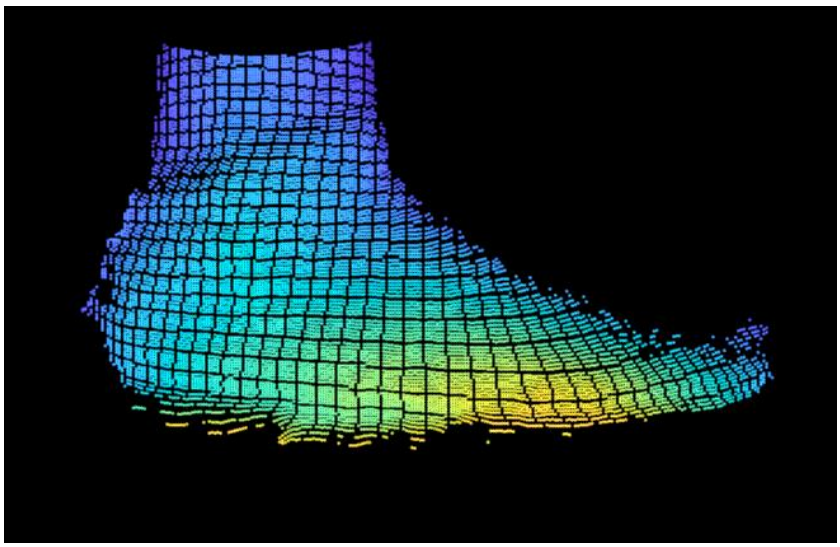

***Figure 4.6**: Point cloud of lateral view of right foot.*

2) **Ground removal:** as shown in the figure above (*Figure 4.6*), the foot point cloud could be affected by errors. These inaccuracies may be caused by the relevance from the depth sensor of the ground (green carpet) or the presence of the shadows due a different weight load on the feet during the static position. It is crucial to remove them since their presence can introduce problems during the merging of different views. For this reason, it is applied an a-priori-chosen threshold based on the percentage of height (e.g. 0.08%) that allows the point cloud to maintain the anatomical information of the foot.

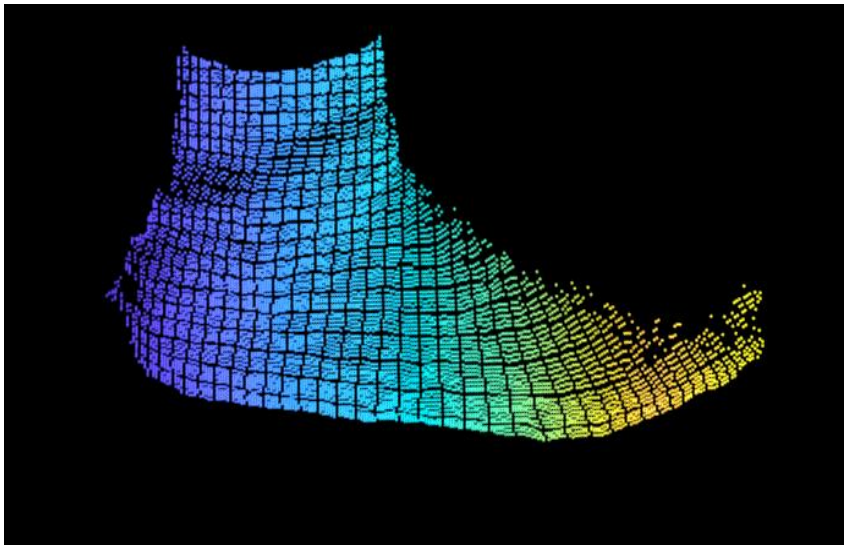

Figure 4.7: *Point cloud after remove the ground layer.*

3) **Identification of lower layer:** first of all, the sole is identified as a plane characterized by the lowest vertical coordinates. To merge the four views, it is essential identify common points in the sole of each view.

4) **Identification of the common points:** for each view there are specific points to be identified.
    - The most anterior point (*ZF*) and the most lateral point (*XFL*) are identified in *FRO*.

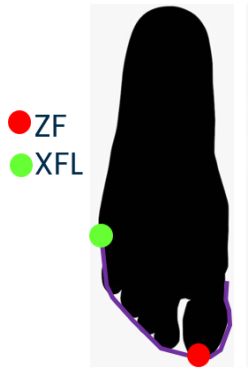

Figure 4.8: *Frontal view (purple line) and the identified points.*

    - The most anterior point (*ZL*), the most lateral (*XL*) in the range between *ZL* and the distance in terms of depth between *ZF* and *XL* and the most posterior point (*ZPL*) are identified in *LAT*.

Figure 4.9: *Lateral view (blue line) and the identified points.*

- The most anterior (*ZM*) and the most posterior point (*ZPM*) are identified in *MED*.

Figure 4.10: *Medial view (orange line) and the identified points.*

- The most distal point (*ZP*) is identified in *POS*.

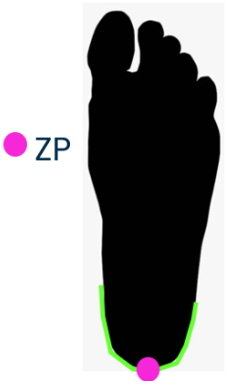

Figure 4.11*: Posterior view (green line) and the identified point.*

5) **Alignment of views:** Using *ZF* of *FRO* as origin, the other three views are aligned following this order:

   - *LAT* is aligned to *FRO* using *XL* as the common point.
   - *MED* is aligned to *FRO* using *ZF* as the common point.
   - *POS* is aligned to *LAT* using *ZP* as the common point.

In particular, the choice to use *XL* as common point between *LAT* and *FRO* is forced since from the lateral view it is not always possible to identify the same most anterior point seen by *FRO*, due the different morphologies of the toes.

Thanks to these alignments, a 3D foot model was created preserving the anatomical features of foot (*Figure 4.12*).

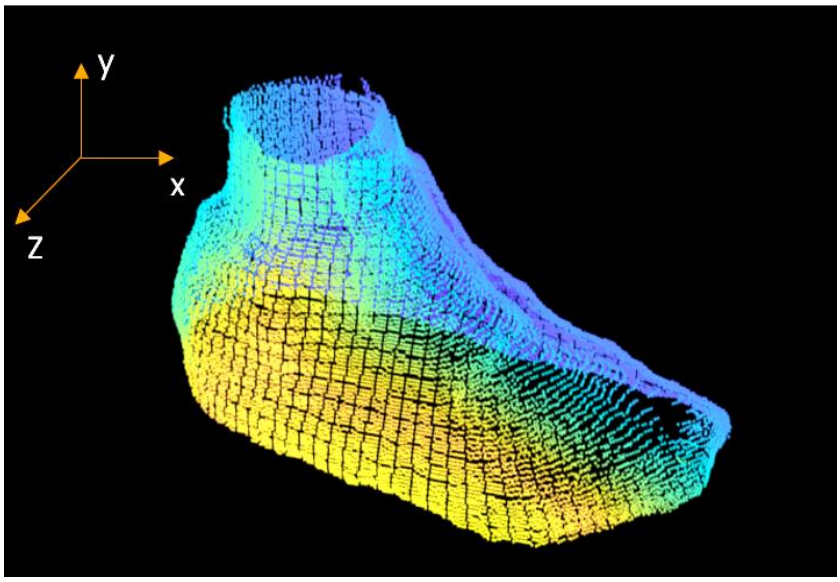

Figure 4.12: *3D foot model (3FM).*

6) **Template calibration:** before acquiring the subject, external anatomical landmarks (*LM*: lateral malleolus; *MTP5*: 5th metatarsophalangeal joint and TOE: the most anterior point on horizontal coordinate on the surface of foot dorsum) have been identified by palpation and marked with a black pen on small colored labels, as it shown in *Figure 4.13*. In order to indicate the positions of landmarks, they are manually identified by clicking on the corresponding locations on the RGB-D images.

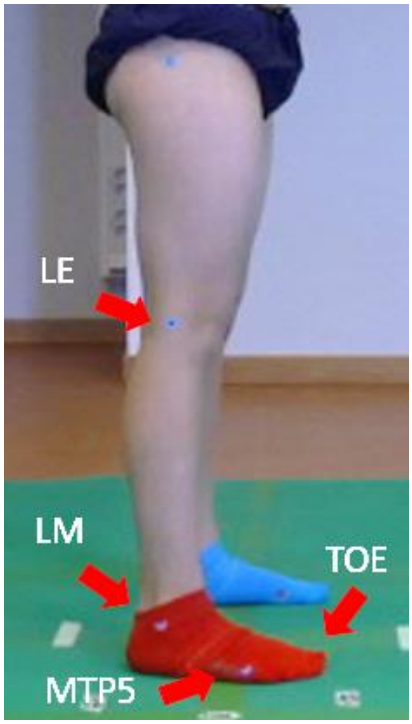

***Figure 4.13**: Illustration of external landmarks (blue labels) applied on the right side of the subject.*

7) **Joints identification:** To obtain the manual identification of landmarks from 2D RGB image and to convert them to corresponding 3D points, it is necessary to apply a conversion factor scale. This helps to establish the relationship between the coordinates in 2D image space and the corresponding 3D space, allowing to correctly map the landmarks on the 3D foot model.
The conversion is defined as:

$$x(p) = \left(\frac{(col(p) - 640)}{k}\right) * abs(d(p));$$

$$y(p) = \left(\frac{(row(p) - 360)}{k}\right) * abs(d(p)); \quad (4.1)$$

$$z(p) = -d(p);$$

Where $x(p)$ and $y(p)$ are vertical and horizontal coordinates respectively, $col(p)$ and $row(p)$ denote the column and row indices of 2D image. The values 640 and 360 correspond to the half width and half height of 2D image, k is the conversion factor from 2D to 3D, $d(p)$ refers to the depth value of depth map with the 2D coordinates of the identified landmark (row(p), col(p)).

Finally, the TOE is identified as the most distal point of the foot at the same coordinate *y* of *MTP5*.

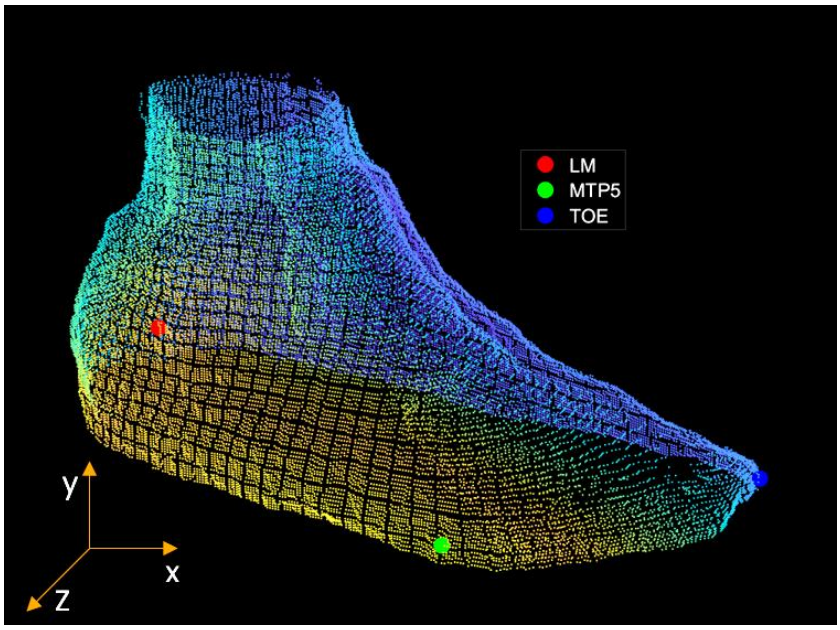

***Figure 4.14**: 3D foot model (3FM) and the identified joint centres.*

8) **Foot model splitting:** the foot model is split, at the coordinate X of *MTP5*, in two segments: Mid-Rear-foot (*MRF*) and Forefoot (*FF*). The cut percentage was saved for applying the same method on dynamic frames.

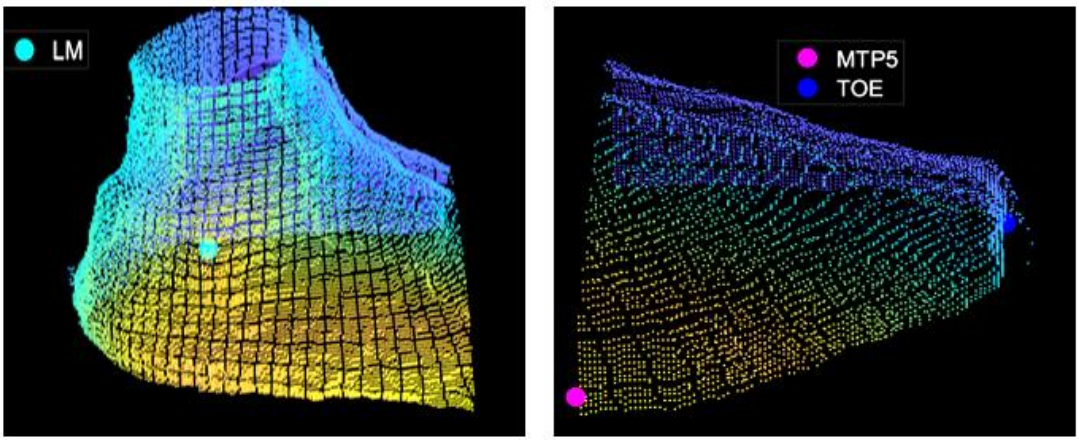

***Figure 4.15:** MRF segment (left) and FF segment (right) with identified joint centres.*

$$Template_{mid-rear\ foot}(x,y,z) = \begin{cases} Template_{foot}(x,y,z), & x_{CA} < x < x_{MTP5} \\ 0, & otherwise \end{cases} \quad (4.2)$$

$$Template_{forefoot}(x,y,z) = \begin{cases} Template_{foot}(x,y,z), & x_{MTP5} < x < x_{TOE} \\ 0, & otherwise \end{cases} \quad (4.3)$$

Where x_{CA} is the x-coordinate of the most posterior point of the foot.

4.2.2 Shank model

To estimate the ankle kinematics on the sagittal plane, the markerless method proposed by Balta et al., 2020 was implemented as follows:

- **Multi-segmental model definition:** a subject-specific lower limb model that consists in four body segments (foot, shank, thigh and pelvis). In this work, only the shank segment will be used, which is connected to the close segments through ankle and knee. The two joints allow for one rotational degree of freedom (*DoF*) and they are centred with the lateral malleolus (*LM*) and on lateral epicondyle (*LE*).
  The shank template and the relative joints coordinates systems were calibrated on the standing acquisition (*Figure 4.16*) by manually selecting the anatomical landmarks to obtain the position vector in the image system ($^{I}LM_{0'}\,c$). In order to accommodate the potential right/left asymmetries, the specific-subject model is defined for both sides.

- **Template calibration:** as explained in the step 6 of the foot model creation, the knee joint is manually identified by expert operator through palpation and marked with a black dot on a small and coloured label in correspondence of the lateral epicondyle (*LE*).

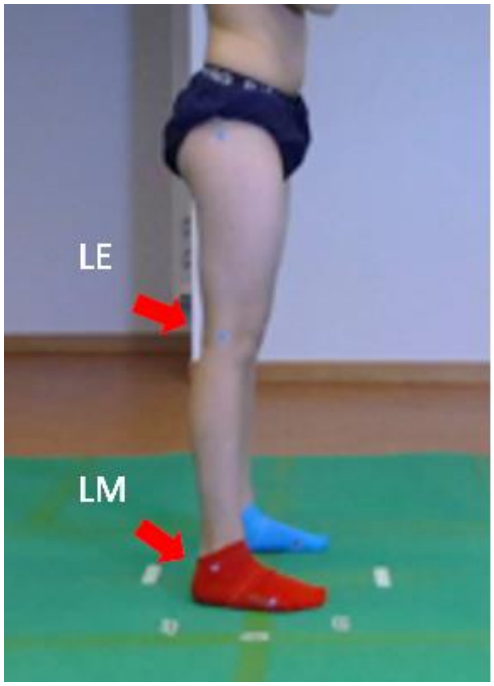

***Figure 4.16**: Sagittal static upright standing acquisition and the 2 landmarks that defined the shank segments (LE and LM), identified by blue labels.*

- **Shank template:** the central shank portion was obtained by extracting the region encompassed within the annulus centred at ${}^{I}LM_0$ with radii equal to 25% and 75% of the Euclidean distance between ${}^{I}LM_0$ and ${}^{I}LM_0$, (L_Shank25 and L_Shank75, respectively). (*Figure 4.17*). Then, the template ${}^{I}Template_{shank}$ and its generic pixel ${}^{I}Template_{shank}(x, y)$ in the image I is obtained as follows:

$$ {}^{I}Template_{shank}(x,y) = \begin{cases} 1, & {}^{I}Mask_{subject}(x,y) = 1 \; and \; L_{Shank25} < x^2 + y^2 < L_{Shank75} \\ 0, & otherwise \end{cases} \quad (4.4) $$

Then ${}^{I}Template_{shank}$ is fitted inside an ellipse and the principal axes using its centroid was calculated.
The shank coordinate system s_0 is defined with its axes coinciding to ellipse inertial principal axes and the transformation matrix is computed.

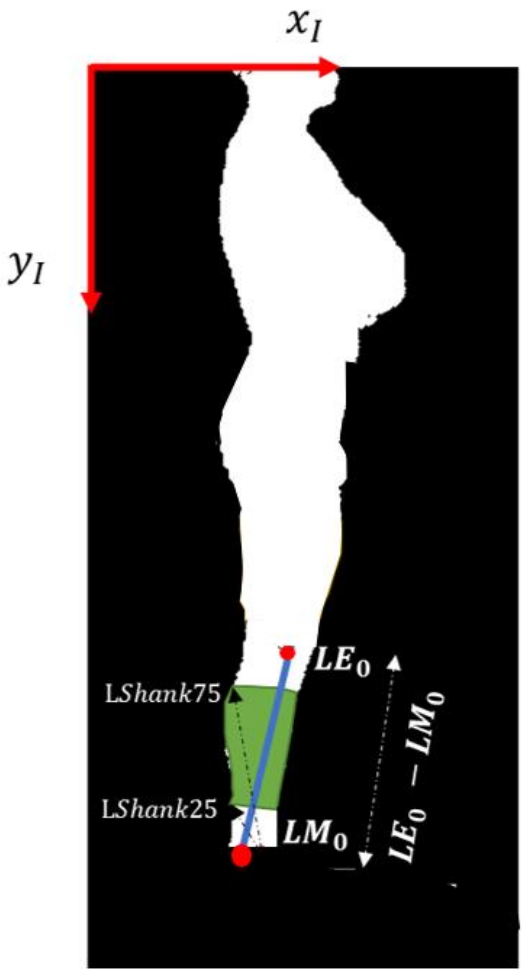

***Figure 4.17**: Lower-limb model for the right side by Balta et al.* [13].

4.3 Foot kinematics estimation

The two main steps to extrapolate the sagittal foot kinematics:

- Foot segmentation
- LM and MTP5 estimation

Each of these two steps are subdivide into other several sub-steps.

4.3.1 Foot segmentation

The proposed foot segmentation follows the steps outlines below:

1) **Segmentation of 2D mask of the foreground foot**: as reported in the previous paragraph, two color filters are implemented to identify the foot in the foreground. When the subject walks from the left to the right, it is applied the red color filter, since the right foot wears the red sock. Conversely, when the walking is in the opposite direction, the color filter detects the blue sock.
   A color filter is based on the RGB values of the pixels, so its effectiveness depends on the lighting conditions of the environment. The choice of the three RGB thresholds is very important to extract the full 2D mask of the foot. The presence of some shadows, or the presence of labels could cause some holes in the mask. For this reason, it is applied an algorithm based on morphological reconstruction to fill all holes in mask.[35]

Figure 4.18: *2D foot mask with holes (left) and filled 2D foot mask (right).*

2) **Identification of Depth mask of the foreground foot**: using the 2D foot mask, it becomes feasible to extract the depth region corresponding to the foot from the full body of subject in the depth image. The depth sensor of Azure Kinect is affected incorrect depth map, when the object captured is in movement, as the foot shown in *Figure 4.19.* For this reason, to avoid a loss of anatomical information, it is crucial restore the missed depth points, in order to improve the estimation of the joint centres position.

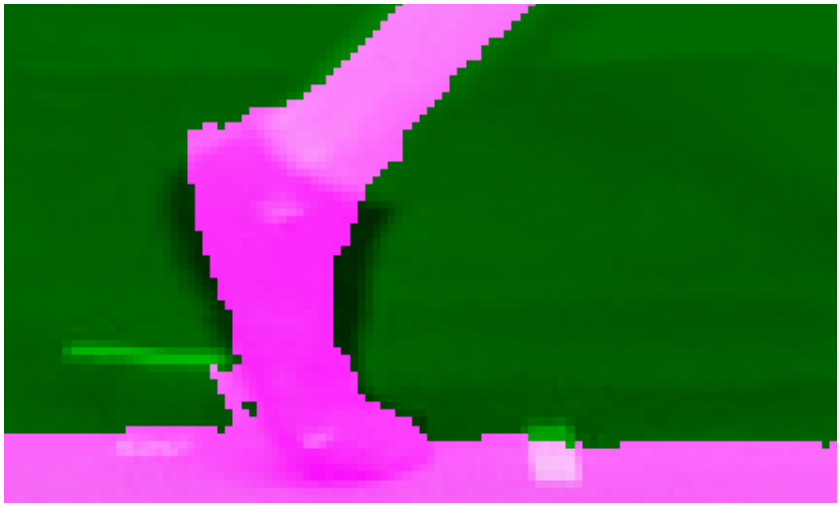

Figure 4.19: *Depth image (pink) overlapped on RGB image (green).*

3) **Depth completion:** Due to the significant portion of the highest moving frames of the gait cycle (pre and swing phase), which may cause an incorrect depth mask, it becomes essential to restore the missed depth points. The developed algorithm suggests a Depth completion to try to solve the problem.
Depth completion in RGB-Depth cameras is a technique that aims to recover dense depth maps from sparse depth measurements [36], since the depth cameras often encounter difficulties in shiny, bright, transparent and distant surfaces [37]. Several approaches can be found in literature: inpainting-based algorithms that leverage the correlation between depth and color information to estimate the missing points [38] or learning- based methods that employ deep learnings models, such as convolutional neural networks (CNNs) or generative adversarial networks (GANs) [36], which learn the mapping between complete depth maps and the incomplete ones.
The purposed method, in order to recover the missing depth points, applies a depth completion using kernel: a small matrix used to perform operations on the neighbouring pixels on the depth map. The process is described in the three following steps:

   A) Selection of target regions:
   To select the missed regions of Depth map, the 2D foot mask is overlapped on the Depth map, as shown in *Figure 4.20*.

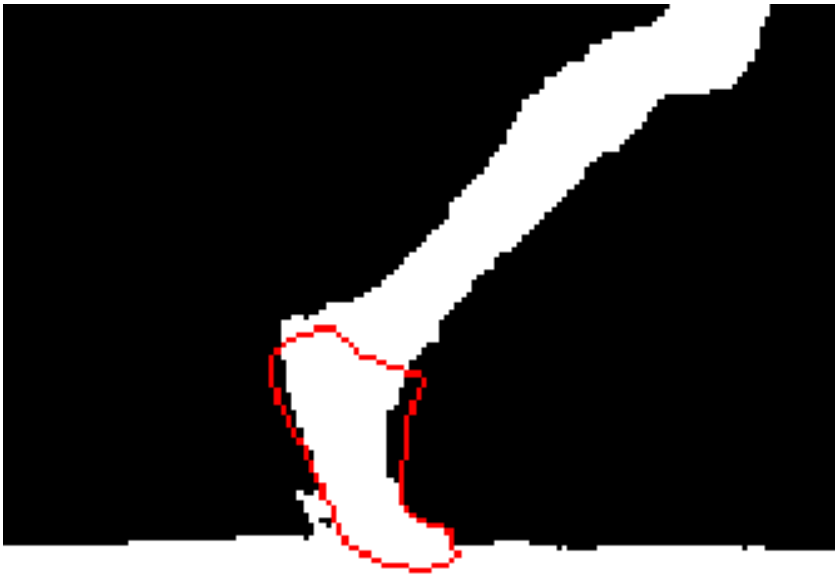

Figure 4.20: *Depth map (white) and 2D foot mask (red line).*

B) *Kernel design:* a 5x5 low-pass kernel is selected. For each pixel in the missed region, the mean value of a 5x5 neighbourhood was computed as reported in (4.5)

$$g(x,y) = w * f(x,y) = \sum_{dx=-a}^{a} \sum_{dy=-b}^{b} w(dx,dy) f(x-dx, y-dy) \qquad (4.5)$$

$$f(x-dx) \neq NaN,\ f(y-dy) \neq Na$$

Where $g(x,y)$ is reconstructed depth value, $f(x,y)$is the original depth value and w is the kernel.

0.04	0.04	0.04	0.04	0.04
0.04	0.04	0.04	0.04	0.04
0.04	0.04	0.04	0.04	0.04
0.04	0.04	0.04	0.04	0.04
0.04	0.04	0.04	0.04	0.04

Figure 4.21: *5x5 mean value kernel.*

C) *Starting point:* it is possible to observe that the majority of the missing points is located in the anterior part of the foot since this part is characterized by the highest speed. Therefore, if the walking direction is towards to right, the kernel starts to slide over the depth map from the top left corner. This allows the kernel to traverse over the existing depth points and then on the missing ones, conversely in the opposite direction.

After the application of kernel, the Depth map is 'filled' on the missing values as shown in the figure below.

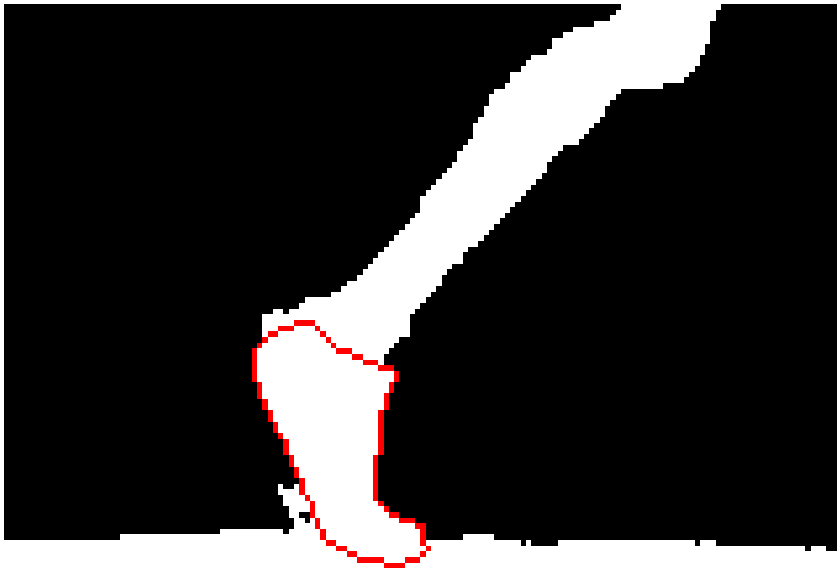

Figure 4.22: *Depth map after depth completion (white) and 2D foot mask (red line).*

Through the reconstructed depth values, the final point cloud of the foot was obtained.

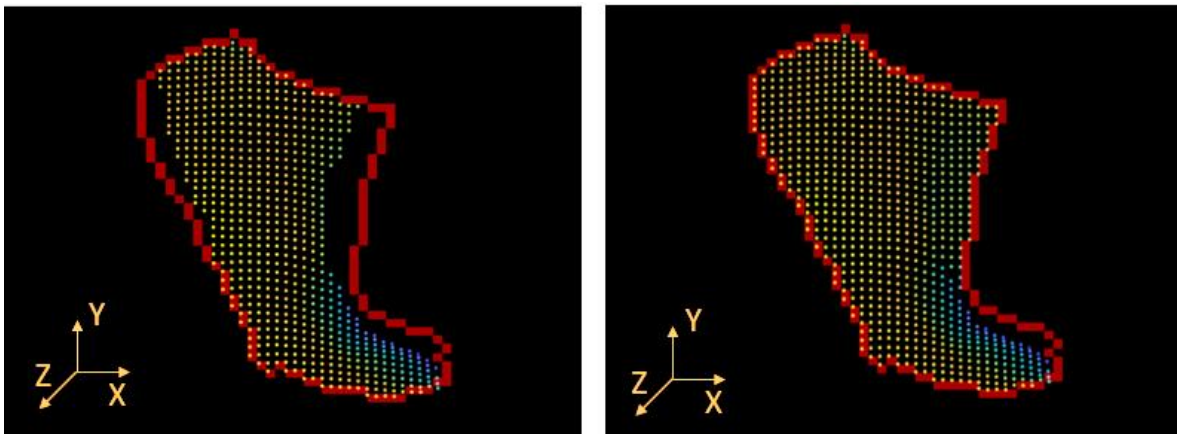

Figure 4.23: *3D foot point cloud during gait cycle before (left) and after (right) depth completion, with the boundary identified by 2D foot mask (red line).*

4.3.2 Iterative Closest Points (ICP)

Iterative Closest Points is an iterative method to registrate 3-dimensional or 2-dimensional shapes. The main goal of ICP is to find the best geometric transformation (translation and rotation), which aligns a generic dataset (P) to model (X). The algorithm works through a series of iterations to minimize the distance between the points of P and X. The main steps concerning in the algorithm include the following:

1. **Computation of closest points**: for each point in the model X, the algorithm searches the corresponding point in the dataset P which has the minimum distance:

$$d(p, X) = min_{x \in X} \; ||x - p|| \tag{4.6}$$

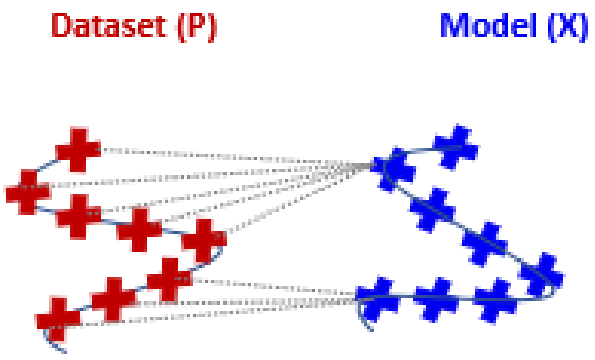

Figure 4.24*: The closest points of the model to the dataset.*

As a result, a set of closest points is obtained (Y).

2. **Computation of the transformation matrix**: the algorithm computes an optimal transformation that aligns the closest points, estimating the roto-translation matrix that allows to minimize the least mean square distance between the closest point.

$$d_k(R,T) = \min |\frac{1}{N_p}\sum_{i=1}^{Np} ||\overrightarrow{y_i} - T(\overrightarrow{p_i})||^2| \tag{4.7}$$

   Where $[T = \begin{matrix} R & \frac{t}{1} \\ 0\,0\,0 & \end{matrix}]$ is the transformation matrix.

3. **Termination of the iteration**: the algorithm continues iterating through the first two steps until the convergence is achieved, $d_k(R,T)$ is below a threshold (chosen a priori) and the iteration finishes.

$$d_k - d_{k+1} < \tau \tag{4.8}$$

4.3.3 Lateral malleolus and fifth metatarso-phalangeal joint estimation

To estimate the sagittal kinematics of the foot, it is necessary to identify the joint centres on each dynamic frame of the gait cycle. In order to do it, it is fundamental to apply the ICP technique between the dynamic point clouds *Dyn* and the 3D foot model point cloud. Anyway, to implement the ICP you need to develop different steps to pre-align the point clouds, since the pre-alignment plays a crucial role to improve the efficiency and accuracy of this technique [14].

The steps to estimate the joint centres are the following:

1. **Foot splitting**: as well as in the creation the 3D foot model in paragraph 4.2.1, two different segments are obtained (*MRF* and *FF*) from the split of the model on the *MPT5*, also *Dyn* point cloud is split into two segments by considering the same percentage.
   Since each *Dyn* point cloud has a different shape and different axis orientation, to find the perfect point of separation it is fundamental to align, in the best possible way to the sagittal axis (X) of *3FM*.
   Using the information of the 2D foot mask's inertia (a measure of the mass distribution of the object around its centroid), it is possible to determine its

orientation as the angle between the x-axis of the image and the principal axis of the object's inertia.
After rotating *Dyn* at the angle output by the function (*Figure 4.25.*), it is possible to split it in two segments (*MRFd* and *FFd*), *as* shown in *Figure 4.26*.
This work proposes to use the same relative lengths of *MRF* and *FF,* which here are standardised, for every dynamic foot shape in every gait cycle phase.

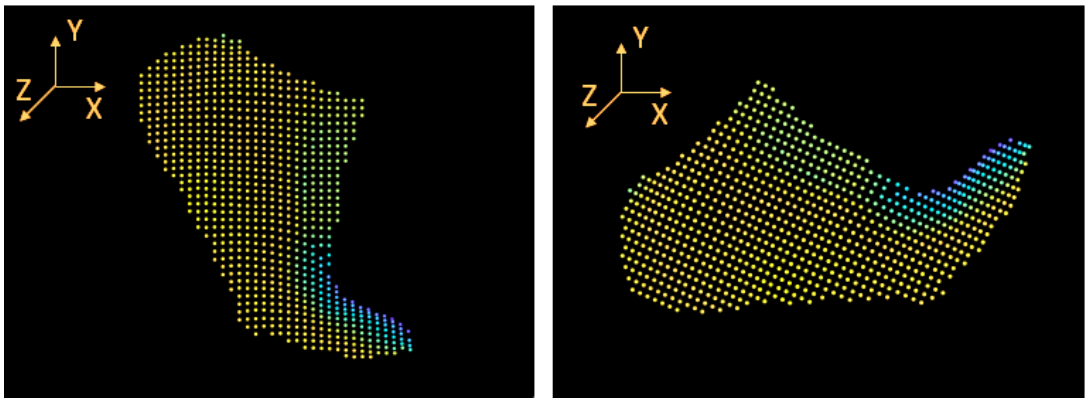

Figure 4.25: *Original dynamic foot point cloud (left), rotated dynamic foot point cloud (right).*

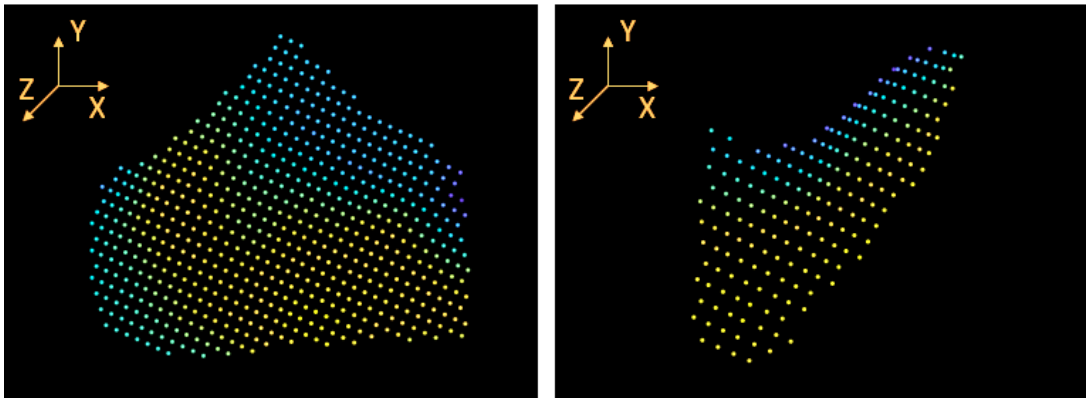

Figure 4.26: *MRFd (left) and FFd (right) after the splitting.*

2. **Segments rotation**: as noticeable in the *Figure 4.26*, *FFd* is not aligned to the horizontal axis as *FF*. The quality of the initial alignment greatly affects the convergence and accuracy of the ICP algorithm. Providing a good initial estimation of the relative pose between the two point clouds can lead to better results [14]. Consequently, it is necessary to coordinate the segments obtained with the segments of model. To do it, the orientation of the 2D foot mask regions corresponding to the area of segments point clouds is computed

to allow the alignment of them. When *FFd* is aligned to the horizontal axis, it is possible to identify the TOE_j as the point with the maximum value on *Y*-axis within a range of approximately of 10 mm around the most distal point on *X*-axis.

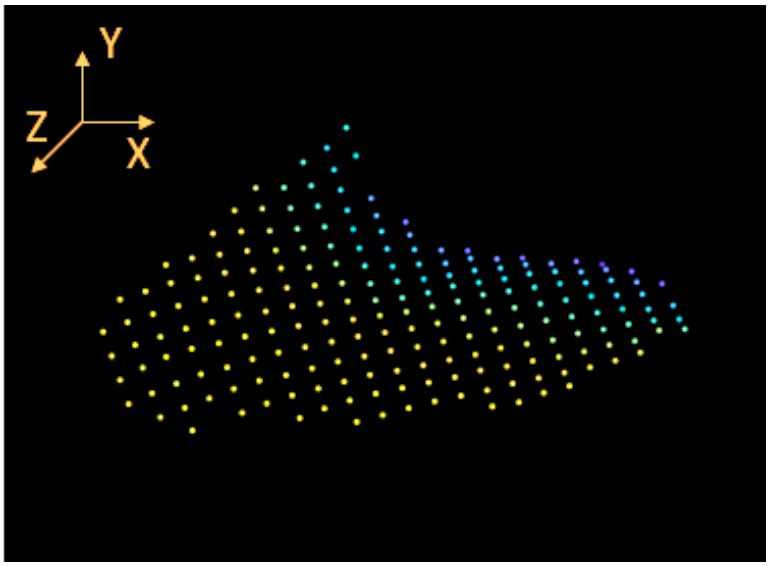

***Figure 4.27**: FFd aligned to the sagittal axis.*

3. **Scaling factor:** What is notable in the dynamic frames is that due to the presence of soft tissue artifacts and the degradation of depth images caused by movement, the foot appears smaller compared to the model. It proposes to apply a factor scale *s*, in order to adapt the dimension of *Dyn* to the model ones so as to maximize the performance of ICP.

$$s = \frac{l_m}{l_{dyn}} \tag{4.9}$$

Where l_m is the model segment length and l_{dyn} is dynamic segment length. It is developed only in ratio of lengths because the height of the model, especially in *MRF*, is obtained by identification of a 3D ROI (point 1 of paragraph 4.3.2) and it could be different from the height *3FM*, which is obtained by color filter.

4. **Segments alignment:** to pre-align each dynamic frame to the model, the centroids of 3D model and dynamic segments are calculated.

The centroid of an object refers to the geometric centre or average position of all the points that make up the object. It is computed as the mean of the coordinates of those points, where each point's position is weighted equally. The centroid is commonly used as a measure of the object's location or as a reference point for various calculations in image processing and computer vision tasks.
The centroids of the segments are aligned by adding an offset (width of the foot) on the frontal axis (*Z*). As mentioned above, Dyn represents the point cloud associated to the depth frame during the dynamic acquisition and it constitutes only the lateral view. Hence, an additional offset is necessary to ensure the alignment of the dynamic segments with the lateral side of the model. This offset helps to avoid to select the closest points on the medial side and ensure a proper alignment.

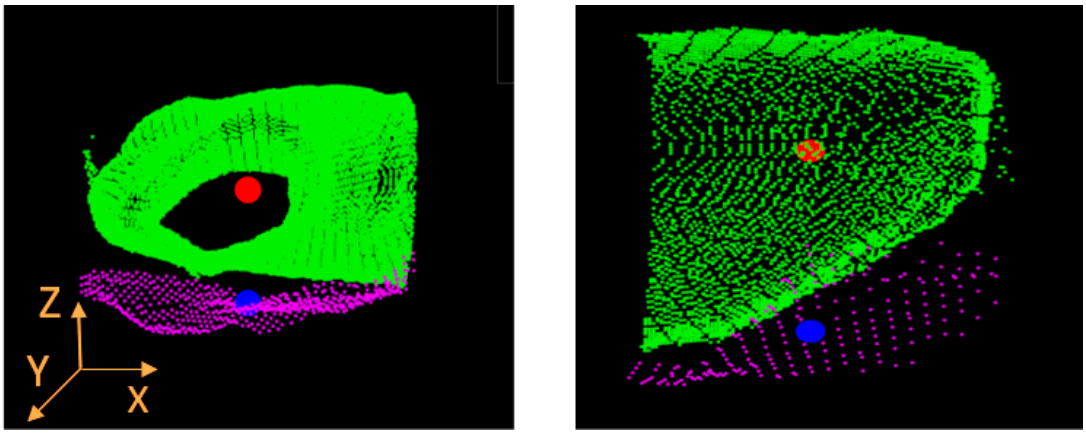

***Figure 4.28**: Pre-alignment with offset of the dynamic segments (purple point cloud and blue dot) with model segments (green point cloud with red dot) view on the frontal plane.*

5. **ICP**: To estimate the position of joint centres (*LM* and *MPT5*), the extraction of the transformation matrix is required to achieve the optimal alignment between the two dynamic segments and the respective model segments. This is possible through the application of ICP algorithms on the dynamic segments, which produces the transformed point cloud along with the transformation matrix *tform*. The metric used is '*Point to Point'* (Euclidean distance between point to point) with the number of maximum iterations is imposed equal 1000.

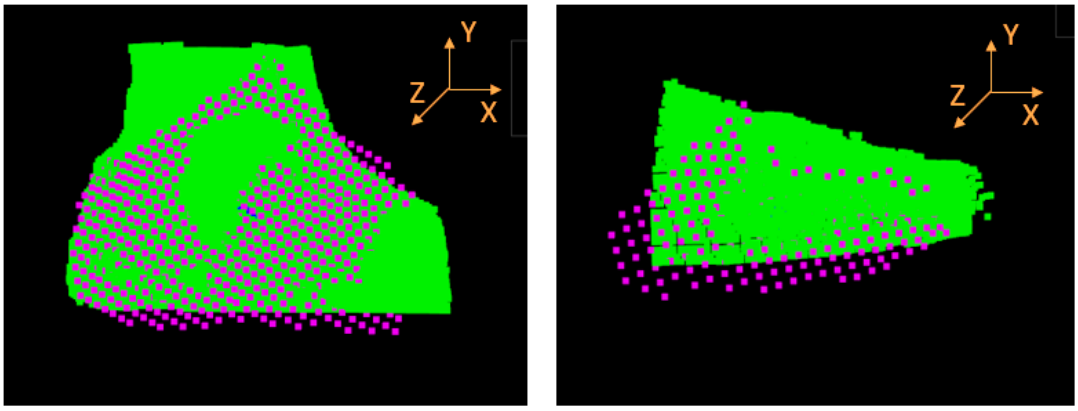

Figure 4.29: *Segment point clouds after alignment by ICP.*

6. **Joint centres placement**: *tform* is a 4x4 homogeneous transformation matrix that includes the information of rotation and translation applied to minimize the difference between the point clouds.
   Applying the inverse matrix to the landmarks *LM* and *MTP5* of 3D foot model it becomes possible to estimate their position in the respective dynamic segments prior to ICP application. Specifically, the landmark LM_j is estimated applying two inverse transformation matrixes to the *LM* (lateral malleolus of the 3D foot model). The first is that output from the alignment of Mid-Rear-foot dynamic segment (*MRFd*) and the respective segment model (*MRF*), while the second is from the alignment *MRFd* and the horizontal axis (point 1 of this paragraph). The same protocol is applied to $MPT5_j$ with the two inverse transformation matrixes applied on *FFd.*

7. **Fixing joint centres**: to mitigate the risk of misestimating the joint position and placing it outside the point cloud, a corrective measure is implemented. This correction entails relocating the joint from its erroneous position to the segment point that has the minimum Euclidean distance from that particular position.

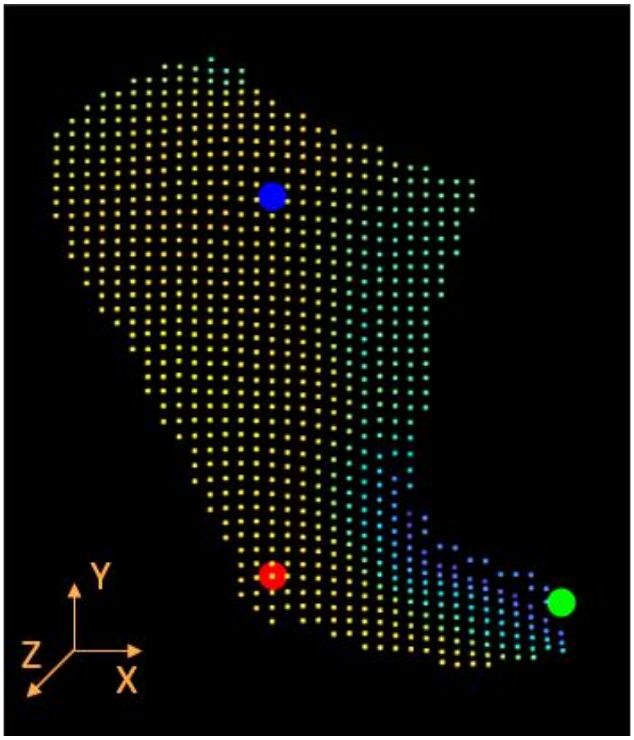

***Figure 4.30**: Dynamic right foot with TOE_j (green dot) and the joint centres estimated: LM_j (blue dot), $MTP5_j$ (red dot).*

8. **LM and MTP5 position estimation on 2D plane:** since the validation is performed on 2D images, the computation of the sagittal angle between *MRF* and *FF* requires the conversion from 3D to 2D coordinates. This conversion can be achieved by applying the inverse Formula 4.1, which allows to obtain the projection of 3D coordinate in the 2D coordinates ($^{I}_{0}LM_j$, $^{I}_{0}MTP5_j$ and $^{I}_{0}TOE_j$), enabling the computation of the sagittal angle.

4.4 Shank kinematics estimation

As the previous paragraph the next big step of this work, it is the shank kinematics estimation, which is achieved through the proposed method by Balta et al. [13]:

- Subject segmentation
- Knee joint centre estimation

4.4.1 Subject segmentation

Subject segmentation involves the identification of the subject in the image, that makes possible to isolate the subject from the background through a series of distinct steps, as described below.

1) **Background subtraction**: for each image acquired in dynamic acquisitions, it is applied a subtraction operation between the frame containing the subject and the background image. Then the difference obtained is converted to grayscale (4.10).

$$D_{x,y,c} = \left| B_{x,y,c} - I_{x,y,c} \right| \tag{4.10}$$

$$D\_gray_{x,y,c} = \sqrt{D_{x,y,R}^2} + D_{x,y,R}^2 + D_{x,y,R}^2 \tag{4.11}$$

Where x and y refer to the image coordinates, c is the color channel of RGB image, I is the image containing the subject and B is the background image.

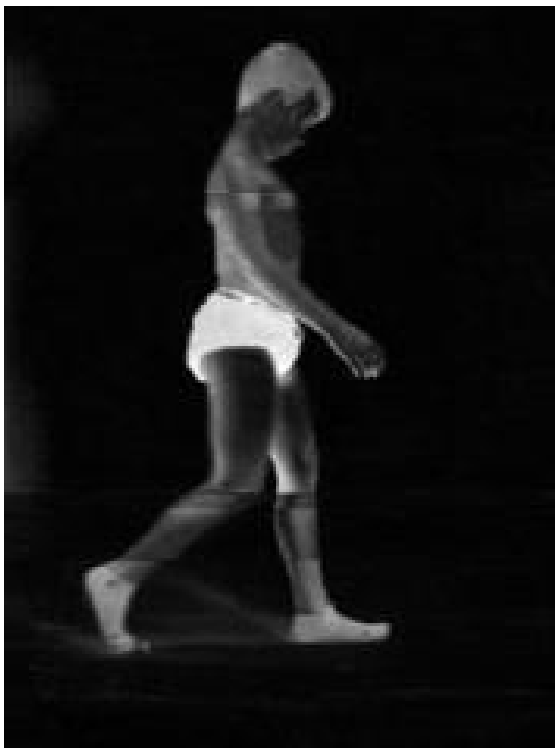

***Figure 4.31**: Difference image in grayscale D_gray.*

2) **Gray thresholding**: a histogram of grayscale is calculated to determine a threshold value, which is used to separate the subject from the background [39] and the weighted mean (Th) of the histogram is calculated using:

$$Th = \frac{\sum_{i=0}^{N} w_i * x_i}{\sum_{i=0}^{N} w_i} \tag{4.12}$$

Where w_i is the histogram count for the i-th class and x_i indicates its grayscale level.

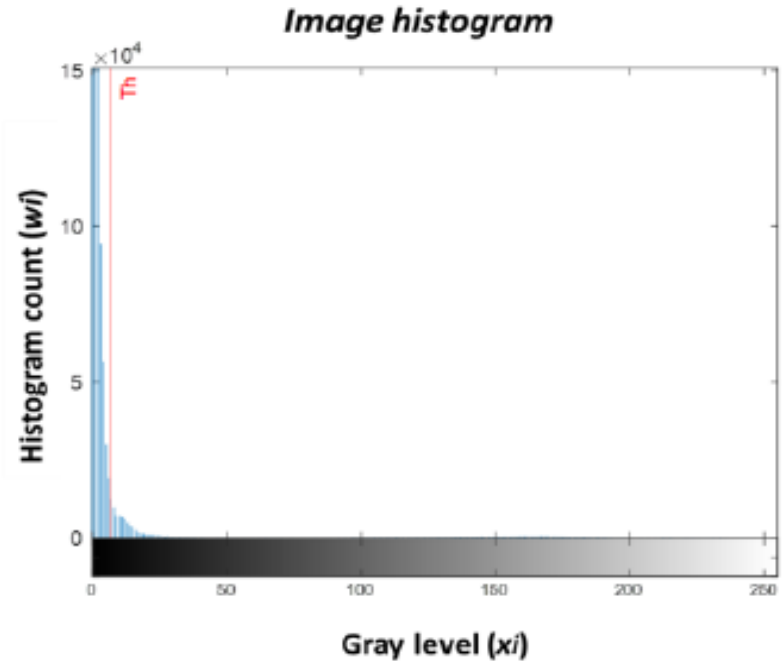

***Figure 4.32**: Histogram of the grayscale image. Red line represents the threshold Th_0.*

The segmentation mask (*Figure 4.33.a*) is generated using the following formula:

$$S_{x,y} = \begin{cases} 0, & p_{x,y} < Th \\ 1, & p_{x,y} \geq Th \end{cases} \tag{4.13}$$

In addition, especially in cases with not homogenous background, it is important to remove all the undesired regions, which may be source of errors.

3) **Mask cleaning**: residual areas belonging to the background was removed by applying a threshold on the depth image (*Figure 4.33.b*).
In addition, to remove the presence of the shadows on the ground, using color filter (paragraph 4.1) applied to the RGB images, the feet are deleted from the mask (*Figure 4.33.c*), leaving only the shadows, which are easy to remove. Then, the feet are reinserted in order to final subject segmentation (*Figure 4.33.d*).

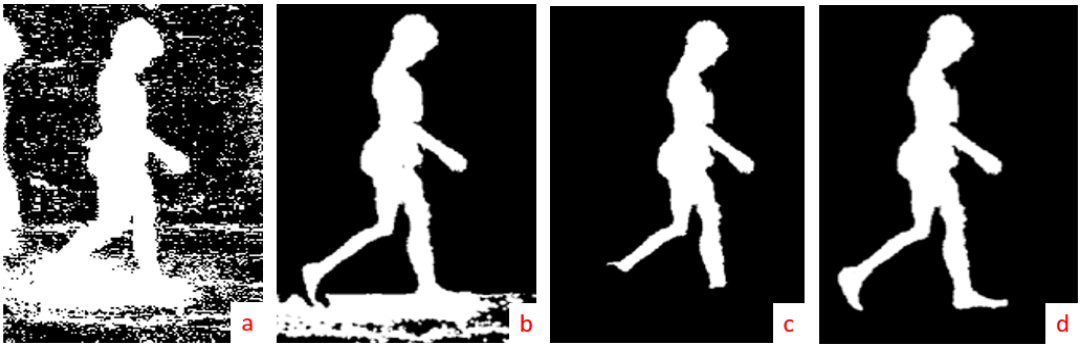

Figure 4.33: *Cleaning mask process on the subject mask: a) represents D_gray after application the thresholding, b) shows the binary and depth mask overlapped, c) represents the subject without the feet, d) shows the subject segmented.*

4.4.2 Knee joint centre estimation

To estimate the knee joint position, first it is necessary to identify the shank segment on the subject.
The separation between the foreground and background shanks was carried out using two alternative strategies depending on whether there was or was not overlap between foreground and background shanks.
To discriminate between overlap/non overlap conditions, a circle centred in ${}^{I}LM_{j}$ (the 2D projection of LM_j extracted in the point 8 of paragraph 4.33) with radius equal to the distance between ${}^{I}LM_{0}$ and ${}^{I}LE_{0}$ (lateral epicondyle on template) was drawn. If there was no overlap, points were grouped in two separated regions, and the foreground shank, being closer to the camera, coincided with the largest area (*Figure 4.34a*).
Conversely, when there was overlap, a single connected region was found, and auxiliary depth sensor data were used to separate foreground and background shanks (*Figure 4.34b*).

The histogram of depth values within the region was obtained and the Otsu method [40] was applied for a binary classification (class 0: foreground shank, class 1: background shank) based on the minimization of the variance between classes.

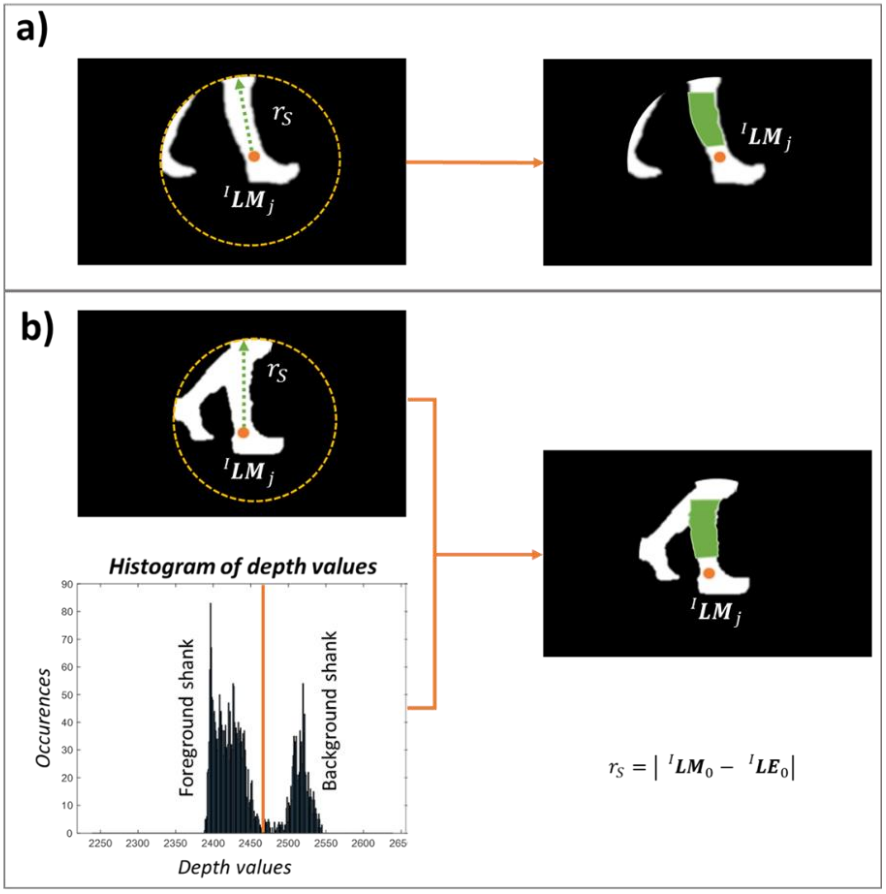

Figure 4.34: *Separation between foreground and background shanks. A circle centred in ${}^{I}_{0}LM_j$ with radius equal to Euclidean distance between ${}^{I}_{0}LM_0$ and ${}^{I}_{0}LE_0$ is drawn. a) No- overlapped shanks. b) Overlap between the two shanks. Below the histogram of depth values inside the region and through Otsu method the two shanks are separated.*

The central portion of the foreground shank is defined in the anulus identified between the 25% and the 75% of the shank length (paragraph 4.2.2).
The shank coordinate system s_j is defined through the axis aligned with principal axes of inertia of the shank mask ellipse (*Figure 4.34*). Moreover, the transformation matrix ${}^{I}T^{0}_{s_j}$, based on the ICP algorithm (paragraph 4.3.2) from s_j to image, is calculated.
The lateral epicondyle position (${}^{I}LE_j$) for each frame is obtained by applying the three subsequent transformations:

$$^{s_0}LE_0 = {}^{s_0}T_I \; {}^{I}LE_0; \; {}^{s_j}LE_j = {}^{s_j}T_{s_0}^{0} \; {}^{I}LE_0; \; {}^{I}LE_j = {}^{I}T_{s_j}^{0} \; {}^{s_j}LE_j \quad (4.14)$$

Where ${}^{s_j}LE_j$ is the lateral epicondyle referred to the dynamic shank template, ${}^{s_j}T_{s_0}^{0}$ is the transformation matrix from the shank template to the dynamic shank template of each frame and ${}^{I}LE_0$ is referred to the shank template.

4.5 Sagittal angles computation

The sagittal are computed as follows:
The ankle joint is computed as the inclination between Lateral Epicondyle (LE) - Lateral Malleolus (LM) segment minus 90 degrees.

$$\theta_{ankle} = arccos\left(\frac{LE-LM}{|LE-LM|} \cdot \frac{MTP5-LM}{|MTP5-LM|}\right) \text{-}90° \quad (4.15a)$$

The ankle joint is computed as the inclination between Lateral Malleolus LM) - Metatarso-phalangeal (MTP5) joint segment.

$$\theta_{MTP5} = arccos\left(\frac{LM - MTP5}{|LM - MTP5|} \cdot \frac{TOE - MTP5}{|TOE - MTP5|}\right) \quad (4.15b)$$

Each gait was made of a different number of frames, so the resulting angles are referred to the percentage of gait cycle (0-100%) through spline interpolation.
The kinematic curve is then filtered using a 3rd-order low-pass Chebyshev filter with a cut-off frequency of 7 Hertz.

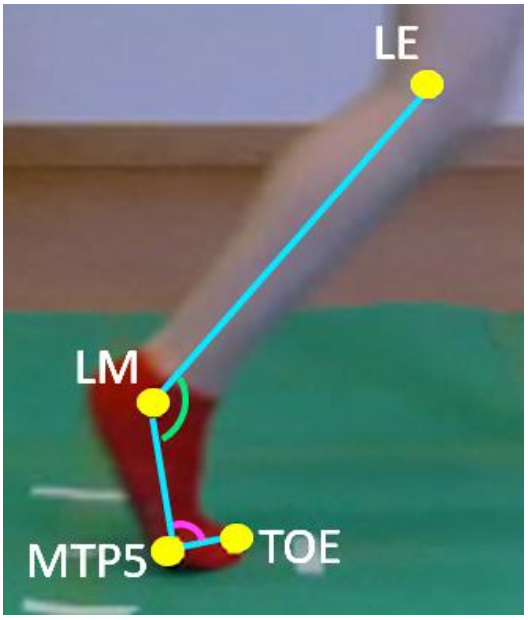

***Figure 5.6**: The sagittal angles of metatarso-phalangeal joint (pink arch) and ankle joint (green arch).*

5 Experimental acquisitions

5.1 Azure Kinect DK

Azure Kinect development kit refers to a comprehensive hardware and software package developed by Microsoft in 2020. It consists of a depth-sensing camera device which combines a depth sensor, high resolution RGB camera, spatial microphone array and an inertial measurement (IMU).

In addition to the hardware, the Azure Kinect DK [41] provides three software developments kits (SDKs): a sensor for lower sensor and device access, Body tracking SDK for tracking bodies 3D and a Speech Cognitive Service SDK[33], [41].

Figure 5.1: *Microsoft Azure Kinect camera.*

This device captures various visual data through a 12 MP RGB camera and 1MP Depth sensor, returning RGB color image, a depth map and an IR image (720x1280 at 30 Fps). The pixels of the last one represents the amount of light reflected from the scene acquired. The image settings, such as the camera exposure, brightness and white balance can be manually managed.
The camera operates based on the Time of Flight (TOF) principle, where it emits an amplitude-modulated continuous wave in the near-infrared spectrum onto the scene.

It then captures the indirect time taken by the emitted light to travel from the camera to the scene and back again.
Processing the TOF measurements, the camera can output a depth map, with three dimensions, where the Z coordinate of every pixel, represents the distance from the camera in millimeters [42].

As noticeable in *Figure 5.2,* there are some cases where the reconstruction of depth may be affected by errors, resulting in invalided pixels. The invalid pixels are represented in the depth map as value equal to zero. The cause of errors could be:

- When the pixels are outside of the active IR illumination mask. This occurs on the corners of the image (blue dots).
- The IR signal on the pixels is saturated (*e.g.* high intensity reflected light from surfaces).
- The IR signal is not quite high to generate depth values (greed dots).
- Multi-path interference on the pixels that received IR signal from different objects in the scene (pink dots).

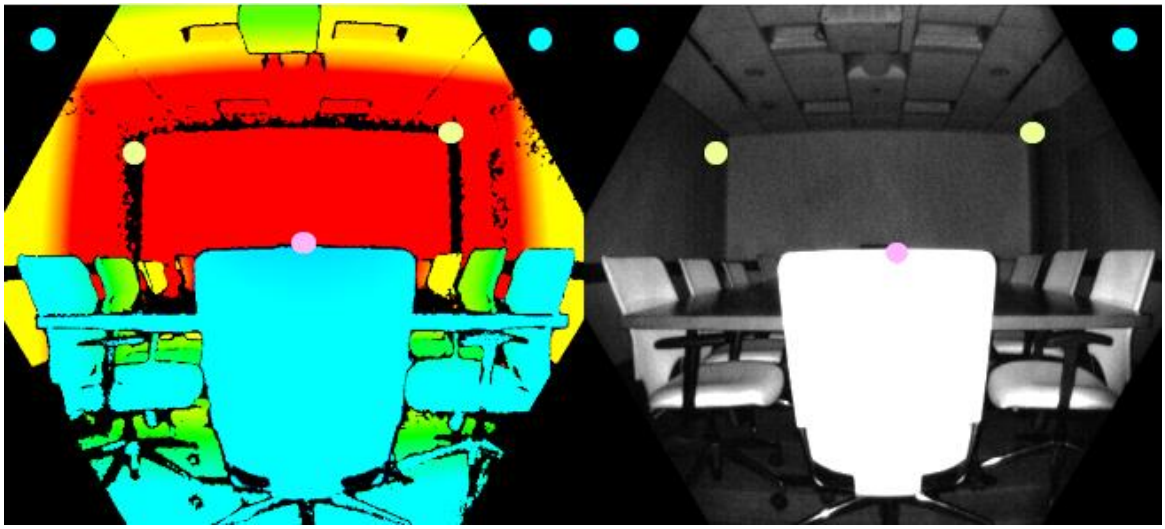

Figure 5.2: *The invalided pixels on the Depth mask (left) and respective IR image (Right).*

5.2 Experimental and Protocol and Setup

The experimental protocol for studying the gait cycle involved of 3 gait cycle for each side in 10 subjects affected by Clubfoot, mean age 13 years (7-17), 5 females

and 5 males. Out of these patients, six subject have Bilateral Clubfoot, three have the right limb affected, and one patient has the condition on the left side.
The patients were requested to walk at their own comfortable speed along a straight 10-meter walkway while wearing ankle socks of different colors (red for the right foot and blue for the left foot) and underwear.
The experimental set-up included:

- One green carpet
- One RGB-Depth camera positioned laterally approximately 2.5 meter from the walkway, at the height of 85 centimetres.
- Two LED lamps set a to the maximum intensity and positioned laterally by the camera.

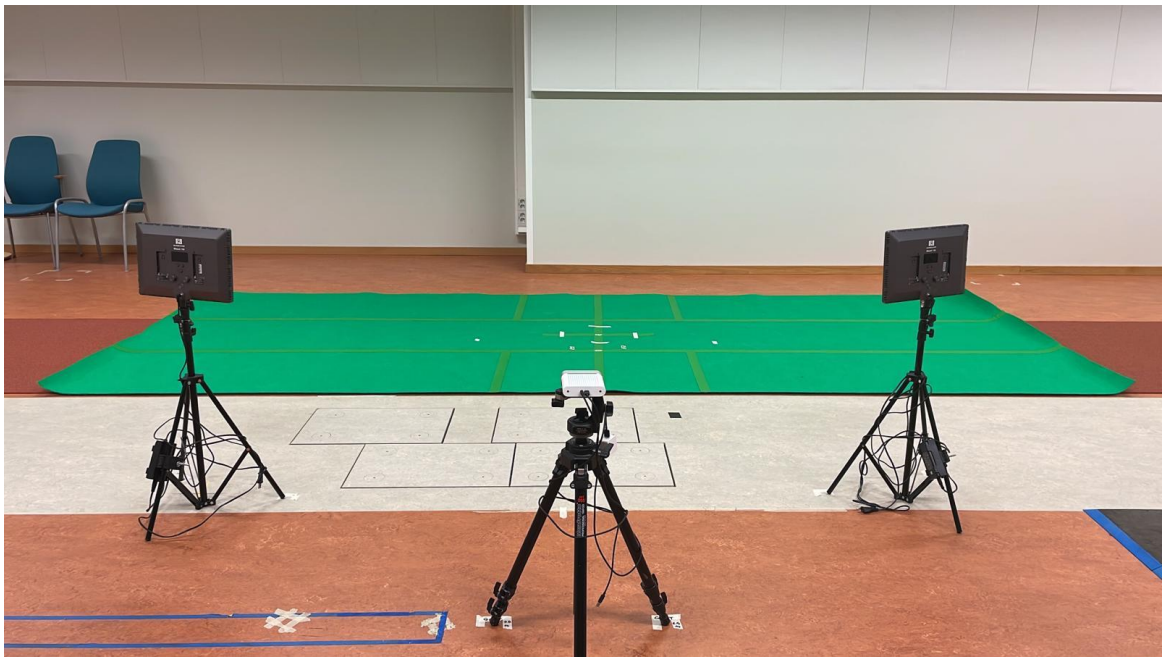

Figure 5.3: *The experimental set up in Skaraborg hospital of Skövde.*

Before starting the trials, subjects were submitted to several static acquisitions. In particular, 4 different static one-side view (frontal, posterior and 2 sagittal) with the subject in an upright standing position (2.5 m distant from the camera). In addition, 4 different static one-side view for both feet (paragraph 4.2.1) with the camera placed on the ground and distant from the foot of 0.6 m.

As shown in *Figure 5.4* on the lateral side of both sides, four joints (great trochanter (*GT*), *LE, LM* and *MTP5*) are identified by palpation of an expert operator and marked with a black pen on small, coloured labels.

Figure 5.4: *Illustration of labelled joints on the lower limb.*

5. 3 Infrared interference

The manual labelling was performed to validate the automatic algorithm purposed.

It was necessary adopt this validation, since the acquisitions were not conducted synchronously between the marker based and Markerless systems due to observed interferences in the depth map reconstruction in the Azure Kinect recordings.

The gait analysis laboratory of Skövde Hospital studies the joint kinematics with the marker based system (Motion capture camera Qualisys). This system works with the same infrared wavelength (850 nm) of the Azure Kinect and it caused several undefined values on the depth mask, taking a great loss of information. When both systems work synchronously, the interference of IR signals causes the inability to generate the depth value (randomly) in the majority part of the scene by the Azure Kinect (*Figure 5.5a*).

In addition, the presence of retroreflective markers produces errors in the depth mask. As shown in the *Figure 5.5b* the pixels, that should represent the markers

positioned on the joint centres, have values equal zero or are classified not-a-number (NaN). This poses significant issue when the maker-placement requires a large number of markers, such as the Oxford protocol [43] utilized by the gait analysis laboratory in Skövde. The protocol foresees the placement of 10 markers on the foot, resulting a very huge loss of anatomical information.

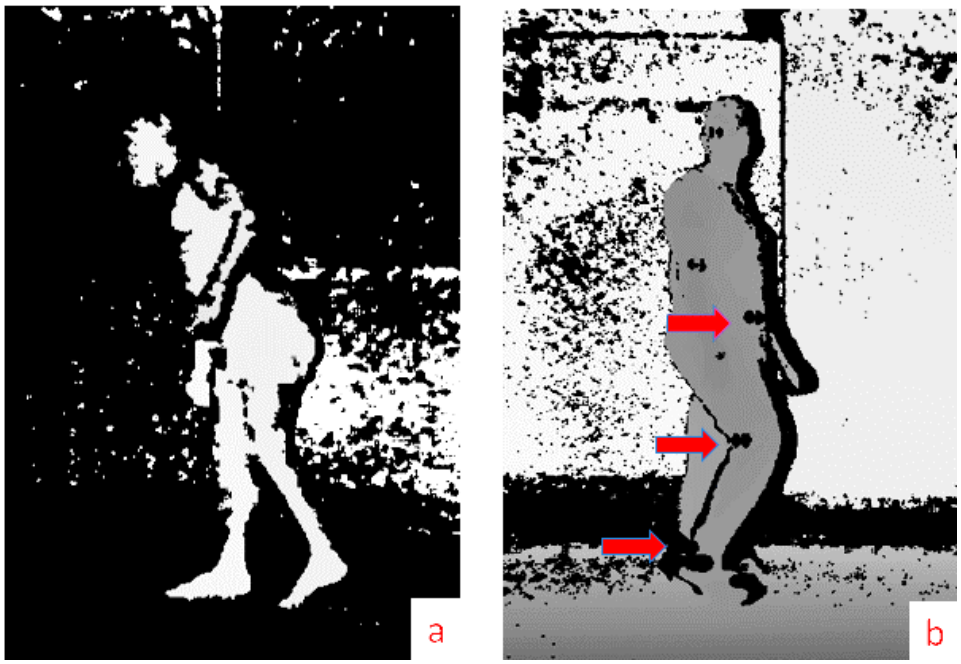

Figure 5.5: *a) Illustration of the IR interference between Azure Kinect and Qualisys system. b) Black invalidated pixels in correspondence of the reflective markers (red arrows)*

5.4 Manual labelling

To avoid the concerns about interference, a manual clicking method on the labelled joints for each frame was employed to validate the output generated by the developed algorithm. Once obtained the coordinates of the labels by the manual clicking on the 2D RGB images, the sagittal shank and foot kinematics is compute using the same formula (*Formula 4.15*) used in the automatic method.

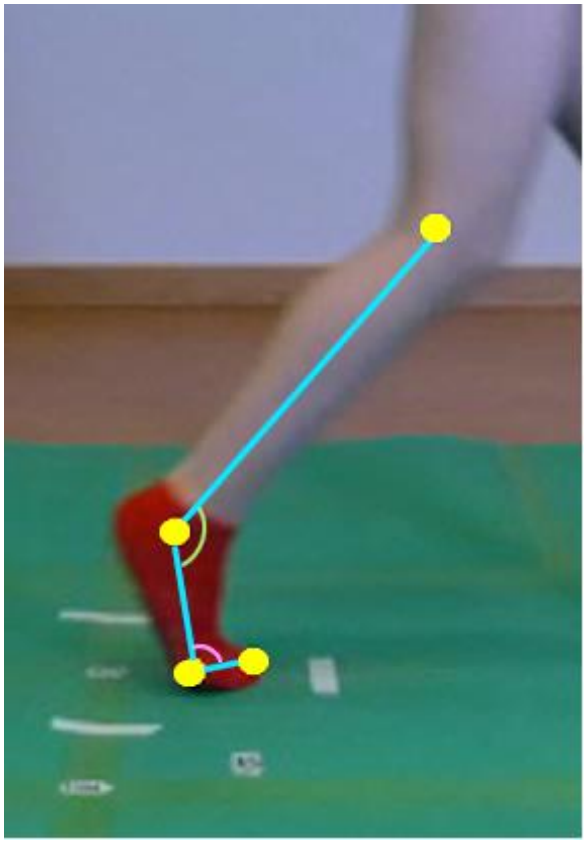

***Figure 5.6**: Labelling by manual clicking on RGB images on the joint centres (LE, LM and MPT5) and TOE (yellow dots). The sagittal angle of the ankle it is represented by the green arch while the sagittal angle of the MTP5 by the pink arch.*

6 Results and Discussions

6. 1 Evaluation of foot modelling accuracy

To validate the reconstruction of the 3D foot model, the mean absolute error and the mean percentage error between the estimate foot length and the measured one were computed. The length of 3D model is obtained as the Euclidean distance between heel and toe position in the 3D foot model.

ERROR (3D Foot model) (mm) (%)	**P_01**	**P_02**	**P_03**	**P_04**	**P_05**	**P_06**	**P_07**	**P_08**	**P_09**	**P_10**	**Mean**
R	2 (0.8%)	19 (8.2%)	16 (6.0%)	18 (7.7%)	15 (6.6%)	8 (4.5%)	21 (7.8%)	11 (5.1%)	26 (10.7%)	11 (4.9%)	12.3±7.6 (5.2±3.0)
L	18 (7.3%)	14 (5.9%)	27 (10.0%)	16 (6.9%)	10 (4.5%)	16 (8.7%)	18 (6.7%)	13 (6.2%)	13 (5.1%)	15 (6.4%)	16±4.4 (5.7±2.4)

Table 6.1*: Absolute and relative errors in term of foot length (mm and in %) of the 3D Foot model.*

As shown in the Table 1, the relative errors fluctuating around 5 % (5.2 ± 3.0 % (R), 5.7 ± 2.5 %(L)). Although the length obtained from the model may not accurately represent the actual length as measured manually, since the extreme coordinates of the point cloud could be different form the extremes measured, it can still serve as an approximate measure of the accuracy of the 3D foot model creation.

6.2 Comparison between 2D markerless kinematics and manual labelling

To validate the estimate lower limb joint kinematics, the sagittal angles of fifth metatarsal-phalangeal joint and ankle obtained from the proposed markerless method is compared with the 2D angles achieved by manual labelling.
The accuracy of the proposed method was assessed in terms of offset between the two curves and waveform similarity by estimating the root mean square errors (RMSE). To facilitate a comparison between the two protocols, the mean value is subtracted from both curves, aligning them and removing any offset differences. This subtraction of the mean value enables a clear assessment of the error between the two angle distributions, without the influence of offset variations. The offsets are obtained through the difference between the mean value of automatic kinematics and the mean value of labelled one.

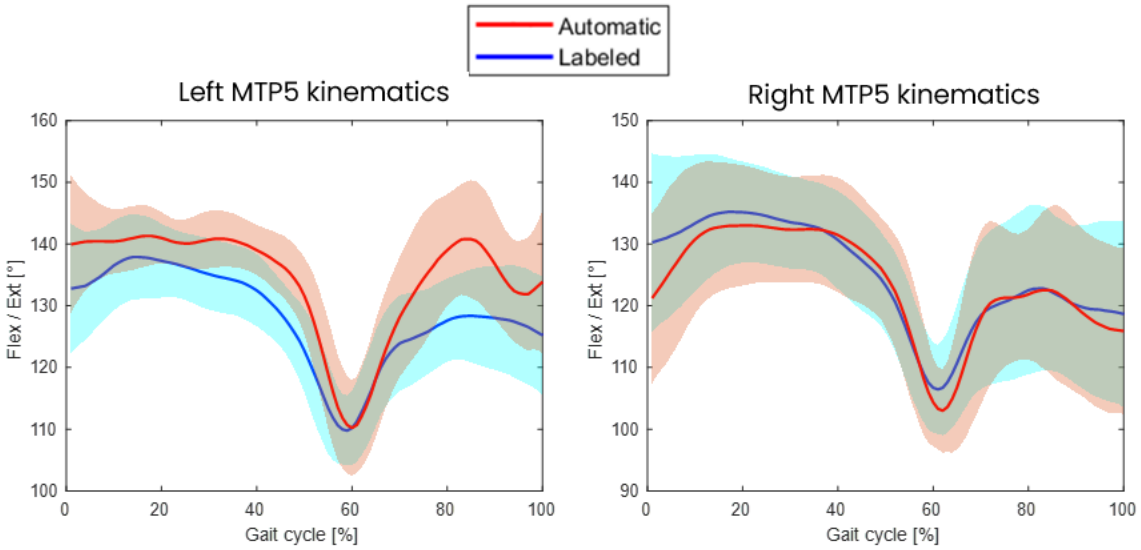

***Figure 6.1**: MTP5 kinematics obtained from the automatic method proposed (red lines) and from manual labelling (blue lines) and respective deviation standard (light red and light blue areas) for three trials for both sides.*

The primary aspect that immediately draws attention in *Figure 6.1* is the high variability in both automatic and labelled curves identified by the deviation standard, represented by light blue and light red areas. This can be attributed to the substantial inter-patient gait variability resulting due to the high heterogeneity of Clubfoot patients. For instance, in *Figure 6.2*, it is evident that the waveforms for

that specific patient deviate from the mean trend of Right MTP5 kinematics shown in *Figure 6.1*. This observation highlights the lack of standardized gait cycle among patients, as individuals with clubfoot are affected differently, causing a high variance between the patients gait cycle.

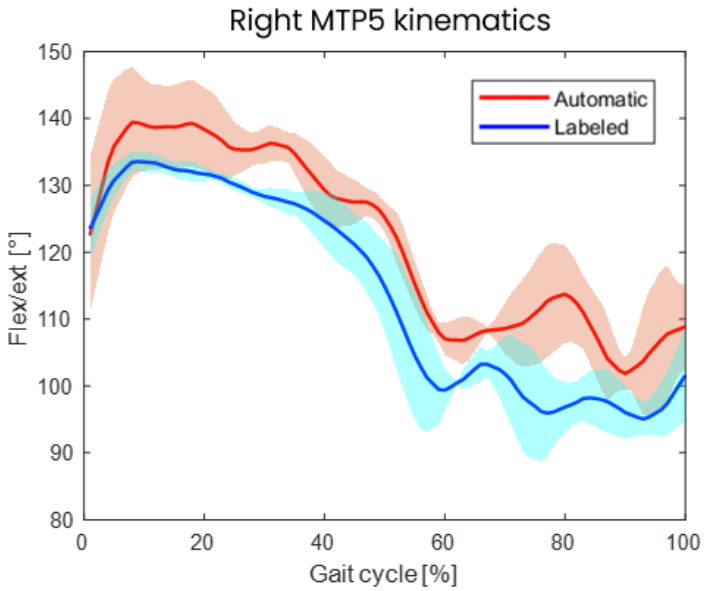

***Figure 6.2**: Right MTP5 automatic (red line) and labelled (blue line) kinematics of single patient.*

The offset between the mean kinematics of the metatarso-phalangeal joint and ankle joint are evaluated and illustrate in *Table 6.2*.

OFFSET *MTP5* (3 trials) (deg)	**P_01**	**P_02**	**P_03**	**P_04**	**P_05**	**P_06**	**P_07**	**P_08**	**P_09**	**P_10**	**Mean**
R	1.2 ±1.7	7.5 ±2.9	1.8 ±1.0	0.4 ±0.3	3.5 ±1.9	7.8 ±3.0	1.9 ±0.9	0.7 ±0.5	2.7 ±0.9	7.8 ±1.2	3.5 ±3.0
L	1.1 ±0.3	6.6 ±2.7	6.0 ±0.5	9.0 ±1.4	10.0 ±4.7	12.8 ±0.5	9.7 ±0.7	5.9 ±1.5	2.3 ±1.5	1.4 ±0.9	6.5 ±4.0

***Table 6.2**: Mean and standard deviation of the offset of 3 trials between the MTP5 joint kinematics obtained from automatic and labelled methods.*

The errors in terms of RMSE between the mean kinematics of the metatarso-phalangeal joint are evaluated and illustrates in *Table 6.3*.

RMSE ***MTP5*** (3 trials) (deg)	**P_01**	**P_02**	**P_03**	**P_04**	**P_05**	**P_06**	**P_07**	**P_08**	**P_09**	**P_10**	**Mean**
R	5.8 ±0.4	3.7 ±0.9	5.0 ±0.8	5.5 ±0.7	4.7 ±0.1	5.3 ±1.3	4.6 ±0.7	4.4 ±0.2	5.2 ±0.9	3.6 ±0.2	4.8 ±0.7
L	5.3 ±0.2	5.2 ±0.7	5.1 ±0.0	5.1 ±0.3	5.7 ±1.1	5.8 ±1.3	5.2 ±0.3	4.4 ±1.0	5.9 ±0.6	5.5 ±0.8	5.3 ±0.5

Table 6.3*: Mean and standard deviation of the RMSE values after removing the offsets of 3 trials between the MTP5 joint kinematics obtained from automatic and labelled methods.*

The mean errors in terms of RMSE and Offset between the mean kinematics of the ankle joint are evaluated and illustrates in *Table 6.4*.

Ankle	**RMSE** (3 trial) (deg)	**OFFSET** (3 trial) (deg)
Right	4.8 ± 0.7	2.3 ±1.7
Left	4.9 ±1.6	3.5 ±2.0

Table 6.4*: Mean and standard deviation of the RMSE values after removing the offsets of 3 trials between the ankle joint kinematics obtained from automatic and labelled methods.*

Upon analysis of *Table 6.4*, it is evident that the errors between the right and left foot are relatively similar, with the right foot displaying a slightly lower error (4.8 ± 0.7 degrees) compared to the left foot (5.3 ± 0.5 degrees). Specifically, *Table 6.2*

demonstrates that the left foot kinematics are affected by a higher offset (6.5 ± 4 degrees) compared to the kinematics of the right foot (3.5 ± 3 degrees). It could be associated with the color of socks. More details are explained in the paragraph (factor influencing the accuracy of kinematics).

6.3 Factors influencing the accuracy of foot modelling

The process of creating the foot model is semi-automated. It must be highlighted the importance of positioning the subject in the correct position during the static acquisitions in order to correctly identified the common points among the four views. One potential source of errors could be attributed to the incorrect position during the acquisition phase. Some of the models with the most significant errors occurs when the subject's shank is not perpendicular to the ground causing a different shape to the side of the foot as shown in *Figure 6.3a*.

Another problem which affects a good accuracy of the model reconstruction is the incorrect position, in particular the alignment of the foot axis with the sagittal plane during the acquisition of lateral and medial side. If this alignment is not respected (*Figure 6.3b*), the creation of the model requires an integration of manual adjustments for translation and rotation offsets since the common points could be different between the four views.

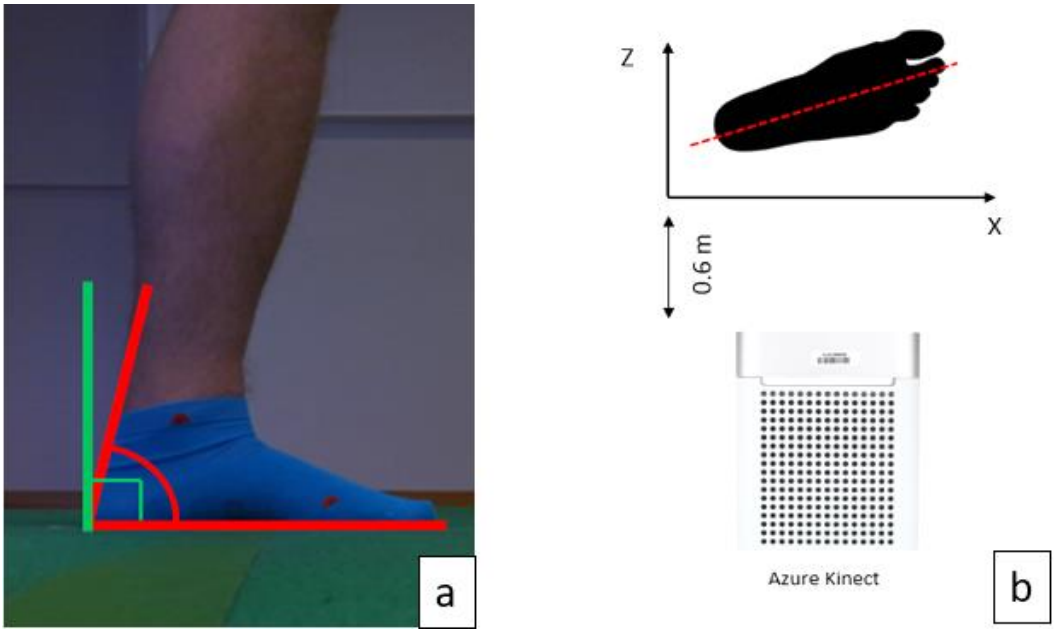

Figure 6.3: a) *The incorrect position on the static acquisition for the medial view of left foot (red lines). The patient should stand as indicated by the green line.*
b) *Representation of misalignment of principal foot axis and (red dashed line) and X-axis during the static acquisition for the lateral view.*

6.4 Factors influencing the accuracy of joint kinematics estimation

In this paragraph it will be analysed some issues which are the cause of residual errors in foot kinematics estimation related to the proposed depth completion technique and ICP algorithm

One of the most factor which influences the joint centres estimation is the application of the depth completion. In particular, the errors are caused by inaccuracies in the measurement of the depth values from the depth sensor during high-speed movements resulting also in a low number of points belonging to the foot and residual errors in foot segmentation due to the blurred RGB images.

During high-speed movements, it has been noticed that the depth sensor fails to accurately reconstruct depth values since the limited exposure time of RGB-D cameras can lead to motion blurs in captured images, potentially causing artifacts such as holes or fake boundaries [44], resulting in improper alignment between the depth image and the RGB image specifically at the foot (*Figure 6.4c*). To address

this issue, a solution is implemented where only the depth values within the RGB segmentation are selected using a color filter for subsequent steps.

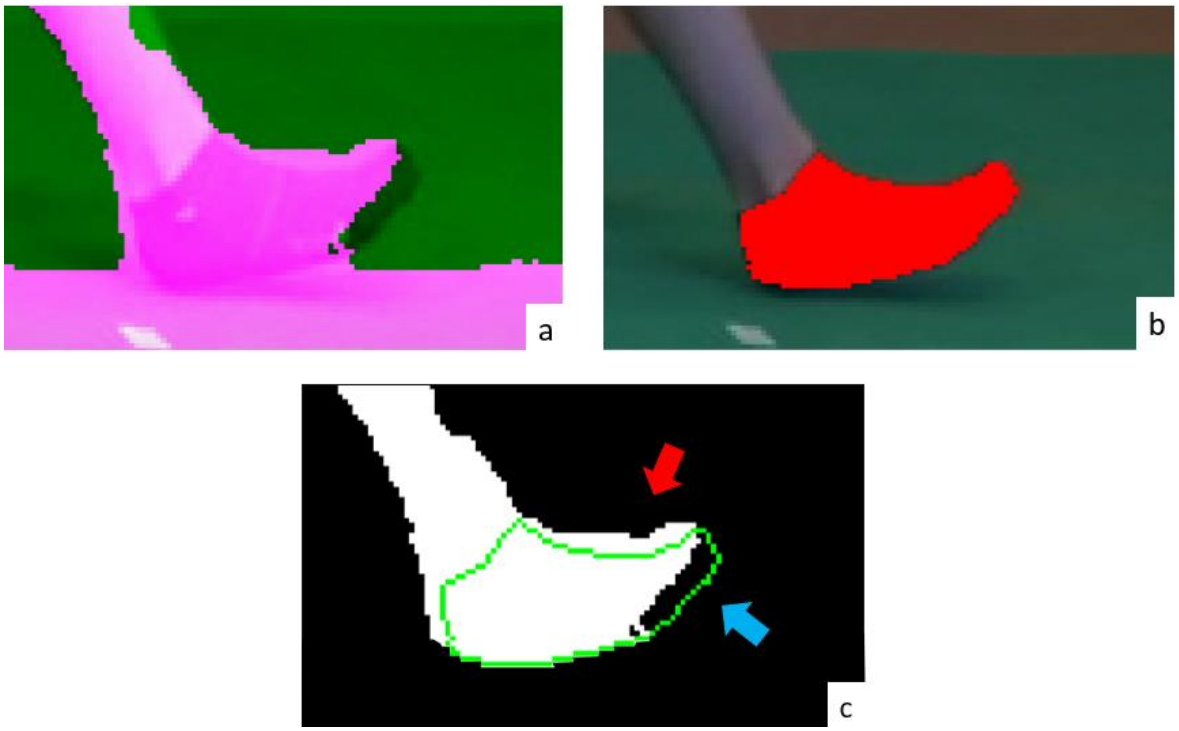

Figure 6.4: *a) Depth map (pink pixels) overlapped on RGB images. b) RGB mask (red pixels) overlapped on RGB image. c) Overlap of RGB mask edge (green pixels) and subject segmented depth map (white pixels). Red arrow indicates the misalignment while blue arrow indicates the missed depth values.*

Another issue arises when there are a significant number of missing points (*Figure 6.5a*), leading to an imperfect reconstruction that does not accurately represent the true 3D shape of the foot (blue arrow *in Figure 6.5b*). As indicated by red arrows in *Figure 6.5b,* another component which leads to a missing of depth values, is the incorrect placement of labels by the operator, too close to the edge of the socks or to the profile of foot. This placement in dynamic phase causes holes in the RGB mask, difficult to fill.

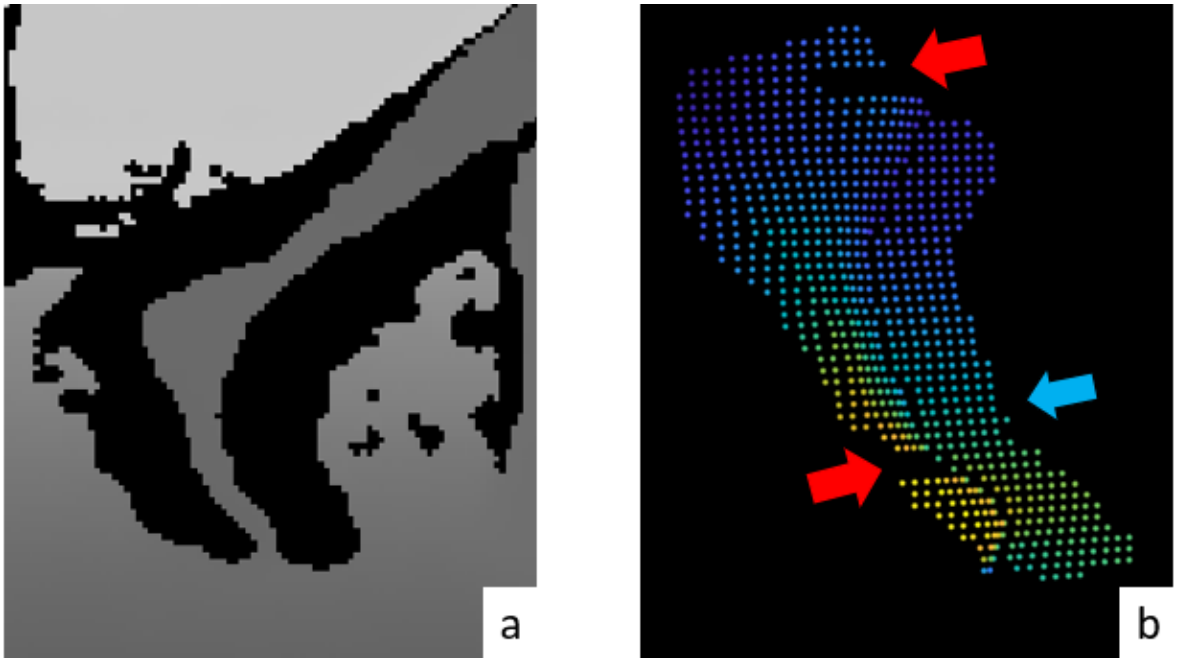

Figure 6.5: *a) Original depth map obtained by the depth sensor. b) Dynamic point cloud with the region rebuilt by depth completion (blue arrow) with the holes caused by the incorrect position of labels (red arrows).*

In addition, the incorrect segmentation of the feet, such as for the right foot (covered by red sock), which in cases of overlap between the two feet (*Figure 6.6a*), leads to major errors in depth completion. Also, for the left foot, which is covered by blue sock, during the stance phase, the color filter encounters difficulties in accurately delineating the foot's edges, due to the similarity in color with the green carpet, or the shadows caused by the foot itself (*Figure 6.6c*). These results in the identification of pixels on the depth map that do not belong to the foreground foot contribute to generate inaccurate point clouds (*Figure 6.6b* and (*Figure 6.6d*) by the depth completion algorithm, which in turn affects the alignment of the segments through the use of ICP.

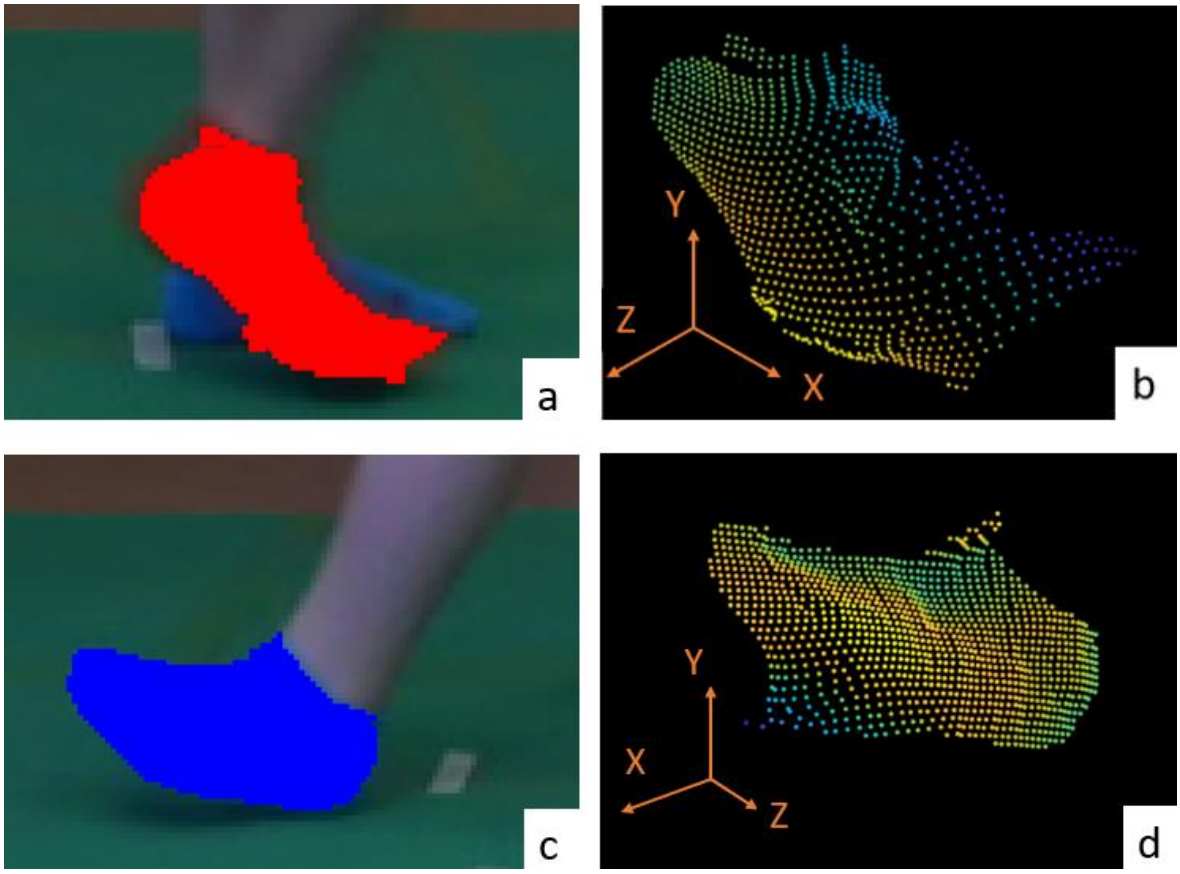

***Figure 6.6**: Incorrect segmentation of the right foot due to feet overlap (a) and of the left foot due to similar color between sock and the carpet, leads an erroneous depth completion generating inaccurate point clouds (b, d).*

Furthermore, in phase of swing the foot undergoes rapid movement causing the blue color of the sock to blur. This blurring effect makes it even more challenging to perform to perform the manual labelling of the small labels for validating the joint kinematics (*Figure 6.6*). This difficult to label in the correct pixels add more errors in phase of validation. These problems could explain the higher errors obtained in the Offset and RMSE of left *MTP5*.

Figure 6.7*: RGB images of left foot during the swing phase showing the difficulties to identify the labels.*

The Iterative Closest Point (ICP) algorithm represents a rigid alignment technique that assumes the objects under examination are rigid segments. However, the human foot cannot be considered a rigid segment as it can undergo shape variations, especially during the swing phase, due to soft tissue artifacts [12]
Moreover, it is important to emphasize that applying the ICP algorithm to feet affected by deformities can pose challenges in achieving accurate alignment between the model and the moving foot.
In this thesis project has been employed a rigid ICP algorithm to address artifacts robustly. However, in cases of insufficient data points or foot deformability, considering the application of non-rigid ICP or deformable shape models may lead to improved estimations of joint kinematics [12].

As stated in the point 2 of paragraph 4.3.3 pre-alignment plays a crucial role to improve the efficiency and accuracy of the ICP [14] since its good fitting is essential to estimate the correct kinematics [12]. In the context of this study, various tests were conducted to determine the optimal initial conditions to provide to the ICP algorithm, resulting that the best pre-alignment involves placing the dynamic point cloud in an external position relative to the model, in order to align the dynamic to the lateral view of the entire model. Despite this precaution, any adaptation issues can be resolved by solely identifying the points of the model belonging to the sagittal view, based on the normal vector of the point cloud. This implies selecting points in the point cloud whose normal vector has the same direction (within a tolerance) as the vector representing the sagittal plane. This approach could help limit the selection of point clouds and reduces errors. The RMS

error of fitting points was employed to get a quantitative measure to assess the fitting quality. Even if it gives a first indication of the measurement, it has been often difficult to interpret it [12].

One of the factors that significantly impacts the quality of the ICP algorithm is the incorrect alignment of the dynamic foot data (*Dyn*) with the horizontal axis. In the foot splitting process, outlined in the chapter 4, the alignment is achieved by orienting the inertial principal axis of the 2D foot mask. However, if this axis of the 2D foot mask does not align with the foot's horizontal axis, it results in an erroneous alignment (*Figure 6.8a*) and introduces a cascade of errors. As a result, due to the incorrect splitting, the Mid-Rear-foot segment may have a shorter length than its actual length (and vice versa for the Forefoot), which leads to errors in calculating the scaling factor.

Another issue to consider is the presence of fewer points in the distal part of the foot compared to the posterior part (Forefoot to Mid-Rear-foot points ratio = 0.26) which can lead to errors in estimating the position of the fifth metatarsophalangeal joint (MTP5).

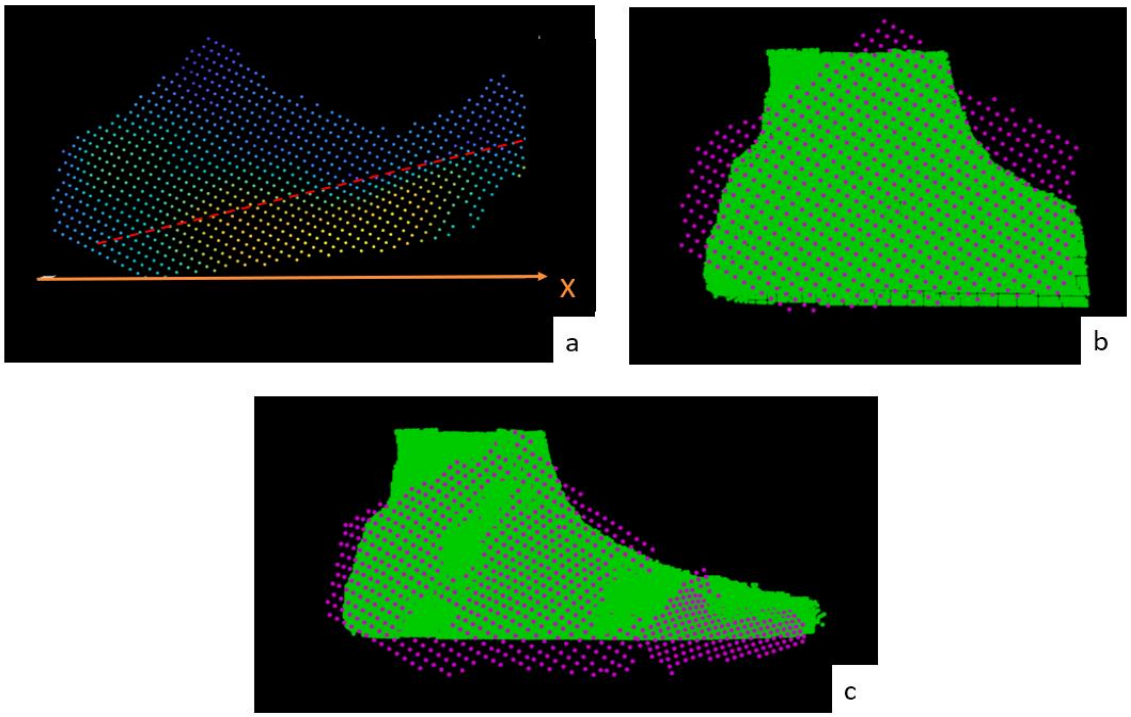

Figure 6.8*: a) Erroneous alignment between principal axis of Dyn (dotted line) and X-axis. b) overlapping of MRFd (pink points) and MRF (green points). c) the overestimated alignment*

7 Conclusions

Despite the marker-based system optoelectronic stereophotogrammetry nowadays represents the gold standard of human motion with its high accuracy of marker position, it has several limitations such as the proper marker placement on the skin of the patient which implies considerable preparing time, the necessity of a prepared personal and the cost of dedicated spaces. Nowadays, video-based markerless system presents a promising and cost-effective alternative to the marker-based system, offering the advantage of speeding up the acquisitions by removing the need of markers placement on the subject's body.

In the recent years, several companies have developed affordable RGB camera integrated with infrared depth sensor (RGB-Depth). These proposes, based on a single RGB-Depth camera (e.g. Azure Kinect body tracking SDK, OpenPose), model the foot as a single segment without articulating the metatarso-phalangeal joint kinematics, which is crucial to guarantee an affective load of the foot and correct progression [11]

This thesis aimed to design a markerless method based on a single RGB-Depth camera to estimate sagittal ankle and metatarso-phalangeal kinematics using a two-segment 3D foot model and explore its clinical applicability on children with foot deformities. The method proposes to expand the 2D Markerless protocol proposed by Balta et al., 2020, including a 2D subject-specific foot model developed by Balta et al., 2020 to 3D one by using a 3D two-segments foot model (Mid-Rear-foot and Forefoot segments).

The proposed 3D protocol is structured in two parts:

1) Creation of a 2-segment 3D foot template by merging four static views of the foot (Frontal, Lateral, Medial, Posterior) by aligning three common points on the foot sole of each view. Then a 2D shank template was obtained as in Balta et al., 2020. Both templates were calibrated by manually identifying lateral epicondyle, lateral malleolus, the fifth metatarso-phalangeal joint and toe on the RGB static image.
2) Estimation of the joint kinematics during a gait cycle. A depth completion technique was implemented to reconstruct, during the gait trials, missing depth information by exploiting RGB information. The positions of the ankle and fifth metatarso-phalangeal joint were reconstructed by matching the 3D foot template to the dynamic point clouds applying ICP algorithm [14] while the position of knee joint centre was extracted implementing the 2D markerless protocol proposed by Balta et al., 2020.

The experimental session was conducted on ten subjects affected by Clubfoot who were asked to walk straight at a self-selected speed for six trials (three for left and three for right) in front of the camera, which was placed laterally to the walkway. The markerless data were not performed synchronously due to the infrared interference of the two systems which had the same IR working wavelength (850 nm). The computed joint angles were validated by comparing them with those obtained from manually labelled anatomical landmarks on RGB images. The accuracy of the proposed MS method was assessed in terms of offset between the two curves and waveform similarity by estimating the root mean square errors (RMSE) after removing their mean values. In addition, percentual errors of the 3D model creation in terms of foot length were computed. The accuracy of the 3D foot model reconstruction was evaluated by computing the difference in terms of the percentage error between the estimated foot length and the one manually measured on the patient.

Taking into account the limited number of views required for the creation of the 3D model, the reconstruction accuracy showed promising results, represented by a percentage error of 5.2% for the right foot and 5.7% for the left foot. The acquisition protocol is specifically designed to minimize any possible discomfort that the subject may experience during the upright static position. However, it is important to note that incorrect positioning of the subject during the static acquisition phase can introduce errors and challenges in aligning the different views. Therefore, ensuring proper subject positioning is crucial to mitigate such issues and enhance the accuracy of the reconstruction process.

By analysing the joint kinematics, the RMSE and offsets associated to the left lower limb are higher than the right one, RMSE (deg): MTP: 4.8 (R), 5.3 (L); ankle: 4.8 (R), 4.9 (L) and Offset (deg): MTP: 3.5 (R), 6.5 (L); ankle: 2.3 (R), 2.5 (L)).
The reported errors are mostly associated with the technological limitations of the RGB-Depth device employed. In particular, during high-speed movements especially during swing phase, the depth sensor of RGB-Depth cameras may fail to accurately reconstruct depth values due to motion blurs, leading to artifacts such as holes or fake boundaries. This limitation could be improved by implementing a depth sensor characterization in order to implement a more suitable depth completion techniques (e.g. inpainting-based or deep learning based models). In particular, the presence of missing points in the reconstruction process affects the accuracy of the 3D shape representation of the foot. This issue is especially represented in the Forefoot segments (Forefoot to Mid-Rear-foot points ratio = 0.26) and can lead to errors in estimating the position of joints, such as the fifth metatarsophalangeal joint (MTP5). Furthermore, the assumption of rigid segments

in the ICP algorithm is not suitable for the non-rigid nature of the human foot which undergoes shape variations and can be influenced by soft tissue artifacts. This limitation suggests the need to explore alternative approaches.
A technical limitation, that needs to be addressed, is the challenge of achieving synchronous acquisition between the Azure Kinect system and stereophotogrammetry.
From a clinical perspective, a limitation of this method is that the acquisition is performed in the sagittal plane with a lateral view. This does not allow the consideration of the kinematics of the first metatarsophalangeal joint, which is a significant indicator of gait quality compared to the fifth joint [11].

In conclusion, considering the rapid technological advancement in depth sensing, the proposed approach seems to be a very promising solution, in terms of preparing and acquisition time and effective cost, to evaluate the gait of subjects with foot deformities.

Bibliography

[1] 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, Nov. 2018, doi: 10.1080/17453674.2018.1525195.

[2] C. A. Gurnett *et al.*, "Asymmetric Lower-Limb Malformations in Individuals with Homeobox PITX1 Gene Mutation," *The American Journal of Human Genetics*, vol. 83, no. 5, pp. 616–622, Nov. 2008, doi: 10.1016/j.ajhg.2008.10.004.

[3] M. B. Dobbs and C. A. Gurnett, "The 2017 ABJS Nicolas Andry Award: Advancing Personalized Medicine for Clubfoot Through Translational Research," *Clin Orthop Relat Res*, vol. 475, no. 6, 1716–1725, Jun. 2017, doi: 10.1007/s11999-017-5290-0.

[4] M. B. Dobbs and C. A. Gurnett, "Update on clubfoot: Etiology and treatment," in *Clinical Orthopaedics and Related Research*, Springer New York, 2009, 1146–1153. doi: 10.1007/s11999-009-0734-9.

[5] "Clubfoot." https://kidshealth.org/en/parents/clubfoot.html (accessed May 01, 2023).

[6] L. Staheli, "Clubfoot: Ponseti Management [3rd Edition]." Global HELP Organization.

[7] Sobel→ Sobel, E; Caselli, MA; Velez, Z (1997). Effect of persistent toe walking on ankle equinus. Analysis of 60 idiopathic toe walkers. *Journal of the American Podiatric Medical Association,* 87(1), 17–22. doi:10.7547/87507315-87-1-17.

[8] "Foot inversion." https://ponseti.pl/foot-movements/?lang=en (accessed May 01, 2023).

[9] 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, 1-11, doi: 10.1155/2015/186780.

[10] L. Wade, L. Needham, P. McGuigan, and J. Bilzon, “Applications and limitations of current markerless motion capture methods for clinical gait biomechanics,” *PeerJ*, vol. 10, Feb. 2022, doi: 10.7717/peerj.12995.

[11] J. J. Allan *et al.*, “First metatarsophalangeal joint range of motion is associated with lower limb kinematics in individuals with first metatarsophalangeal joint osteoarthritis,” *J Foot Ankle Res*, vol. 13. 13-33, no. 1, Jun. 2020, doi: 10.1186/s13047-020-00404-0.

[12] I. Van den Herrewegen *et al.*, “Dynamic 3D scanning as a markerless method to calculate multi-segment foot kinematics during stance phase: Methodology and first application,” *J Biomech*, vol. 47, no. 11, 2531–2539, Aug. 2014, doi: 10.1016/j.jbiomech.2014.06.010.

[13] Balta, Diletta; Salvi, Massimo; Molinari, Filippo; Figari, Giulio; Paolini, Gabriele; Croce, Ugo Della; Cereatti, Andrea (2020). *[IEEE 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA) - Bari, Italy (2020.6.1-2020.7.1)] 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA) - A two-dimensional clinical gait analysis protocol based on markerless recordings from a single RGB-Depth camera.* doi:10.1109/MeMeA49120.2020.9137183.

[14] Besl, P.J.; McKay, H.D. A method for registration of 3-D shapes. 1992, 14(2), 0–256, doi:10.1109/34.121791.

[15] A. Kharb, V. Saini, Y. K. Jain, and S. Dhiman, “A review of gait cycle and its parameters,” 2011. [Online]. Available: www.IJCEM.orgIJCEMwww.ijcem.org

[16] G. Cicirelli, D. Impedovo, V. Dentamaro, R. Marani, G. Pirlo, and T. R. D’Orazio, “Human Gait Analysis in Neurodegenerative Diseases: A Review,” *IEEE J Biomed Health Inform*, vol. 26, no. 1, 229–242, Jan. 2022, doi: 10.1109/JBHI.2021.3092875.

[17] S. Affatato, “Biomechanics of the knee,” in *Surgical Techniques in Total Knee Arthroplasty (TKA) and Alternative Procedures*, Elsevier Ltd, 2015, 17–35. doi: 10.1533/9781782420385.1.17.

[18] E. S. Grood and W. J. Suntay, “A joint coordinate system for the clinical description of three-dimensional motions: Application to the knee,” *J Biomech Eng*, vol. 105, no. 2, 136–144, 1983, doi: 10.1115/1.3138397.

[19] Ge Wu; Sorin Siegler; Paul Allard; Chris Kirtley; Alberto Leardini; Dieter Rosenbaum; Mike Whittle; Darryl D D’Lima; Luca Cristofolini; Hartmut Witte; Oskar Schmid; Ian Stokes. *ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion-part I: ankle, hip, and spine., 2002, 35(4), 543–548.* doi:10.1016/s0021-9290(01)00222-6.

[20] G. Wu, S. Siegler, P. Allard, C. Kirtley, A. Leardini, D. Rosenbaum, M. Whittle, D. D’Lima, L. Cristofolini, H. Witte, O. A. Schmid, and I. Stokes. *ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion–part i: ankle, hip, and spine*. international society of biomechanics. Journal of biomechanics, 35:543–8, 05 2002.

[21] “Foot and ankle system.” https://www.physio-pedia.com/Foot_and_Ankle_Structure_and_Function (accessed May 01, 2023).

[22] Ahn, T.-K. Kitaoka, H. B.; Luo, Z.-P.; An, K.-N. *Kinematics and Contact Characteristics of the First Metatarsophalangeal Joint. Foot & Ankle International,1997, 18(3), 170–174.* doi:10.1177/107110079701800310.

[23] U. Della Croce, A. Leardini, L. Chiari, and A. Cappozzo, “Human movement analysis using stereophotogrammetry Part 4: Assessment of anatomical landmark misplacement and its effects on joint kinematics,” *Gait and Posture*, vol. 21, no. 2. Elsevier Ireland Ltd, 226–237, 2005. doi: 10.1016/j.gaitpost.2004.05.003.

[24] L. Chiari, U. Della Croce, A. Leardini, and A. Cappozzo, “Human movement analysis using stereophotogrammetry. Part 2: Instrumental errors,” *Gait and Posture*, vol. 21, no. 2. Elsevier Ireland Ltd, 197–211, 2005. doi: 10.1016/j.gaitpost.2004.04.004.

[25] A. Leardini, A. Chiari, U. Della Croce, and A. Cappozzo, “Human movement analysis using stereophotogrammetry Part 3. Soft tissue artifact assessment

and compensation," *Gait and Posture*, vol. 21, no. 2. Elsevier Ireland Ltd, 2005, 212–225. doi: 10.1016/j.gaitpost.2004.05.002.

[26] V. Camomilla, A. Cappozzo, and G. Vannozzi, "Three-Dimensional Reconstruction of the Human Skeleton in Motion," in *Handbook of Human Motion*, Springer International Publishing, 2017, 1–29. doi: 10.1007/978-3-319-30808-1_146-1.

[27] A. Cappozzo, U. Della Croce, A. Leardini, and L. Chiari, "Human movement analysis using stereophotogrammetry. Part 1: Theoretical background," *Gait and Posture*, vol. 21, no. 2. Elsevier Ireland Ltd, 2005, 186–196. doi: 10.1016.

[28] R. Kanko, E. Laende, S. Selbie, and K. Deluzio, "Inter-session repeatability of Theia3D markerless motion capture gait kinematics",2020. doi: 10.1101/2020.06.23.155358.

[29] L. Li, "Time-of-Flight Camera-An Introduction.", *Technical White Paper, Texas Instruments,* SLOA190B – January 2014. Revised May 2014.

[30] "Time of flight." https://www.seeedstudio.com/blog/2020/01/08/what-is-a-time-of-flight-sensor-and-how-does-a-tof-sensor-work/ (accessed May 10, 2023).

[31] Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: A survey of deep learning-based methods," *Computer Vision and Image Understanding*, vol. 192, Mar. 2020, doi: 10.1016/j.cviu.2019.102897.

[32] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," Dec. 2018.

[33] "Microsoft Azure Kinect Sensor SDK." https://www.microsoft.com/en-us/research/project/skeletal-tracking-on-azure-kinect/ (accessed May 10, 2023).

[34] H. D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects.", *Pattern recognition,* 34, 2259-2281,2001.

[35] Soille, P. Morphological Image Analysis, 2004. doi:10.1007/978-3-662-05088-0.

[36] Y. Xu, X. Zhu, J. Shi, G. Zhang, H. Bao, and H. Li, “Depth Completion from Sparse LiDAR Data with Depth-Normal Constraints,” Oct. 2019.

[37] Y. Zhang and T. Funkhouser, “Deep Depth Completion of a Single RGB-D Image.”, arXiv:1803.09326v2 [cs.CV] 2 May 2018.

[38] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab, “Deeper Depth Prediction with Fully Convolutional Residual Networks.”

[39] 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, Jun. 2018, doi: 10.1186/s12938-018-0518-0.

[40] *A Threshold Selection Method from Gray-Level Histograms, 1979, 9(1), 62–66.* doi:10.1109/tsmc.1979.4310076.

[41] “Microsoft. Azure Kinect DK.” https://learn.microsoft.com/en-us/azure/kinect-dk/about-azure-kinect-dk (accessed May 10, 2023).

[42] “Microsoft. Azure Kinect DK hardware specifications.” https://learn.microsoft.com/th-th/azure/kinect-dk/depth-camera (accessed May 10, 2023).

[43] “Oxford Foot Model 1.4 ii Version 1.4,” 2012. [Online]. Available: http://www.vicon.com/support/.

[44] Y. Gao, Y. Yang, Y. Zhen, and Q. Dai, “Depth error elimination for RGB-D cameras,” *ACM Trans Intell Syst Technol*, vol. 6, no. 2, Apr. 2015, doi: 10.1145/2735959.