POLITECNICO DI TORINO

Master of Science's Degree in Biomedical Engineering

Politecnico
di Torino

Master of Science's Degree Thesis

3D Markerless clinical gait analysis based on RGB and Depth sensing technology for children affected by Cerebral Palsy and Clubfoot

Supervisors:

Prof. Ing. Andrea CEREATTI

Dott. Ing. Diletta BALTA

Candidate:

Adriana DI BIASE

March 2025

Abstract

Marker-based stereophotogrammetry (MB) is the gold standard for gait analysis, but its clinical use is limited due to high costs, long set-up times, and patient discomfort. Recent advances in depth sensing and machine learning have made markerless (ML) gait analysis a promising alternative for clinical applications, especially when ease of setup is essential. The development of RGB-Depth technology further enhances ML capabilities by integrating color and depth data, thus allowing for a color point cloud reconstruction. In the literature, various single-camera ML algorithms have been proposed, including deterministic and deep learning-based approaches. However, these algorithms rely on 2D video analysis, requiring manual identification of anatomical landmarks and failing to capture out-of-plane movement. Recent advances in computer vision have led to 3D statistical models, such as the Skinned Multi Person Linear (SMPL) model, which realistically represent diverse body shapes and poses. However, its use in clinical gait analysis remains limited. This thesis proposes an original ML protocol based on a single RGB-D camera and the 3D SMPL model. Five cerebral palsy (CP) and nine clubfoot patients performed three self-selected gait trials per side, and a static posture of each participant was acquired from three different camera views (frontal, posterior, and sagittal). A SMPL consisting of foot, shank, thigh, and pelvis interconnected by ankle, knee, and hip joints was calibrated to each participant's static posture. Then, the SMPL was aligned to each dynamic frame of the gait cycle using the articulated iterative closest point algorithm to estimate 3D joint kinematics and extract seven clinical gait features. Validation was performed against a 3D MB clinical gait analysis protocol. Assuming movement repeatability, gait trials were recorded separately to avoid IR sensor interference between systems. The accuracy of the ML protocol was evaluated using Mean Absolute Difference (MAD) with respect to the MB system, while reliability and variability of both ML and MB protocols were assessed via Intraclass Correlation Coefficient (ICC) and Gait Variability Standard Deviation (GVSD). MAD values were: for the knee kinematics, 3.5° during stance, 2.8° in swing; for the ankle kinematics, 4.1° in stance, 4.5° in swing; for the hip kinematics, 5.4° in stance. ICC for ML and MB were: for the knee kinematics, 0.84 vs 0.89 in stance, 0.82 vs 0.96 in swing; for the ankle kinematics, 0.91 vs 0.86 in stance, 0.87 vs 0.90 in swing; for the hip kinematics, 0.90 vs 0.95 in stance. Mean GVSD values for ML and MB were: 4.6° vs 3.6° for the knee, 3.4° vs 1.9° for the ankle, 3.2° vs 2.3° for the hip. Residual MAD values are mainly due to asynchronous acquisitions and different anatomical axis definitions between ML and MB protocols. ICC values for the two protocols

are comparable, with the largest discrepancy observed in the knee kinematics during the swing phase due to the depth sensor's limited ability to reconstruct depth values at high speed. Mean GVSD values for ML and MB protocols are comparable, with an average difference of 1°, proving ML protocol to have similar variability to MB. In conclusion, the proposed ML protocol provides 3D joint kinematics by leveraging the SMPL model to enhance automation and overcome 2D analysis limitations. The proposed ML protocol shows strong reliability (ICC > 0.8), making it a promising solution for future clinical applications.

Acknowledgements

I would like to sincerely thank Professor Cereatti for giving me the opportunity to take part in this thesis project. I am really proud to have embarked on this journey under his guidance.

A special thank you also goes to Dr. Ing. Diletta Balta, who has been a point of reference for me. Her passion for her work has been truly inspiring, and I am deeply grateful for her constant support and availability.

I would also like to thank all the members of the DET office, whose presence has made this journey more of a pleasure than a duty.

Finally, I want to express my deepest gratitude to my parents, my family, and all my friends. Thank you for always being by my side with a word of comfort and encouragement. One thank you is simply not enough.

Table of Contents

Acronyms

AI Artificial Intelligence

AICP Articulated Iterative Closest Point

CI Confidence Interval

CNN Convolutional Neural Network

CP Cerebral Palsy

DoF Degrees of Freedom

DQBS Dual Quaternion Blending Skinning

FF Forefoot

GMFCS Gross Motor Function Classification System

GVSD Gait Variability Standard Deviation

IC Initial Contact

ICC Intraclass Correlation Coefficient

ICP Iterative Closest Point

IR Infrared

LBS Linear Blend Skinning

LED Light-Emitting Diode

MAD Mean Absolute Difference

MB Marker-based

ML Markerless

MRF Mid-rear Foot

PCA Principal Component Analysis

PSD Pose Space Deformation model

RGB-D RGB-Depth

SDK Software Development Kit

SMPL Skinned Multi-Person Linear model

TEV Talipes Equinovarus

ToF Time-of-Flight

WPSD Weighted Pose Space Deformation model

Chapter 1

1. Introduction

1.1. The importance of human motion monitoring

Gait Analysis is a field of research of growing interest, focusing on the quantitative and objective measurement of human movement. Its application in clinical settings has made it a useful tool for screening purposes, for the prediction of neurological disorders, as well as for the optimization of therapeutic strategies. Marker-based (MB) optical stereophotogrammetry is currently the most accurate and widely used approach in instrumented movement analysis [1]. Recognized as the gold standard in the field, it provides submillimeter accuracy in tracking the position of markers attached to the subject's skin and offers high temporal resolution, reaching down to milliseconds. Despite its high accuracy, the use of stereophotogrammetry is limited by several practical challenges. These include lengthy preparation times for participants, difficulties in applying the markers, physical and psychological discomfort for participants, the requirement for skilled personnel and specialized, well-equipped laboratories, as well as the high associated costs. Several technologies have emerged over the years to overcome those limitations. Particularly, markerless (ML) motion capture systems offer an alternative method for the measurement of kinematic data with several practical benefits, including reduced costs and time, easy setup and the elimination of the need to place markers on the patient, and consequently increased system portability. However, these markerless systems aren't currently in widespread use in clinical settings as their accuracy and practicality has yet to be fully investigated [2]. The potential of these systems remains crucial in applications where portability and ease of use are priorities such as screening, monitoring protocols and evaluating treatment, especially in setups that rely on a single-camera system [3]. With the recent advancement of video-based pose estimation, video registration from easily accessible devices is emerging as a valuable tool for gait analysis. In response, several manufacturers have introduced affordable tracking systems between 200 and 400 €/$ that combine a conventional color camera with an infrared depth sensor (RGB-D). By combining the information of the RGB image with depth data, these systems can be used to generate enhanced depth color images (2D+), thus allowing for a color point cloud reconstruction.

1.2. Clinical relevance of the work

Clinical gait analysis plays a crucial role in understanding and interpreting the pathophysiological characteristics of human locomotion. By providing objective and detailed information such as spatial-temporal parameters and joint kinematics, it helps to identify and quantify gait deviations, monitor their progression over time, and evaluate the effectiveness of treatments [4]. This makes gait analysis an indispensable tool for improving diagnostics and supporting clinical decision-making. Instrumental methods can provide accurate and reliable information to physicians, with proven benefits to patients through more informed decisions. However, they have several disadvantages that limit their applications in the clinical setting. As mentioned above, through the elimination of the time-consuming marker placement procedure, motion capturing experiment can be performed in a more convenient way. This makes markerless systems capable of capturing human movement more naturally, without the physical and psychological limitations imposed by markers, using low-cost, portable instrumentation. The ability to non-invasively and realistically record movement in less structured environments offers a more comfortable experience for patients, while providing accurate data useful for long-term monitoring and adaptation of rehabilitation treatments [5]. This is particularly useful in contexts, such as the clinical evaluation of patients with cerebral palsy (CP) or foot deformities, where continuous and repeated monitoring over time is required, especially when working with very young children [3].

In Sweden and most Nordic countries, CP patients are monitored over time through the Cerebral Palsy Follow-up Program (CPUP), which aims to prevent the occurrence of severe deformities through continuous and standardized surveillance, if necessary combined with early treatment, and thus to optimize the quality of life of people with cerebral palsy. CPUP provides for biannual clinical examinations for children under 6 and annual examinations for those aged between 6 and 18. However, the program does not currently include clinical instrumental gait assessment, which is a valuable resource for improving monitoring and detecting changes over time. Although highly beneficial, MB techniques cannot be implemented in these contexts due to the extensive instrumentation required, the need for specialized laboratories, high costs, and prolonged examination times. Hence the need for ML approaches such as the one proposed in this thesis, to be applied in gait cycle clinical evaluation.

1.2.1. Cerebral Palsy

Cerebral palsy, which occurs in 2-3 out of 1,000 live births, has multiple etiologies resulting in central nervous system injury that affects movement, posture, and balance [6]. It is the most common cause of disability among children and involves several motor disorders due to primary deficits, such as muscle spasticity, muscle weakness, loss of selective motor control, and secondary deficits, such as muscle contractures and bone deformities. If left untreated, soft tissue contractures, muscle weakness, pain and compensation mechanisms result in the persistence of lifelong gait abnormalities [7]. Within the group of affected children, 75% are ambulatory, presenting a wide variety of impairments and degrees of severity. Clinical classifications are based on the topography of impairments (hemiplegia, diplegia, quadriplegia), the type of motor disorders (spastic, athetotic, dystonic, hypotonic, ataxic, and a mixed group), and functional capacities assessed using the Gross Motor Function Classification System (GMFCS) [8].

Gait pattern classifications differ between unilateral and bilateral spastic CP, with most based on the observation of sagittal plane kinematics. For patients with unilateral CP, four groups of gait patterns were identified with increasing severity of impairments and a progression of impairments from distal to proximal (ankle to pelvis) going from the first to the fourth group [9]:

1. First group: foot drop (difficulty in lifting the front part of the foot) during the swing phase, resulting in a lack of first rocker at the moment of initial contact. Associated with weakness or hypoactivity of the tibialis anterior muscle relative to hyperactivity of the gastrocnemius and soleus.
2. Second group: foot drop during the swing phase and permanent plantar flexion in the stance phase, with hyperextension of the knee. Associated with static or dynamic contractures of gastrocnemius and soleus.
3. Third group: in addition to deviations of group 1 and 2, it involves limited knee flexion during the swing phase, hip hyperflexion and increased lumbar lordosis
4. Fourth group: it presents all the deviations of the previous groups, plus limited hip and knee motion. It involves compensating for the restricted hip motion with an increase in pelvic lordosis during the terminal phase of stance.

Four main groups were also identified for patients with bilateral CP based on sagittal plane kinematics, considering ankle, knee, hip and pelvis [10]:

1. True equinus: the ankle is in plantarflexion during the entire stance phase, with extended hips and knees. It may be masked by a knee recurvatum.
2. Jump gait: it is characterised by equinus at the ankle, knee and hip flexion, anterior tilt of the pelvis and increased lumbar lordosis.
3. Apparent equinus: the range of dorsiflexion at the ankle is normal, but the hip and knee are in excessive flexion throughout the stance phase leading to walking on the toes and giving an impression of equinus.
4. Crouch gait: excessive dorsiflexion at the ankle in combination with excessive flexion at the knee and hip joints.

1.2.2. Club Foot

Clubfoot or talipes equinovarus (TEV) is a congenital three-dimensional deformity of leg, ankle and foot immediately recognizable at birth. It involves abnormal positioning of the foot and ankle joint due to structural defects in several tissues of the foot and shank [11].

This malformation has an incidence of approximately 1 in 1,000 live births and may affect one foot only (unilateral TEV) or both (bilateral TEV). In 50% of cases, the deformity is unilateral, with a predilection for the right foot [12]. The condition affects males more frequently than females, with a male to female incidence ratio of 2:1 [11]. Clubfoot is not passively correctable and, if it is left untreated, it could provoke infections, foot and leg deformities, pain, and limits mobility [13].

Gait features in clubfoot patients show typical alterations [14] [15]:

- Toe walking (equinus gait): patients exhibit a predominantly toe walking, with little or no heel contact with the ground. This is due to the limited dorsiflexion of the foot during the stance phase.
- Foot inversion: during the swing phase, the foot may rotate internally, with the sole tilting towards the midline of the body. This alteration leads to an abnormal positioning of the foot, reducing overall stability during walking.
- Intoeing: there is often an internal angle of progression of the foot, resulting from internal rotation of the tibia or structural deformities in the axis of the foot.
- Limping: due to limited range of motion and abnormal posture, the subject with a significant clubfoot deformity may show an important limb.

- Shortened stride length: stride length is often shortened due to limited mobility and flexibility of the affected foot.

1.3. Optoelectronic stereophotogrammetry

The optoelectronic stereophotogrammetric multi-camera capturing system is the gold standard for motion analysis, tracking markers placed on the body [16]. Through this instrumental methodology, it is in fact possible to reconstruct the 3D position of the markers applied to the subject instant by instant. In this way, optoelectronic stereophotogrammetry can be exploited to perform kinematic analysis of anatomical segments on which the markers are placed.

The markers must be positioned to ensure they are visible to the cameras and can be either active or passive. Passive markers are generally hemispherical plastic holders coated with reflective material while active markers are infrared (IR) light-emitting diodes (LED) that generate the light signal themselves. The former are used in combination with infrared stroboscopic illumination produced by a series of LEDs mounted around the lens of each camera. The use of infrared light, combined with an IR filter placed on the camera lenses, improves the accuracy of marker detection, reducing interference from visible light and enabling accurate marker recognition through threshold-based image processing techniques [17]. In contrast, active markers produce light at a given frequency, so these systems do not require illumination, and, as such, the markers are more easily identified and tracked. Unlike passive markers that reflect incident light, active markers emit light at a specific frequency, eliminating the need for illumination. Active markers are easier to identify and track as each can emit light at a specific frequency, allowing the system to easily distinguish them even if they are close together. Additionally, they are illuminated one at a time, making the marker correspondence in the images more reliable. These features ensure accurate and stable real-time three-dimensional tracking, improving system reliability and minimizing identification errors [18]. However, each active marker requires a power supply and a control unit. Additionally, if the markers are activated at the same frame rate as the cameras, the effective frame rate for each marker is reduced, as it is divided by the number of markers in use.

The marker tracking procedure involves several steps, illustrated in the following. The first step is the calibration process which aims to determine the intrinsic parameters, such as focal length and optical centers, and relative positions of the cameras, as well as to define a global reference frame for expressing the coordinates of the markers. After proper calibration, the cameras capture 2D images, which are then processed to identify the markers. Techniques such as linear

thresholding are used to distinguish markers based on brightness, while morphological operators help refine their shapes by eliminating noise or filling gaps. Once identified, the markers are labelled to ensure consistent recognition across all images acquired over time, and their two-dimensional coordinates in the image planes are extracted. The second step involves converting the 2D coordinates of the markers, obtained from the cameras, into 3D coordinates using triangulation technique. This process estimates the depth by observing the target from different perspective [19] and determines the exact 3D position of each marker in the global reference frame, based on the calibrated positions and orientations of the cameras. At least two cameras are needed to perform this task, but usually up to 50 cameras are used, especially for the analysis of complex movements.

Once the 3D coordinates are calculated, it becomes possible to define a local reference frame for each body segment, which allows the relative positions and orientations of the joints (joint poses) to be determined. In the final step, the calculated 3D positions and joint poses are integrated with a biomechanical model of the human body. This model, which represents the body as a system of rigid segments connected by joints, is used to compute important motion parameters, such as the trajectories of body segments, their angular velocity, acceleration, and the kinematics of joints.

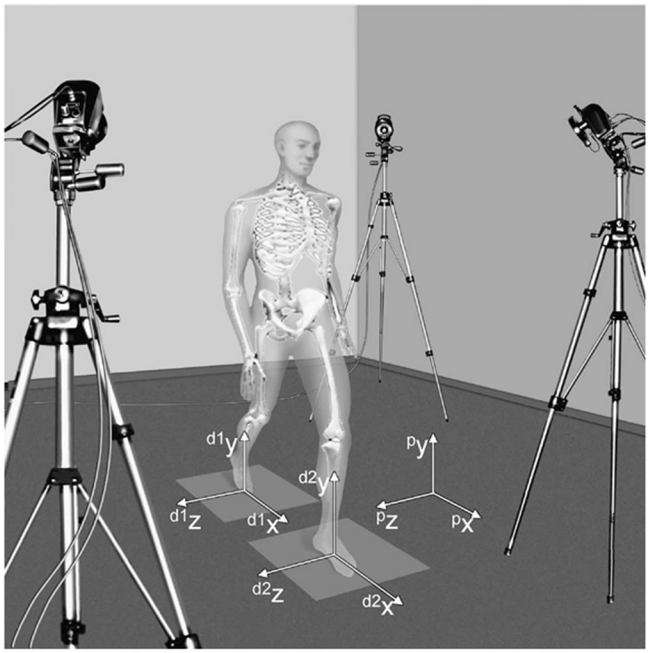

***Figure 1**: Illustration of an optoelectronic stereophotogrammetric multi-camera capturing system for 3D biomechanical analysis.*

Despite its measurement accuracy, optoelectronic stereophotogrammetry still has certain limitations due to [17]:

- Instrumental errors: result from instrumental noise and inaccuracies in volume calibration. Noise can be reduced with low-pass filtering, while inaccuracies in calibration depend on the number of cameras and the algorithm used. An insufficient number of cameras or an inadequate algorithm can cause errors in the definition of the acquisition volume, compromising the accuracy of the system. However, this type of error is usually very small, almost negligible.
- Soft tissue artifacts: occur when markers applied to the skin follow the movements of the skin itself, which does not exactly follow bone movements due to its greater deformability and mobility than the underlying bone structures. Since this error has the same frequency as bone movements, these cannot be removed by filtering. However, the effect can be reduced by choosing strategic positions for markers to minimize relative displacements and by using mathematical methods to estimate bone position.
- Markers misplacement: occur when markers are not positioned correctly with respect to anatomical landmarks, this type of error is strongly influenced by operator experience. Markers misplacement may result in wrong clinical interpretations of the estimation.
- Limited availability: Due to the required instrumentation, optoelectronic stereophotogrammetry is only available in specialized laboratories with a limited acquisition volume, which allows the analysis of only a limited number of gait cycles and precludes the monitoring of patients during daily activities. Its limited availability is also associated with high costs and long preparation times for the subject to be analyzed.

1.4. Working principles of RGB-D technology

RGB-D cameras, with their ability to capture depth and generate 3D point clouds, have revolutionized computer vision, finding application in fields such as gaming, robotics, and healthcare.

Microsoft Kinect marked a pivotal moment in the evolution of RGB-D sensors, making this technology widely accessible and popular. From the launch of Kinect v1 in 2010, through Kinect v2 in 2013, to Azure Kinect in 2019, Microsoft laid the foundations for the development of RGB-D technology, originally designed for gaming but later expanded to a wide range of research and innovation applications.

***Figure 2**: Progression of Microsoft Kinect RGB-D sensors: Kinect v1, Kinect v2, and Azure Kinect. Each iteration introduced advancements in depth sensing technology, improving accuracy, resolution, and tracking capabilities.*

At the core of this recent technology are fundamental working principles that enable both depth and 3D point cloud reconstruction. This section provides an overview of these principles.

1.4.1. Depth image reconstruction

RGB-D sensors employ three primary technologies for depth measurement, based on distinct physical and computational principles. These approaches include passive stereo reconstruction, active structured light, and time-of-flight (ToF) detection, providing different solutions tailored to specific application requirements [20].

- **Passive stereo reconstruction:** Like human vision, stereo approaches use two cameras positioned at a distance like the spacing between human eyes to capture observations of the scene from two slightly different viewpoints. By analyzing these two perspectives, stereovision algorithms can reconstruct depth information of the scene.

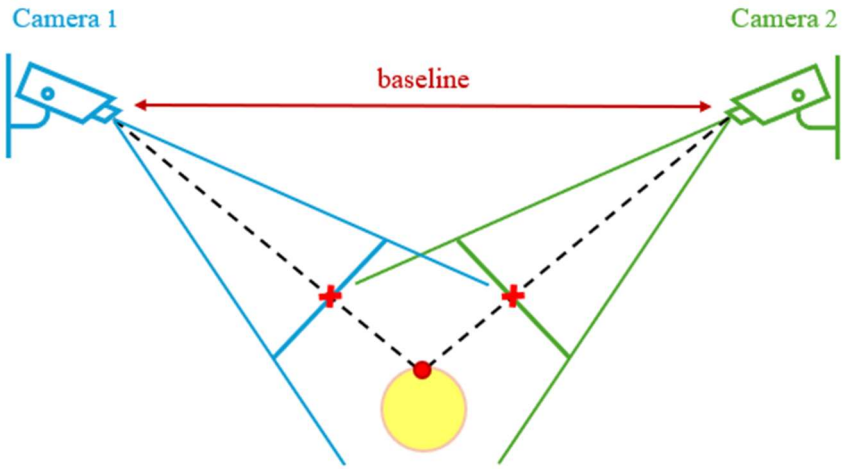

***Figure 3**: Stereo reconstruction. Stereo approaches use two cameras that capture the same scene from different perspectives, creating slight disparities in the images obtained. Once computed the corresponding points in both images, the 3D position of the point can be found using triangulation, i.e., by intersecting two rays cast through the detected point correspondences.*

The term 'passive' refers to the fact that passive sensors determine the depth of a scene without actively modifying or interacting with it, such as by emitting light or signals. This approach has been mainly implemented in devices like Intel RealSense cameras.

The most challenging process to perform in stereovision algorithms is the search for correspondences between camera views, i.e. identifying the pixels that in the two acquired images correspond to the same 3D point within the scene. A key step that simplifies this search is the rectification process. After determining the intrinsic parameters of the cameras and their relative positions through calibration, the stereo image pairs are realigned so that the conjugated epipolar lines (the lines along which corresponding points in both images lie) become horizontal and parallel [21]. This transformation reduces the search for matches from a two-dimensional area to a single direction along horizontal lines, making correspondence matching more efficient.

At this point, techniques such as feature matching, which analyses unique features within the images, and such as block matching, which compares large blocks of pixels across images, allow matches to be identified. Differences in position between matched features or blocks, known as disparities, are exploited for depth calculation: large disparities indicate a closer point, while smaller disparities indicate a more distant point within the scene.

Having identified on the images the corresponding point pair, the 3D coordinates of the point within the scene are reconstructed by triangulation, which involves using the baseline distance between the cameras and their focal lengths.

Passive stereo reconstruction works well in textured areas, where variations in intensity and color help find correspondences. However, in featureless regions, stereo matching may fail, leading to gaps in depth information or reduced accuracy. Active depth sensing approaches aim at alleviating this problem.

- **Structured Light:** This technique relies on the principles of stereo reconstruction to generate a depth map. Being an active system, it replaces one of the two cameras typical of stereovision with a projector, which functions as an inverse camera, emitting light instead of capturing it. The projector casts a known pattern using infrared light onto the scene, introducing artificial features that simplify point matching and allow depth estimation even in areas without texture or color variations.

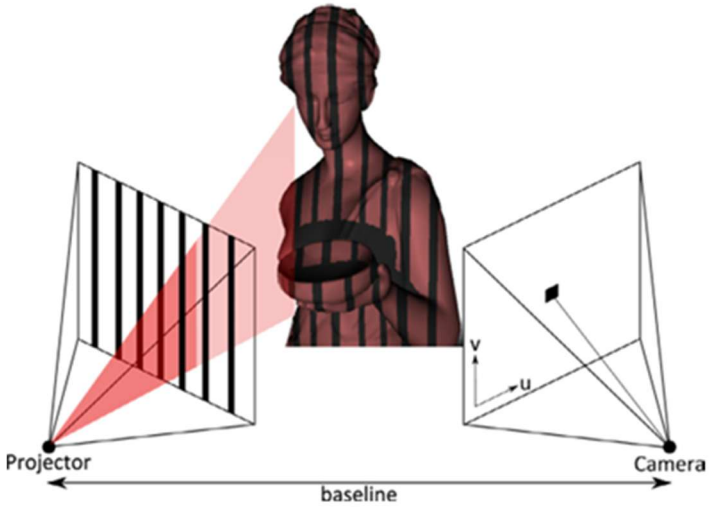

***Figure 4**: Structured light principles. This technique employs a camera and an infrared projector that casts a predefined pattern onto the scene. The projected pattern introduces artificial features, facilitating point matching and enabling accurate depth estimation via triangulation [22].*

Since it relies on triangulation, like stereo reconstruction, its accuracy decreases as the distance from the scene increases. Another limitation is its sensitivity to sunlight's infrared rays, which can saturate the sensor, making the projected pattern indistinguishable and resulting in missing depth values. This technology was exploited by Kinect v1 and allowed the introduction of controller-free gaming and basic gesture recognition.

- **Time of Flight (ToF):** In contrast to stereo vision and structured light, Time of Flight cameras are based on a different physical measurement principle. ToF sensors operate by emitting a pulse of infrared light, which travels until hitting an object and being reflected to the camera sensor which detects its arrival. The system measures the time it takes for the light pulses to return after striking the target, allowing for the calculation of the object's distance based on the speed of light. Additionally, it can estimate distance by analyzing the phase delay between the emitted and reflected signals.

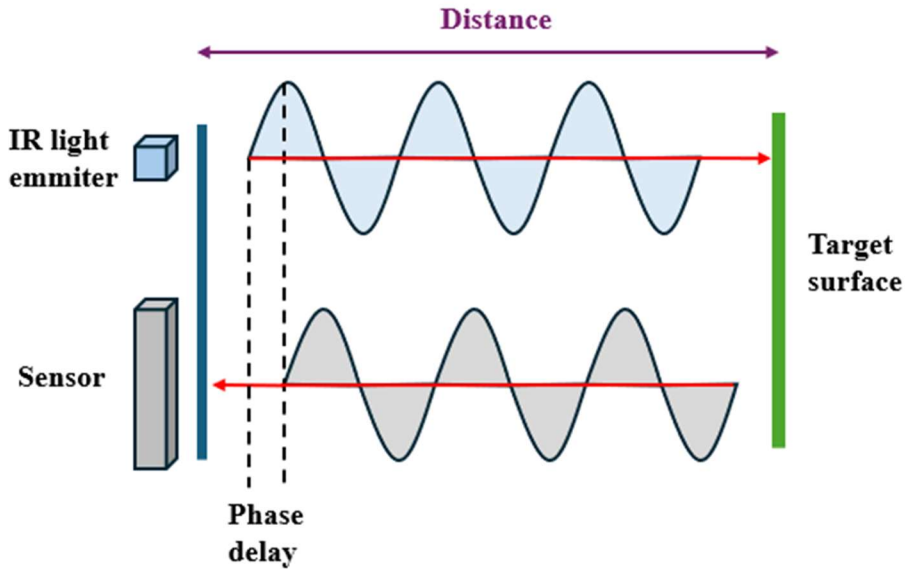

***Figure 5**: ToF principles. An IR light emitter projects a pulse of light towards a target, which reflects the signal back to the sensor. The system estimates depth by analyzing either the phase delay or the time delay between the emitted and reflected signals.*

By adopting this technology, the Kinect v2 provided significant advantages with respect to the Kinect v1 including higher output resolution of the depth data, a wider field of view for both color and depth cameras, lower noise level, less sensitivity to lighting, and more constant depth measurement precision across the operating range [23]. In addition, the Kinect v2 can track up to 25 joints of 6 users simultaneously [24] proving to be better at detecting users' positions and movements.

The release of the Azure Kinect introduced further advances, expanding the application of these technologies in different areas. While previous models of Kinect were originally developed for gaming and entertainment, the Azure Kinect was adopted in broader areas, from 3D reconstruction to healthcare. With upgraded ToF depth sensors, it offers better quality depth data with higher spatial resolution, suitable for a range of environments and lighting conditions. Currently, the Azure Kinect hardware includes: a 12-megapixel color camera, a 1-megapixel ToF depth camera, an inertial measurement unit and a seven-microphone circular array [23].

The transition from Structured Light to Time-of-Flight (ToF) represents a significant advancement toward more versatile solutions. ToF technology offers ease of use, fast acquisition, and robust performance, even in outdoor environments. Unlike Structured Light, which is sensitive to ambient light and requires complex processing to interpret pattern deformations, ToF directly measures the light’s return time, enabling accurate and rapid depth estimation. This makes it particularly suitable for real-time applications,

such as human motion monitoring. Additionally, thanks to its greater range of accuracy, ToF delivers more consistent and reliable results over extended distances, making it a preferred choice for various operational scenarios.

1.4.2. 3D point cloud reconstruction from depth data

From the depth data, RGB-D cameras can reconstruct a 3D point cloud of objects within the camera's field of view. A point cloud is a data structure that represents a set of multi-dimensional points, where each point is defined by its 3D spatial coordinates. To reconstruct 3D coordinates, it is necessary to know the intrinsic parameters of the camera, obtained through calibration, such as the focal length and the position of the principal point, which is the projection of the optical center onto the image plane. Given these parameters, the point cloud is generated using the following equations:

$$X = \frac{p_x - pp_x}{f_x} * D(p_x, p_y)$$

$$Y = \frac{p_y - pp_y}{f_y} * D(p_x, p_y)$$

$$Z = D(p_x, p_y)$$

Where X, Y, and Z represent the 3D coordinates of the object corresponding to the position of a point referred to the 3D image reference system. The variables p_x and p_y denote the locations of a pixel on the 2D image plane. The focal lengths f_x and f_y correspond to the camera's focal length along the x-axis and y-axis and pp_x and pp_y represent the coordinates of the principal point along the x-axis and y-axis, respectively, and are crucial parameters for mapping 2D pixel locations into the 3D image reference system. $D(p_x, p_y)$ refers to the depth value obtained at the pixel location p_x, p_y.

1.5. Aim of the thesis

The aim of this thesis is to propose a 3-dimensional deterministic ML gait analysis protocol using a single RGB-D camera to estimate the lower limb 3D joint kinematics for clinical applications. Already proposed in 2023 in [25], such an approach is now extended with 3D kinematic analysis instead of 2D, overcoming the limitations of two-dimensional modelling. In particular, the new methodology allows for the compensation of out-of-plane movements and changes in the subject's position relative to the camera that can occur during gait, improving

the accuracy in the estimation of joint kinematics of the lower limb in the sagittal plane. The method is primarily designed for children with cerebral palsy and foot deformities, addressing the need for non-invasive gait analysis.

The goal is to overcome practical limitations associated with MB approaches and to validate the proposed system against stereophotogrammetry, which represents the gold standard for instrumental gait analysis.

The proposed protocol consists of four steps:

1. *3D static posture creation of the subject* - At this preliminary stage, the subject was identified from the acquired images, isolating the region of the lower limbs. The point-clouds for the lower limbs, extracted from the RGB-D camera, were subsequently merged to generate a complete three-dimensional reconstruction of the subject's static posture for both right and left lower limbs.
2. *Skinned Multi-Person Linear (SMPL) model calibration* – The 3D SMPL model used in this protocol, was adapted to the specific characteristics of the subject by means of a scaling and translation procedure to ensure fitting with the previously obtained subject's 3D static posture. In addition, through the application of the Articulated Iterative Closest Point (AICP) algorithm, the pose of the subject-specific SMPL model was modified to fit that of the 3D static posture. This was achieved by minimizing the distances between the model points and those of the 3D static pose, resulting in an accurate alignment for the subsequent dynamic analysis.
3. *Joint centers trajectories tracking* - The objective of this phase was to estimate the positions of the subject's lower limb joint centers (hip, knee and ankle) during walking. To this end, the subject-specific SMPL model was fitted to each acquired dynamic frame using the AICP algorithm.
4. *Estimation of joint kinematics in the sagittal plane*- From joint trajectories, joint kinematics in the sagittal plane were extracted by calculating the angles between the segments connecting the joint centers.

Chapter 2

2. State of the Art

2.1. Markerless approaches for clinical gait analysis

With the recent advancements in computer vision and deep learning, a wide range of algorithms for markerless human pose estimation have been developed, which typically extract features directly from single or multiple 2D image planes. These techniques are generally divided into two categories: deterministic and AI-based approaches [26].

2.1.1. Deterministic approaches

Deterministic methods for human pose estimation are based on precise mathematical formulas, well-defined anatomical rules and clear principles, guaranteeing replicable and reliable results. Unlike AI-based approaches, which improve their performance as more data become available, deterministic methods maintain their effectiveness without the need for large amounts of information. This is advantageous in clinical settings, where data acquisition can be complex, especially in the case of rare conditions or restricted pathological populations. Their applicability allows for standardized and comparable evaluations over time, ensuring consistency in results and facilitating patient monitoring. Due to their rigorous nature, these methods offer a high degree of reliability and can be used with confidence, as they are based on established anatomical relationships, without the risk of unexpected variations.

Deterministic approaches can be divided into model-based and model-free methods. The former employ a predefined model based on a-priori information such as specific motion and context. The process of pose recovery in model-based methods involves two sequential phases: modelling and estimation [27]. The modelling phase consists of defining a biomechanical and articulated model of the human body, consisting of segments representing bones and joints. The model is characterized by specific parameters, such as limb lengths, joint constraints and possible ranges of motion, which must be tuned accurately to realistically simulate the movement of the human body. The estimation phase involves fitting the biomechanical model to the acquired data through adjustments and optimizations of positions and orientations. This is achieved using algorithms that minimize the error between the model and the observed data. At the end of this process, the adapted model can be used to compute biomechanical parameters of interest, such as joint angles, velocities, and accelerations. Most deterministic model-based

approaches using a single-camera setup are limited to 2D model reconstruction, allowing motion analysis within a single plane. While 3D models can be obtained, they typically rely on multi-camera systems, which, despite providing greater accuracy, are often impractical for clinical applications due to their complexity, cost, and spatial constraints.

Model-free approaches attempt to capture skeletal and movement characteristics in the absence of an a-priori model, starting from data and exploiting the pre-defined anatomical proportions of humans. These methods analyze the subject's silhouette extracted from the acquired images, utilizing predefined anatomical proportions to infer the underlying skeletal structure. By exploiting geometric relationships between body segments, such as limb length ratios, they can approximate joint locations within the silhouette without requiring explicit deformation of a 2D or 3D template, as required in model-based approaches.

In the context of clinical applications, despite their greater computational complexity and the need for detailed anatomical models, model-based methods are preferable as they ensure higher accuracy in representing body kinematics. This is primarily due to the incorporation of anatomical constraints, which enhances the reliability of motion analysis, particularly in the estimation of internal joint positions that are not directly visible from the external silhouette. Unlike model-free approaches, which rely on silhouette extraction and assumed body proportions, model-based methods integrate biomechanical principles to prevent unnatural joint configurations.

Mündermann and Ceseracciu ([26], [28]) proposed a model-based markerless approach based on visual hulls from multiple image streams to provide an accurate 3D estimate of human motion. Their approach involves reconstruction of the human pose from multiple RGB cameras mounted around the subject simultaneously capturing at 75 fps from different views. By applying thresholds on intensity and color, the subject was separated from the background in each image captured by the cameras. The silhouettes of the subject extracted from each camera were then combined to obtain a volumetric model of the subject, known as the visual hull. The latter represents the maximal volume consistent with the silhouettes acquired from different angles, providing an approximation of the subject's body shape. At this point, a subject-specific 3D model, consisting of 15 body segments and 14 joints connecting them, was fitted to the Visual Hull in each frame using the Articulated Iterative Closest Point (AICP) [29] algorithm to accurately estimate joint motion. AICP algorithm is a generalization of the standard Iterative Closest Point (ICP) algorithm for articulated body, providing for the application of joint

constraints that guarantee realistic movements. Treating the visual hull as a point cloud, the protocol involves point cloud registration and then using AICP to minimize the distance between the model's joints and the visual hull surface. The alignment was progressively refined through iterations until a precise fit of the model to the movements acquired via visual hull was obtained. Finally, the positions of the joint centers were extracted and the corresponding joint kinematics calculated.

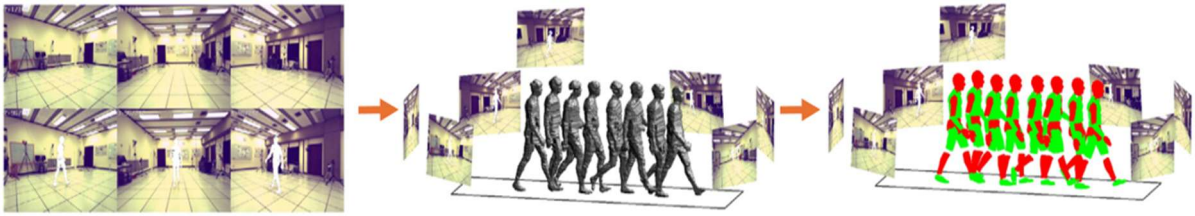

***Figure 6**: Model-based ML protocol proposed in [26]. Video sequences capture the subject from various viewpoints, from which visual hulls in motion are reconstructed from multiple angles, allowing an articulated body model to be matched to these hulls for joint localization.*

Despite their 3D reconstruction capability, clinical gait analysis systems employing multiple cameras may be more limiting than those based on a single camera, due to the complexity of installation, specific calibration procedure and need for considerable dedicated space.

In this context, following the approach proposed in [26], Castelli and Panztar-Castilla proposed model-based ML methodologies using a single camera. In particular, Castelli et al. [30] introduced an ML protocol for the extraction of 2D lower limb joint kinematics using a single RGB camera. The method involves extracting the subject from the images and from its silhouette, a subject-specific multi-segmental model of the lower limb (foot, tibia, femur and pelvis) is constructed, calibrated on a static sagittal acquisition through the manual identification of key anatomical points. For each acquired dynamic frame, the subject silhouette is extracted with a bottom-up approach and from this the key points are identified. These points are then used to align the reference system of the subject-specific model with that reconstructed for each frame via singular value decomposition, ensuring accurate joint kinematics calculation. However, since this method was validated against the MB system on twenty healthy subjects, its clinical validity for pathological patients remains uncertain. Panztar-Castilla et colleagues [31] proposed an improvement to this approach, using a single RGB-Depth camera rather than a standard RGB camera. The use of depth data enabled improved body segmentation and joint centers tracking by no longer relying solely on the contours of the extracted silhouette, thereby reducing errors due to body segments overlapping during walking. Moreover, Panztar-Castilla et al validated the proposed protocol on patients with cerebral palsy (CP), demonstrating its

potential clinical applicability for long-term follow-up and monitoring of CP patients. However, the protocol relies on the use of a green background to facilitate subject segmentation, resulting in additional equipment and less system's portability. Additionally, although proper validation of the ML system against the MB stereophotogrammetric system was conducted, this was limited to knee joint kinematics only, failing to provide a complete view of the entire gait cycle, which is essential for accurate diagnosis and treatment planning.

To address the limitations in the clinical applicability of previously proposed algorithms, Balta et al. proposed in [25] an ML method that utilizes a single RGB-D camera and a 2D muti-segmental model for estimating 2D joint kinematics and spatial-temporal parameters. Compared to the Pantzar-Castilla approach, this method does not require the use of a uniform background for subject segmentation but introduces an automatic segmentation algorithm based on a greyscale histogram approach, which allows the applicability of the protocol to be extended to different environments. Also in this case, a subject-specific 2D kinematic model was introduced, consisting of 4 body segments (foot, shank, thigh and pelvis) connected by revolute joints (ankle, knee and hip) for a total of 6 degrees of freedom (DoF). Joint centers trajectories were estimated by aligning the 2D model of the lower limb with dynamic data using ICP, exploiting depth data and applying a statistical analysis with the Otsu method to handle the overlap of body segments. As the size and shape of the body segments vary during the gait cycle due to soft tissue deformation and changes in the subject's position relative to the camera, a multiple calibration procedure was implemented based on three sets of body segment models, the first defined from standing posture, the second and third from frames selected during the load and swing phases of the gait cycle, respectively. Consequently, the procedure for identifying joint centers trajectories was repeated using the additional templates, thus resulting in three different trajectories for each center.

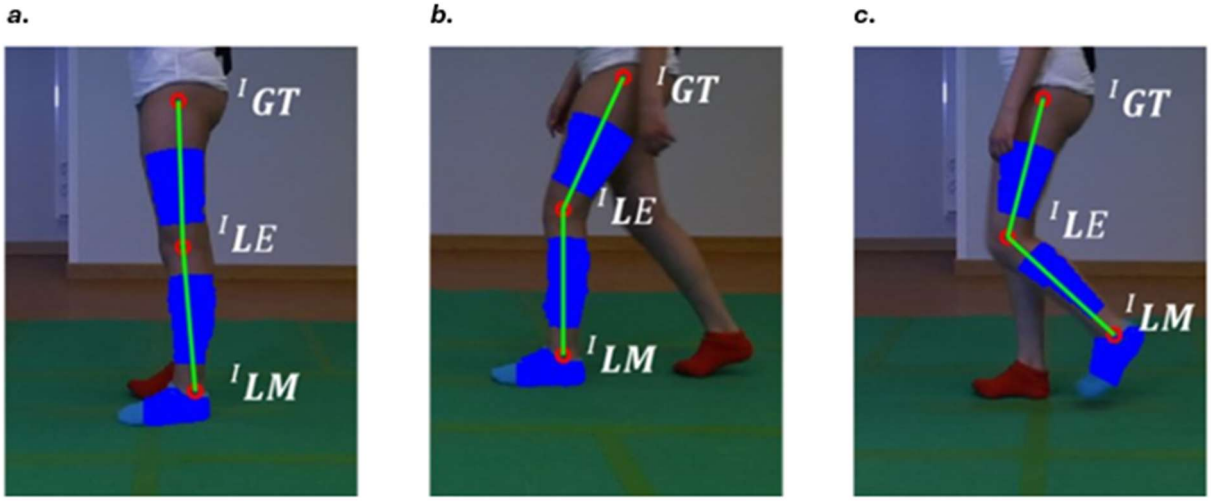

***Figure 7**: Multiple calibration procedure, based on manual identification of anatomical landmarks (greater trochanter GT, lateral epicondyle LE, lateral malleolus LM). Three different body segment templates were derived from the calibration process during (a) static phase, (b) loading phase and (c) swing phase [25].*

However, joint centers trajectories are reconstructed from a 2D model derived from a 2D RGB image, as depth information is only exploited for the segmentation of foreground segments, thus implying projection of human body 3D motion to a 2D space and leading to ambiguities that can only partly be solved by multiple calibration. Despite this limitation, the proposed ML method was tested on 18 CP patients and validated against the MB stereophotogrammetric system, demonstrating good-to-excellent reliability in the assessment of hip, knee, and ankle joint kinematics as well as spatiotemporal parameters.

2.1.2. AI based approaches

AI-driven gait analysis has shown promising performance, but its implementation in clinical practice is still in its infancy. Artificial intelligence (AI) based approaches automatically extract motion features from visual data and efficiently handle complex, high-dimensional, and temporal data, learning motion patterns directly from large datasets to generate accurate predictions. Like deterministic methods, these approaches can either exploit the use of a predefined model or estimate the human pose directly from the data. Although they generally adopt a model-free strategy, there are cases where AI-based methods incorporate a model-based approach to achieve greater accuracy in the estimation of joint position.

Several methods proposed in the literature exploit artificial intelligence, which can be implemented mainly through two types of approaches: proprietary black-box methods, which include solutions based on software development kits (SDKs) integrated with proprietary hardware or commercial software (e.g. IPsoft iPi Biomech, MediaPipe Studio), and open-source methods, which are based on deep learning techniques.

A key example of existing proprietary black-box method is the Azure Kinect body tracking SDK which accurately reconstructs, even in real time, the positions of the 3D joint centers by exploiting deep learning, a convolutional neural network (CNN) and the optimization of a 3D skeletal model with 32 joints. In addition to providing the 3D coordinates of each of the 32 joints, the SDK also returns the confidence level for each of them, indicating the reliability of the estimate. The procedure that the Azure Kinect body tracking SDK employs involves several steps, illustrated in the following.

First, a 2D estimation of joint center positions is obtained using a CNN that processes infrared images. These bidimensional positions are then combined with depth sensor data to transition from 2D to 3D, obtaining an initial rough 3D pose estimation. Predefined information regarding anatomical proportions and biomechanical constraints is then applied to ensure a plausible estimate of the human pose. At this stage, a kinematic model is introduced and fitted to the estimated pose using an optimization algorithm that minimizes the distances between the 3D joint positions predicted by the CNN and those of the skeletal model. During the optimization of the skeletal model, regularization terms are applied to ensure realistic poses, maintain consistent proportions, limit joint movements to plausible anatomical values and ensure temporal stability between consecutive frames. This helps when CNN's predictions might be less reliable due to occlusions or ambiguities in depth data.

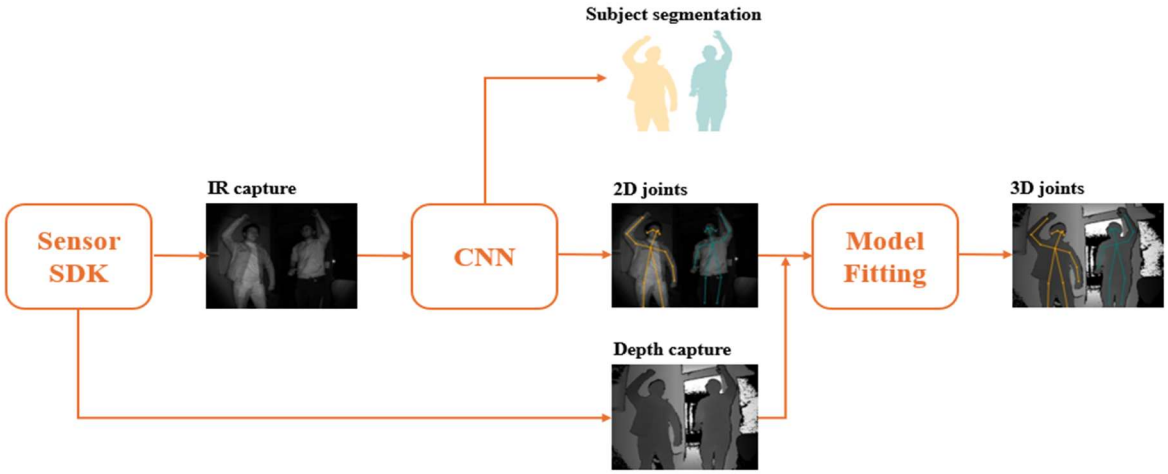

***Figure 8**: 3D reconstruction process of joint positions by Azure Kinect body tracking SDK [32].*

Despite their potential, proprietary black-box methods have several disadvantages that limit their application in clinical gait analysis, including their lack of transparency, interpretability and customization for pathological data, which are key aspects in clinical settings.

Furthermore, being initially designed for gaming, they do not conform to medical standards and terminologies and are often tied to specific hardware, making their generalization difficult.

On the other hand, open-source methods exploit deep learning offering continuous improvement due to the increasing availability of large datasets and the potential of Convolutional Neural Networks (CNN). Compared to black-box methods, these guarantee high adaptability as they allow parameters tuning for the representation of clinical cases of interest. To date, most AI-based algorithms for human pose estimation are based on neural networks relying on deep learning. In particular, CNNs have been widely used to estimate 2D keypoints from a single image, enabling 2D markerless human pose estimation.

Mathis et al. proposed in [33] an advanced machine learning ML keypoints tracking framework, DeepLabCut. Although originally designed for video identification of predefined key points on animals, its application can be extended to human pose estimation thanks to the implementation of transfer learning. This technique involves the use of a pre-trained CNN on a broad dataset that does not necessarily contain information related to the current task, and then using a much smaller set of targeted information to fine-tune the algorithm for the specific application, such as human pose estimation. Thanks to this distinctive feature, DeepLabCut has established itself for its ability to obtain accurate results with a significantly reduced amount of data, guaranteeing gait analysis accuracy comparable to MB methods and distinguishing itself for its high flexibility in adapting to different contexts. To accomplish this task, it is essential to select from a large dataset only the most relevant images, in which the body parts to be tracked are manually and consistently annotated. These images should be as diverse as possible to effectively capture the breadth of the experimental data to be analyzed and enhance the model's ability to generalize across different scenarios. Techniques like k-means clustering can be used to group similar images, facilitating the selection of a diverse and representative subset of images for training. Finally, the pre-trained network is fine-tuned using the selected images, enabling it to accurately recognize the annotated body parts in similar contexts.

From these 2D deep learning-based key point tracking frameworks several approaches for lifting 2D to 3D human pose estimation.

In this context, Pavllo et al. [34] introduced an approach based on a 1D temporal CNN and semi-supervised training. This training approach represents a significant innovation as it enables the network to learn 3D human poses even in the absence of annotated 3D training data, which are costly and limited compared to 2D data. In fact, it combines supervised training on

annotated 3D datasets with self-supervised training, in which unlabeled data is used for learning by projecting the predicted 3D pose to 2D and minimizing the discrepancy with the detected key points. As a first step, 2D keypoints for an unlabeled video with an off-the-shelf 2D keypoint detector are identified. A temporal dilatated convolutional network, which takes as input the temporal sequence of 2D joints coordinates (key points) extracted from videos, is then used to capture relationships between past and future frames and model long-term dependencies. By leveraging this temporal information, the algorithm can predict stable 3D pose estimates, correcting errors in tracking and interpolating hidden joints in occlusions.

Another method was proposed by Omran et al. [35], who introduced Neural Body Fitting, one of the first approaches integrating the training of CNNs with the fitting of a parametrized 3D human model, the Skinned Multi person linear model (SMPL), in a single framework. A standard semantic segmentation CNN is firstly used to obtain a colour-coded part segmentation map (12 semantic parts). This segmentation serves as input for an encoding CNN which predicts the body shape and pose parameters, β and θ, of the SMPL model which, in turn, produces a 3D mesh of the acquired subject. The training process admits both full 3D supervision and weak 2D supervision, when images with only 2D annotations are available. This is implemented by reprojecting the predicted 3D joints coordinates onto the 2D image plane to evaluate and minimize the loss function in 2D space, thus reducing dependency on annotated 3D dataset. By leveraging a 3D human body model, Neural Body Fitting approach maintains structural consistency and body proportions, unlike purely data-driven methods, which may produce geometrically incoherent keypoints.

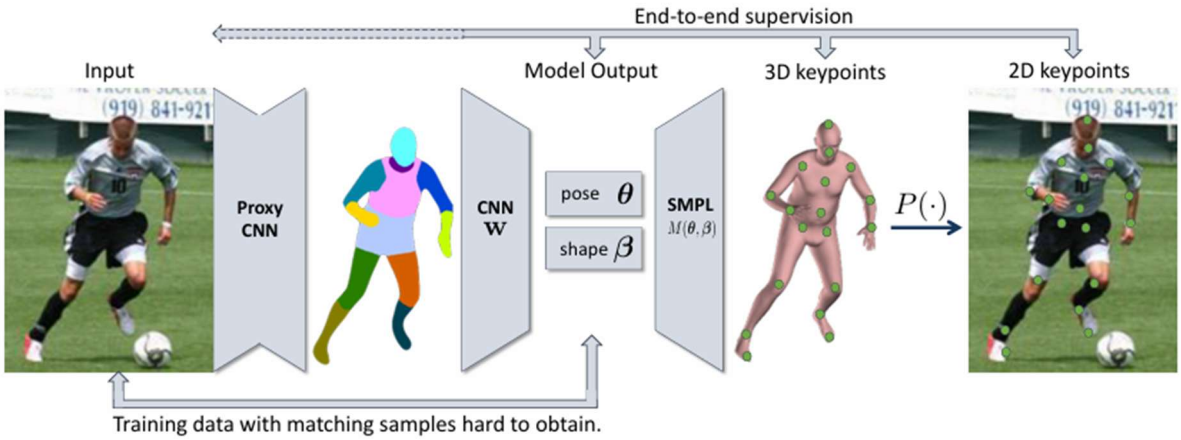

Figure 9 : *Pipeline of the Neural Body Fitting approach proposed by Omran et al. [35].*

Despite representing promising approaches in gait analysis, most deep-learning based algorithms are often trained on synthetic and generic motion data, which are not necessarily

specific to gait and do not conform to established clinical standards for gait analysis. A further limitation of open-source methods is the lack of data from pathological patients in the training sets, which limits their representativeness and compromises their clinical validity.

2.2. 3D animation and deformation models

To better understand the context of 3D model-based ML human pose estimation, this paragraph is presented to analyze some of the 3D models used in the representation and animation of body structures.

The most common approach to accurately represent the movements of the human body within 3D deformation models is to transfer the movements of an underlying skeleton to the deformation of the external mesh. One of the main challenges of animated feature films, computer games, and interactive applications is to create believable and compelling skin deformations. Traditionally, bone transformations describe the position and orientation of the joints, and the skin deformation is computed by linearly blending bone transformations to the skin. This technique is Linear Blend Skinning (LBS), also known as "skeletal subspace deformation" [36] and it's the basic and most well-known algorithm for direct skeletal shape deformation [37]. In this approach the skin is represented as a polygonal mesh, while the skeleton is modeled as a rooted tree, with its nodes corresponding to the joints.

The fundamental parameters governing deformation in the LBS model include rest pose shape, bone transformation and skinning weights. These elements work in synergy to map the movements of the skeleton to the surface of the model, allowing for a dynamic and deformable representation of the body structure.

- Rest pose shape represents the undeformed reference configuration of the model and is typically defined by a polygon mesh. During the animation process, the rest-pose vertices, $\boldsymbol{v_1}, \ldots, \boldsymbol{v_n} \in \mathbb{R}^3$ ($or\ \mathbb{R}^4$in homogeneous coordinates) may change position, but the connectivity of the mesh remains unchanged, ensuring the structural consistency of the model.
- Bone transformations are represented by a list of matrices, $\boldsymbol{T_1}, \ldots, \boldsymbol{T_m} \in \mathbb{R}^{3x4}$. Each matrix, $\boldsymbol{T_i}$ corresponds to the spatial transformation that aligns the rest pose of the *i*-th joint with its current animated pose. These components are the only ones allowed to vary during the animation process.

- Skinning weights, $w_{i,1}, \ldots, w_{i,m} \in \mathbb{R}$, are associated to each rest pose vertex $\boldsymbol{v_i}$. Each weight $w_{i,j}$ represents the amount of influence of joint j on vertex i.

During the animation process a rest pose vertex, $\boldsymbol{v_i}$, is transformed as follows:

$$\boldsymbol{v_i'} = \sum_{j=1}^{m} w_{i,j}\, \boldsymbol{T_j}\, \boldsymbol{v_i} = (\sum_{j=1}^{m} w_{i,j}\, \boldsymbol{T_j}\,)\boldsymbol{v_i}$$

Where $\boldsymbol{v_i'}$ represents the transformed vertex, $\boldsymbol{v_i}$, after applying the bone-induced transformations of the underlying skeleton using the LBS method and m is the total number of joints of the skeleton. Based on this formula, $\boldsymbol{v_i'}$ is obtained as the weighted linear combination (blend) of bone transformation matrices $\boldsymbol{T_j}$.

Since the group of 3D rigid rotations (SO (3)) is not a linear space but a curved manifold, a linear combination of rigid rotations is no longer a valid rigid rotation. Consequently, LBS performs well when the transformations to be blended are similar, but it fails when there are large relative rotations, such as in joints with a wide range of motion, like shoulders, wrists, or elbows. When LBS attempts to interpolate these rotations, unnatural compression or twisting of the mesh may occur, resulting in the well-known 'candy-wrapper artefact'. In addition to this problem, LBS has only a limited number of parameters, which makes it incapable of accurately representing complex deformations [38].

Advanced methods like Dual Quaternion Blending Skinning (DQBS) address these issues by using non-linear transformations to preserve rigidity. This approach is based on the use of dual quaternions, a mathematical structure for representing rigid 3D transformations, combining rotations and translations in a compact and efficient form. A dual quaternion can be written as the sum of two ordinary quaternions:

$$\hat{\boldsymbol{q}} = \boldsymbol{q_0} + \varepsilon \boldsymbol{q_\varepsilon}$$

Where $\boldsymbol{q_0}$ is the non-dual part, $\boldsymbol{q_\varepsilon}$ the dual part and $\boldsymbol{\varepsilon}$ is a dual unit satisfying $\varepsilon^2 = 0$.

Alternatively, a dual quaternion can be expressed as:

$$\hat{\boldsymbol{q}} = \hat{w} + i\hat{x} + j\hat{y} + k\hat{z}$$

Where $\hat{w}$ is the scalar part (dual number), $(\hat{x}, \hat{y}, \hat{z})$ is the vector part (dual vector) and i, j, k are the usual quaternion units.

A unit dual quaternion is a dual quaternion satisfying the unitary norm condition and can represent both rotations and translations in the 3D space. Specifically, when the dual part is null ($\boldsymbol{q_\varepsilon = 0}$), the unit dual quaternion represents a pure 3D rotation. For translations, a unit dual quaternion $\hat{\boldsymbol{t}}$, defined as $\hat{\boldsymbol{t}} = 1 + \frac{\varepsilon}{2}(t_0 i + t_1 j + t_2 k)$ corresponds to a translation by vector (t_0, t_1, t_2). A rigid transformation is a composition of rotation and translation. This composition corresponds to the multiplication of unit dual quaternions, which always produces a unit dual quaternion [39].

Based on this, the first step in switching from the LBS to the DQBS method is to convert the previously described transformation matrices, $\boldsymbol{T_1, \ldots, T_m}$, into unit dual quaternions. The next task is to perform dual quaternion linear blending, i.e. to compute a blended unit dual quaternion $\boldsymbol{q}$ with respect to the given weights. This can be done by taking their linear combination followed by a normalization (to obtain a unit dual quaternion). The resulting unit dual quaternion can be converted back into a matrix, which is then used to transform a vertex from its rest-pose $\boldsymbol{v_i}$ to its deformed one $\boldsymbol{v_i'}$. Through this process, the DQBS method always returns a rigid transformation, successfully eliminating the "candy-wrapper" artefacts present in the LBS method. However, a side effect of dual quaternion blending is the "bulging artifact", in which the vertices near a joint move outward during a bend, causing an unnatural bulge. This happens because spherical blending constrains the vertices to move on spheres with a fixed radius, ignoring the natural compression of the volume [38].

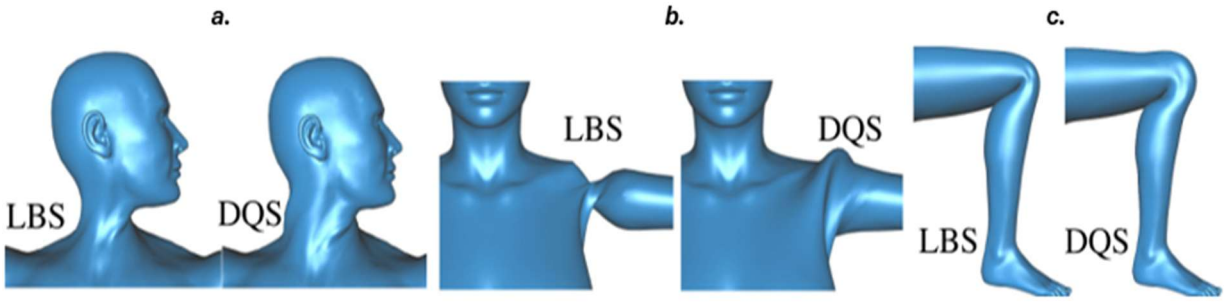

***Figure 10**: Illustration of deformation artifacts in Linear Blend Skinning (LBS) and Dual Quaternion Skinning (DQBS). LBS suffers from the "candy-wrapper" artifact when twisting, causing unnatural distortions around the rotation axis (a; b). This issue is eliminated by nonlinear blending techniques such as DQBS. However, DQS introduces an unnatural joint-bulging artifact while bending (c), absent in LBS. This comparison highlights that LBS does not produce bulging while bending and DQS does not suffer from the "candy-wrapper" artefact while twisting [40].*

These skinning approaches, known as geometric methods, have inspired an evolution towards more sophisticated models known as example-based or data-driven methods that permit to handle more complex skinning effects. These methods take as input a series of sculpted example poses and interpolate them to obtain the desired deformation. Among these example-based

methods, one of the first to emerge was pose space deformation (PSD) [41]. This technique introduces deformations of the mesh, in the sense of vertex displacements, with respect to a basic shape to solve typical artefacts of basic blend skinning. Specifically, PSD defines corrective shapes for specific key poses, calculates the distance between the current pose and the nearest key poses, uses a function, (e.g. Radial Basis function) to weight the example poses non-linearly based on distance and finally combines the corrective shapes using these weights to produce the right deformation for the current pose. An improvement of PSD was introduced with Weighted Pose Space Deformation (WPSD) [42], which largely reduces the number of required example poses to generate natural deformations. Unlike traditional PSD, in which the deformation is globally influenced by the example poses, WPSD assigns each vertex a specific weight that modulates the influence of the joints according to their local relevance, improving the quality of deformation. As an alternative to using sculpted example poses and manually adjusting corrections, several example-based approaches employed scanned or photographed data. Allen et al in [43] used recorded 3D scans to learn key poses deformations and, similarly to the PSD approach, interpolated them to generate the desired deformation. However, requiring a very large training dataset, this method fails to generalize well to new poses. To overcome this limitation, Kry et al. [44] used Principal Component Analysis (PCA) to extract the most significant variations in the positional deformations associated with each joint, reducing the dimensionality of the problem and improving the ability to generalize. The approaches introduced in [43] and [44] proposed learning poseable single-shape models, however, to realistically represent the variability of human morphology, poseable models capable of covering the entire space of variation in body shape were then developed. One of the most promising methods in this context is SCAPE, which represents body shape and pose-dependent shape in terms of triangle deformations rather than vertex displacements. Using training scans with different body shapes and poses, SCAPE builds a statistical model of shape variation, exploiting triangle-level deformations to integrate multiple transformations, including body shape variation, rigid part rotation and pose-dependent deformation.

Once training is complete, the SCAPE model can capture a wide range of poses and body shapes by relying on a triangulated mesh template $\boldsymbol{T}$, that is pre-segmented into rigid parts connected in a kinematic tree structure. Within this framework, all triangles belonging to a single part share the same rotation.

In SCAPE, each triangle $\boldsymbol{T_f}$ is individually deformed according to:

- A rigid rotation based on the pose, $\boldsymbol{R_f(\theta)}$.
- A deformation representing the specific shape of a person's body, $\boldsymbol{D_f}$.
- A deformation that accounts for pose-dependent shape changes like muscle bulging and skin wrinkling and which corrects for deviations between the rigidly posed model and the true shape, $\boldsymbol{Q_f(\theta)}$.

Let triangle $\boldsymbol{T_f}$ contain the points $\boldsymbol{x_{f,1};\ x_{f,2};\ x_{f,3}}$. The deformations are applied in terms of the triangle's local coordinate system, obtained by translating point $\boldsymbol{x_{f,1}}$ to the global origin. Thus, the deformations will be applied to the triangle edges $\boldsymbol{v_{f,j} = x_{f,j} - x_{f,1}}$; $j = 2,3$ [45].

Consequently, each triangle is deformed according to this formula:

$$\boldsymbol{v'_{f,j} = R_f(\theta) D_f Q_f(\theta) v_{f,j}} \text{ ; } j = 2,3$$

After deformation, these triangles are assembled, or 'stitched', to define the vertices of a watertight mesh, $\boldsymbol{M(\theta, D, Q)}$. Since each rigid part is independently rotated, the final stitched body surface can collapse, crease or fold near joints. To address this issue, the BlendSCAPE model was introduced, where the rotation of each triangle is expressed as a linear blend, $\boldsymbol{B(\theta)} = \sum_i w_i \boldsymbol{R_i}$, combining the rotations $\boldsymbol{R_i}$ of the $i-th$ part of the kinematic tree. Clearly, $\boldsymbol{B(\theta)}$, alone cannot fully capture realistic body pose deformations, but it reduces the workload of $\boldsymbol{Q(\theta)}$, which handles residual inaccuracies from simplified part interactions and soft tissue details.

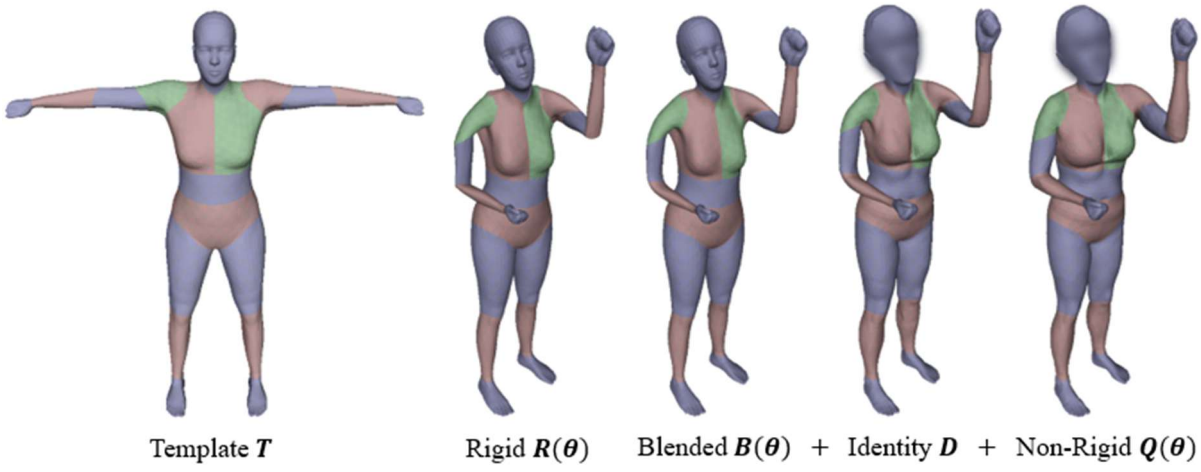

***Figure 11**: Illustration of template deformation using BlendSCAPE. Unlike SCAPE, BlendSCAPE performs a linear blend of rigid rotations ($\boldsymbol{B(\theta)}$) to achieve smoother transitions between body segments. The final deformed model is obtained from the contributions of $\boldsymbol{B(\theta)}$ for joint motion, $\boldsymbol{D}$ for identity shape and $\boldsymbol{Q(\theta)}$ for non-rigid deformations, providing a more realistic representation of the human body in motion [46].*

Although the BlendSCAPE model can capture a wide range of poses and shapes and model non-rigid deformations due to pose, it has some limitations related to its approach based on local triangle deformations. Training models by treating triangle deformations independently reduces computational cost, but at the same time compromises the ability to preserve long-range relationships between distant points in the mesh. Moreover, the representation of body shape and pose-dependent shape in terms of triangular deformations makes these models not compatible with existing animation software, which typically relies on skeletal rigging and skinning techniques. These limitations have led to the development of models such as Skinned Multi-Person Linear model (SMPL), which overcome these constraints by adopting a vertex-based approach, providing a more accurate representation of global anatomical relationships and shape-pose interactions and better adaptability to animation and deep learning applications [47].

Chapter 3

3. Materials and methods

3.1. Participants

Fourteen patients enrolled in the Swedish CPUP program, consisting of 6 females and 8 males were involved in this study. Eight participants were affected by clubfoot: five had bilateral clubfoot, while the remaining three had unilateral right clubfoot. Similarly, five participants had Cerebral Palsy (CP), with four presenting bilateral CP and one having unilateral right CP. The last patient had right tibial torsion.

3.2. Experimental setup

The experimental setup involved the use of An Azure Kinect camera (RGB images: 720 × 1280 pixels at 30 fps, Depth images: 640 × 576 pixels at 30 fps) from Microsoft was used, positioned 2.5 meters lateral to the center of the walkway and 5 meters from the background. The image coordinate system (I) of the video camera was aligned to the sagittal plane identified by the direction of progression and the vertical direction. The RGB Exposure of the camera was set manually to avoid motion blur due to Auto-exposure. In addition to the camera, to ensure clarity and better lighting conditions, two LED lamps were employed. Participants were asked to wear minimal clothing and colored ankle socks, specifically a red sock for the right foot and a blue sock for the left foot.

Four static acquisitions with the subject standing still were performed: one frontal, two sagittal (right and left) and one posterior. Participants were then instructed to walk on a 5-metre-long green carpet at a comfortable, self-selected speed. Six gait trials were acquired for each subject (3 left and 3 right gait cycles). Where necessary, assistance from an expert operator or parent was provided to participants to enable them to maintain an upright position during the static acquisition phase and to ensure stability during walking. The validation against the gold standard was conducted using a stereo-photogrammetric system composed by 12 cameras (Oqus 400 Qualisys medical AB, Gothenburg, Sweden, fs = 100 Hz). Thirty-eight retro-reflective markers according to the modified Helen-Heyes model [48] were used and attached on the skin of the subject. To calculate lower-limb joint angles, Visual 3D software (C Motion Inc., USA) was used. It is important to highlight that acquisitions were not performed

synchronously by the MB and ML systems to avoid IR interference, as the wavelength of the Azure Kinect IR sensor is the same as the Qualisys system (850 nm). Consequently, the same trial was repeated twice to be acquired separately with the MB and ML systems under the hypothesis of repeatability of the gesture.

3.3. Multi-segmental model definition

Skinned Multi-Person Linear model (SMPL) by the Max Planck Institute (2014) is a skinned vertex-based model that realistically represents a wide variety of body shapes and poses by using 3D scans of different subjects in several poses. It uses a data-driven statistical approach to capture and replicate the natural variations in human body shapes and poses with high accuracy. Like BlendSCAPE, the SMPL model decomposes body shape into identity-dependent and non-rigid, pose-dependent shapes; unlike BlendSCAPE, it uses a vertex-based skinning approach with corrective blend shapes to improve deformation accuracy. Each blend shape is represented as a vector of concatenated vertex offsets. These vectors are added to a mean shape to achieve specific body shapes and poses. The mesh presents an unchanged topology for men and women and a spatially varying resolution. It is realized as a clean quadrilateral structure, segmented into distinct parts, including initial blend weights and a skeletal rig. For simplicity of notation, meshes and shapes are indicated as vectors of vertices using bold upper-case letters (e.g. $\boldsymbol{X}$), while bold lower-case letters (such as $\boldsymbol{x_i}$) are used to represent individual vertices specifically [47].

3.3.1. Pose blend shapes

The initial mean template shape consists of a vector of N = 6890 concatenated vertices in the zero pose $\boldsymbol{\theta^*}$**.** The body pose is defined by a standard skeletal rig composed of 23 joints and is therefore represented by a pose vector $\boldsymbol{\theta}$ of 72 scalar parameters of which 69 (23*3) denotes the axis-angle representation of the relative rotation of part k with respect to its parent in the kinematic tree around the three axes and the remaining 3 parameters are the root orientation.

The various poses are generated from the pose vector $\boldsymbol{\theta}$, as defined above, by applying the Pose blend shape function B_p, that takes as input a vector of pose parameters, $\boldsymbol{\theta}$, and accounts for the effects of pose dependent deformations. It describes the vertex deviations from the rest template, $\boldsymbol{T}$:

$$B_p(\boldsymbol{\theta}; \boldsymbol{P}) = \sum_{\boldsymbol{n=1}}^{\boldsymbol{9K}} (\boldsymbol{R_n}(\boldsymbol{\theta}) - \boldsymbol{R_n}(\boldsymbol{\theta^*}))\mathbf{P_n}$$

Where $\boldsymbol{R_n(\theta)}$ is a vector of 207 elements (23x9) obtained by applying the function $\boldsymbol{R}$ which takes a pose vector $\boldsymbol{\theta}$ and returns a vector of concatenated part relative rotation matrices, $\boldsymbol{\theta^*}$ is the resting pose, $\boldsymbol{K}$ is the number of joints of the SMPL model and $\boldsymbol{P}$ is a matrix of all 207 pose blend shapes, i.e. containing vectors of vertex displacements $\mathbf{P_n}$ which are applied to the vertices following the rotation of a joint [47].

3.3.2. Shape blend shapes

Similarly, to represent body shapes of different people a linear function Bs is used. It takes as input a vector of shape parameters, $\boldsymbol{\beta}$, and outputs a blend shape sculpting the subject identity.

$$B_S(\boldsymbol{\beta};\boldsymbol{S}) = \sum_{\boldsymbol{n=1}}^{|\boldsymbol{\beta}|} \beta_{\boldsymbol{n}} \boldsymbol{S_n}$$

Where $\boldsymbol{\beta}$ is a shape vector which could contain from 10 to 300 values and it could be interpreted as an amount of expansion or shrink of a human subject, $|\ \boldsymbol{\beta}\ |$ is the number of linear shape coefficients and $\boldsymbol{S_n}$ represent orthonormal principal components of shape displacements such as in which direction the displacements should be applied due to changes of the shape factor.

Different body shapes are associated with different positions of the joint centers. The 3D coordinates associated with them are then obtained as a function of the body shape $\boldsymbol{\beta}$. Starting from the SMPL model, the locations of the joint centers J are estimated as a weighted sum of the surrounding vertices obtained by applying a joint regression matrix, $\boldsymbol{J}$. The latter was provided by the Max Plank Institute (learned from the training set) and defines a scattered set of vertex weights for each joint [47].

$$\mathrm{J}(\boldsymbol{\beta};\boldsymbol{J};\boldsymbol{T};\boldsymbol{S}) = \boldsymbol{J}(\boldsymbol{T} + B_S(\boldsymbol{\beta};\boldsymbol{S}))$$

***Figure 12**: Estimation of joint centers positions. The left knee joint is computed as the weighted average of red vertices [47].*

3.3.3. Application of transformations

In the animation process, transformations like translation, rotation, and scaling are applied to the skeleton segments of the SMPL model. Rather than assigning these transformations directly to individual segments, they are blended using weights. As a result, each vertex on the mesh is affected by several transformations simultaneously, with the influence of each transformation determined by its respective blend weight.

To change the shape of the SMPL model the corrective blend shapes of the functions illustrated above B_p and B_s are added in the resting pose as a first step:

$$\boldsymbol{T_P}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \boldsymbol{T} + B_s(\boldsymbol{\beta}) + B_p(\boldsymbol{\theta})$$

Consequently, the individual $i - th$ vertex to which the blend shapes have been applied is determined as follows:

$$\boldsymbol{t_{P,i}}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P}) = \boldsymbol{t_i} + \sum_{\boldsymbol{m=1}}^{|\boldsymbol{\beta}|} \beta_m s_{m,i} + \sum_{\boldsymbol{n=1}}^{\boldsymbol{9K}} (\boldsymbol{R_n}(\boldsymbol{\theta}) - \boldsymbol{R_n}(\boldsymbol{\theta}^*)) p_{n,i}$$

Where $s_{m,i}$, $p_{n,i}$ are the elements of the shape and pose blend shapes corresponding to template vertex $\boldsymbol{t_i}$.

Based on this definition; to generate different poses, a vertex $\boldsymbol{t_i}$ is transformed during the dynamic processing as follows:

$$t_i' = \sum_{k=1}^{K} w_{k,i}\, G_k'(\boldsymbol{\theta}, J(\boldsymbol{\beta}, \boldsymbol{J}, \boldsymbol{T}, \boldsymbol{S}))\; t_{P,i}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{T}, \boldsymbol{S}, \boldsymbol{P})$$

Therefore, to compact the notation, a generic SMPL model can be described as follows:

$$\boldsymbol{SMPL_{model}} = \boldsymbol{W} * \boldsymbol{G}(\boldsymbol{\theta}) * (\boldsymbol{T} + \boldsymbol{S} * \boldsymbol{\beta} + \boldsymbol{R}(\boldsymbol{\theta}) * \boldsymbol{P})$$

Where $\boldsymbol{W}$ is the matrix of the weight, $\boldsymbol{G}(\boldsymbol{\theta})$ is the transformation matrix of all SMPL joints, $\boldsymbol{T}$ is the mean shape, $\boldsymbol{S} * \boldsymbol{\beta}$ represents the displacement of each vertex of the mean shape due to the change of the shape factor, $\boldsymbol{R}(\boldsymbol{\theta}) * \boldsymbol{P}$ represents the displacement of each vertex of the mean shape due to the change of the pose factor.

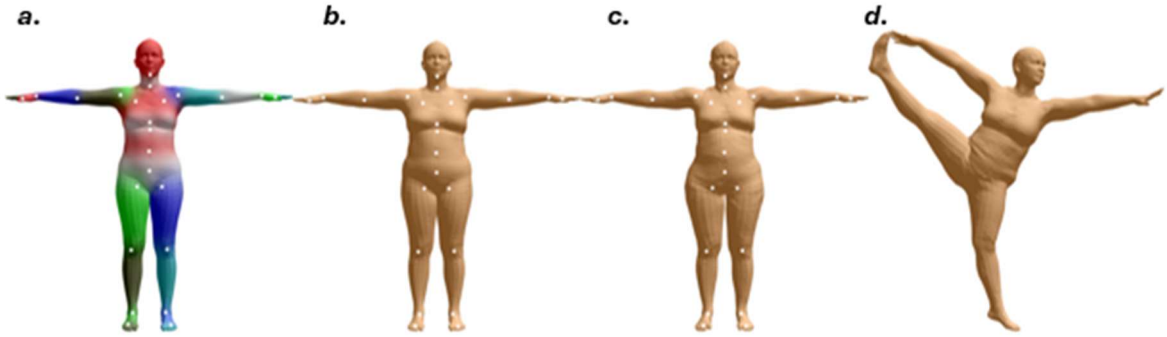

Figure 13: *a) Template mesh with blend weights indicated by color and joints shown in white ($\boldsymbol{T}, \boldsymbol{W}$). b) Template mesh with identity-driven blendshape contribution only ($\boldsymbol{T} + B_s(\boldsymbol{\beta}), J(\boldsymbol{\beta})$) ; both vertex and joint locations vary linearly with the shape vector. c) Template mesh with both shape blend shapes and pose blend shapes contributions ($\boldsymbol{T_P}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \boldsymbol{T} + B_s(\boldsymbol{\beta}) + B_p(\boldsymbol{\theta})$). d) Deformed vertices reposed by dual quaternion skinning for the split pose ($\boldsymbol{SMPL_{model}} = \boldsymbol{W} * \boldsymbol{G}(\boldsymbol{\theta}) * (\boldsymbol{T} + \boldsymbol{S} * \boldsymbol{\beta} + \boldsymbol{R}(\boldsymbol{\theta}) * \boldsymbol{P})$) [47].*

As far as the training of the SMPL model is concerned, as the model decomposes shape and pose these were trained separately, simplifying the optimization. For the pose parameters, a multi-pose dataset of 1786 recordings from 40 individuals was utilized. Gradient-based refinement was applied to optimize the parameters W and P, minimizing the difference between the 3D model and the edges of the captured 3D scans. The multishape dataset, on the other hand, consists of records from the CAESAR dataset, totalling 1700 records for males and 2100 for females, from which the mean template shape ($\boldsymbol{T}$), regressor matrix ($\boldsymbol{J}$) and blend weights for the shape parameters were automatically derived. Principal component analysis (PCA) was employed to capture the most significant variance in vertex offsets in the rest pose, constrained by a limited number of shape directions (**Figure 14**) [47].

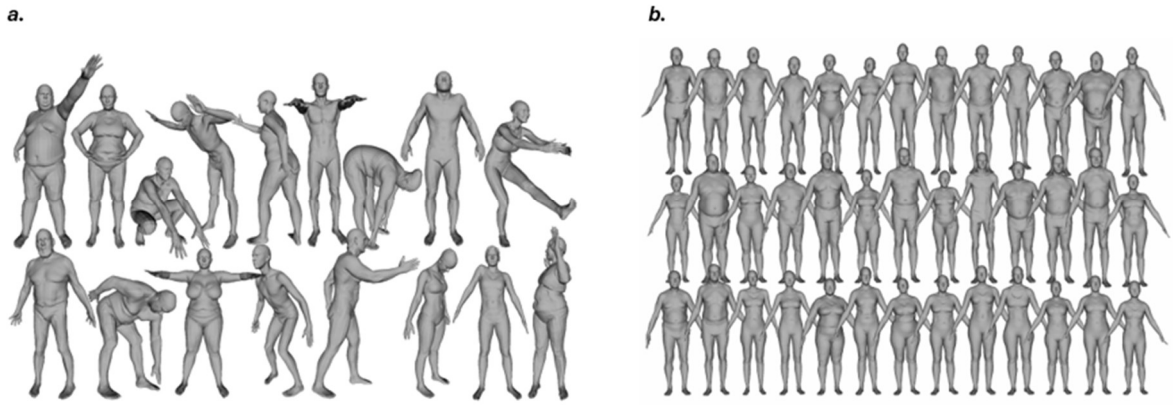

Figure 14: *a) Multi-pose dataset. b) Multi-shape dataset [47].*

3.4. Subject-specific model calibration

A generic SMPL model of the lower limb, consisting of the thigh, shank, and foot, was employed for gait analysis. The anatomical segments in the model are connected by the joint centers at the hip, knee, and ankle. The objective is to generate a subject-specific 3D static model using markerless technology by adapting the mean template of the SMPL lower-limb model to the experimentally derived subject's 3D posture, using the following formula:

$$\boldsymbol{SMPL_{subject-specific}} = \boldsymbol{W} * \boldsymbol{G}(\boldsymbol{\theta_{ss}}) * (c_{ss} * \boldsymbol{T} + \boldsymbol{S} * \boldsymbol{\beta_{ss}} + \boldsymbol{R}(\boldsymbol{\theta_{ss}}) * \boldsymbol{P})$$

Where $\boldsymbol{c_{ss}}$ is the scaling factor to adapt the size of the mean shape of the lower limb model, $\boldsymbol{T}$, to the subject's one. The pose parameters, $\boldsymbol{\theta_{ss}}$, were changed to fit the pose of $\boldsymbol{T}$ to that of the subject. Moreover, a displacement can be applied to each vertex of the model to further enhance its shape and improve alignment with the subject's observed data.

First, the subject was preliminarily identified using the depth image by applying two thresholds based on the subject's known distance from the camera. Then, the lower limb of interest was identified and isolated in the RGB image by manually identifying two points on the image: one to differentiate the right leg from the left, as they were processed separately, and another to select the region extending from the pelvis downward. The green carpet on which the subject was placed during the acquisitions was removed by applying a color filter on the RGB image. Binary masks of left and right foot were then obtained in the same way by exploiting the red and blue color of the ankle socks supplied to the subject, enabling more accurate foot segmentation.

Point clouds from the four different views of the subject were also extracted from the RGB-D camera and were processed to exclude the 3D points corresponding to pixels previously

removed in the depth images. For each side (left and right), the 3D static posture of the subject was reconstructed by merging the frontal, sagittal, and posterior point clouds by identifying four points of interest (i.e. $TL, H, M1, M2$) which were then aligned. $M1$ (in the frontal and sagittal views) and $M2$ (in the posterior and sagittal views) were identified as intersection points between the segment corresponding to the upper pelvis and the segment corresponding to the upper thigh. The heel point H on the posterior depth image was identified as the point with the minimum depth value within the foot. On the sagittal depth image, H was determined as the most lateral point within the foot region. Instead, the point TL in the frontal depth image was identified as the most lateral point in the foot region, in the sagittal plane, TL was identified as the point with the minimum **x** coordinate for the right foot and the maximum **x** coordinate for the left foot, along with the maximum **y** coordinate in the anterior region of the foot. The latter was obtained by selecting points within a **z** value range with a width equal to the difference between the **z** coordinates of the tip of the foot and the most lateral point of the foot, obtained previously on the frontal plane.

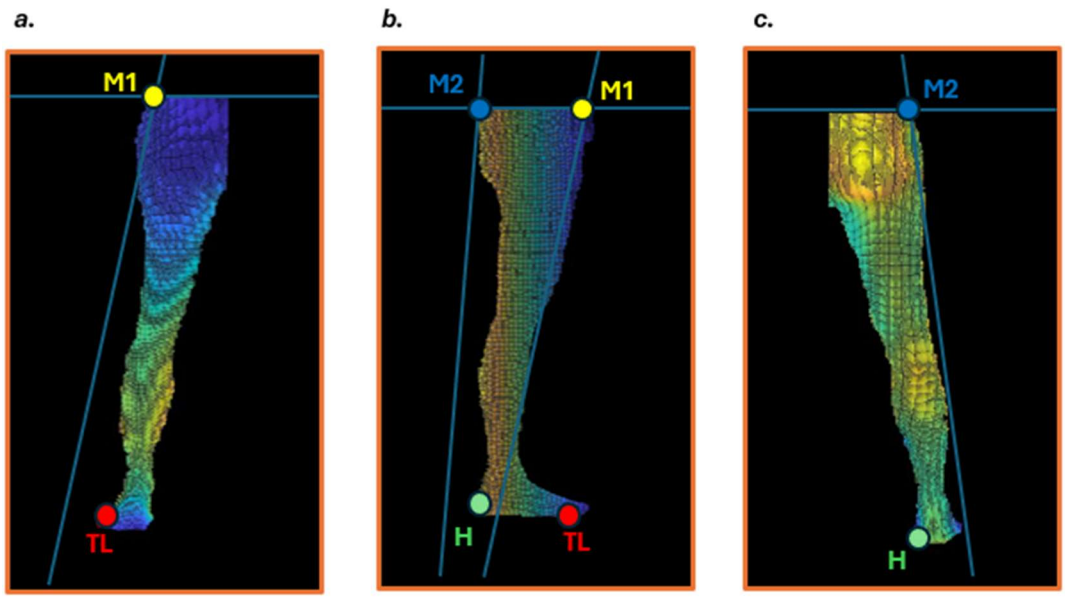

***Figure 15**: a) Representation of the frontal point-cloud and its points of interest. b) Representation of the right sagittal point-cloud and its points of interest. c) Representation of the posterior point-cloud and its points of interest*

Thanks to the points obtained, the alignment of the sagittal and posterior point clouds with respect to the frontal one was successfully performed, with the latter remaining unchanged. The steps taken were:

1. Alignment of the sagittal point cloud to the frontal: using M1 points (corresponding between sagittal and frontal), a translation vector was calculated to move the sagittal point cloud to the correct position relative to the frontal. Subsequently, the TL points

were used to determine the necessary rotation angles in the frontal and sagittal planes, resulting in a precise alignment of the sagittal point cloud in 3D space with the frontal one. The sagittal point cloud was then translated again to ensure appropriate positioning with the frontal point cloud after the applied rotations.

2. Alignment of the posterior point cloud to the sagittal: M2 points (corresponding between posterior and sagittal) were used to calculate the translation vector to correctly position the posterior point cloud with respect to the previously aligned sagittal one. Subsequently, the H points (heel points) were used to calculate the rotation angles, in the frontal and sagittal plane, required to align the posterior point cloud with the sagittal one. The posterior point cloud was then translated again to ensure appropriate positioning with the sagittal point cloud after the applied rotations.

Once aligned, the three point clouds (frontal, sagittal and posterior) were merged to obtain a complete 3D static posture of the subject, as shown in **Figure 16**.

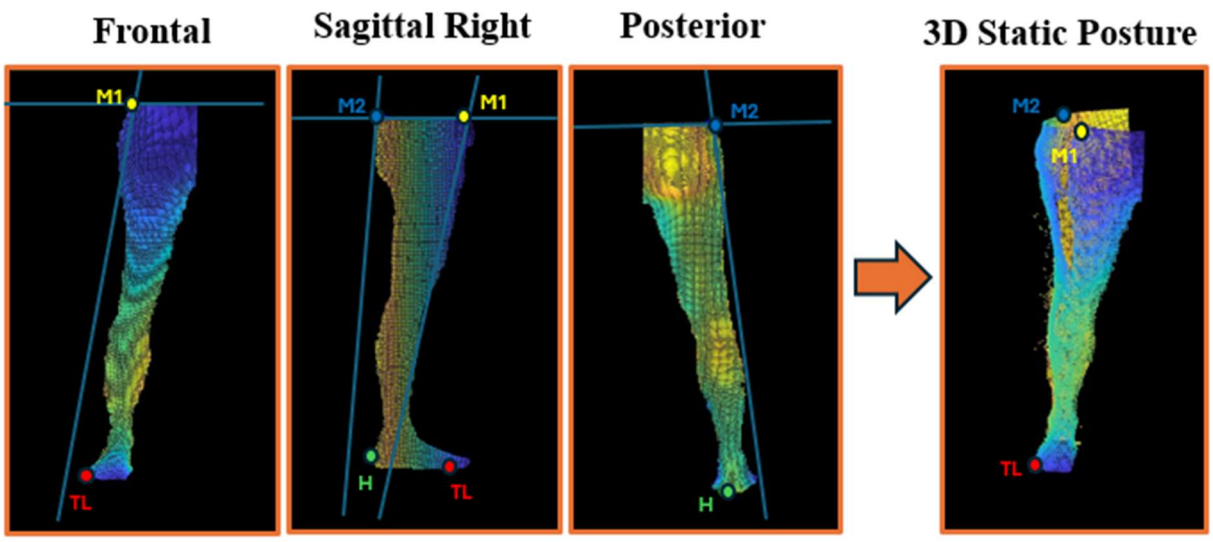

***Figure 16**: Creation of a 3D static posture by aligning four common points among different views.*

At this point, the scaling factor c_{ss} calculated as the ratio between the length of $\boldsymbol{T}$ and the length of the 3D static posture was applied to the template $\boldsymbol{T}$.

After that $\boldsymbol{T}$ was translated so that it was aligned with the subject's 3D static posture by matching the most anterior and upper point on the hip. To further refine the alignment, the template was rotated in the frontal and sagittal plane by a hip angle, calculated using trigonometric formulas based on the coordinates of the most anterior and upper point of the hip and the toe point.

3.4.1. Articulated Iterative Closest Point

To compute the pose of the subject's lower limb, represented by $\boldsymbol{\theta_{ss}}$, the Articulated Iterative Closest Point (AICP) was implemented.

The Iterative Closest Point algorithm (ICP) is a widely used method for registering 3D point sets. It aligns points from a source surface to a target surface by minimizing the distance between corresponding points. The AICP algorithm [29] is a generalization of the ICP algorithm specifically designed for articulated structures, maintaining the essential characteristics of the original method. The concept behind AICP is to subdivide the articulated body into distinct parts and perform a rigid and complete alignment of these in each iteration as in the original ICP but with the addition of constraints that ensure the structure remains intact.

An articulated body model M is composed of rigid parts $\{p_1; p_2; \dots\ p_{NP}\}$. Each part p_i has a joint j_i through which the part is connected to another part. We restrict our attention to open kinematic structures. We arbitrarily choose one of our rigid parts to be the root node p_r. By convention, the corresponding joint jr, which has no parent, is connected to the world.

In this context, we will consider the hip, knee, and ankle joints as being approximately equivalent to spherical joints. A spherical generic joint j_i has 3 d.o.f such that:

$$\boldsymbol{\theta_i} = \{\alpha, \beta, \gamma\}$$

Where α, β, γ are the rotation angles around the x, y, and z-axes, respectively

For each part p_i and joint j_i there is a rigid transformation $\boldsymbol{G_i(\theta_i)}$, which specifies the part's rotation w.r.t. its parent. The absolute pose of a part p_i in the world coordinate system $\boldsymbol{G_i^W(\theta_i)}$ can be obtained by concatenating the transformations along the kinematic chain from the root part to p_i as follows:

$$\boldsymbol{G_i^W(\theta_i) = G_r^W * \dots * G_i(\theta_i)}$$

Where $\boldsymbol{G_r^W}$ is a rigid transformation which specifies the root's translation and rotation w.r.t. the world reference system.

For each p_i, there is a set of N_i points $\{\boldsymbol{m_1^{pi}} \dots\ \boldsymbol{m_{Ni}^{pi}}\}$ such as the position of a point $\boldsymbol{m_j^{pi}}$ in world coordinate system is $\boldsymbol{G_i^W m_j^{pi}}$. By removing a single joint j_i , any open articulated structure M can be split into two branches, one of which contains the root p_r. We will call the branch

containing the root the base branch $\boldsymbol{M_b^i}$, and the other one the outer branch $\boldsymbol{M_o^i}$, containing the end effector. In the special case of "splitting" at the root p_r, we define $\boldsymbol{M_b^r} = \boldsymbol{M_b^r}$, i.e., both branches correspond to the entire model.

Pellegrini and colleagues proposed that, rather than estimating all pose parameters using an iterative local optimizer, attention should be focused on a small subset of parameters that can be solved in closed form. The process involves iterating over different subsets to achieve optimal alignment: the articulated body was divided into the base branch $\boldsymbol{M_b^i}$ and the outer branch$\boldsymbol{M_o^i}$, as defined earlier, and only the outer branch was aligned through a rigid transformation that respects the joint constraints at j_i.

Given a dataset D of world points $\{\boldsymbol{d_1}; \boldsymbol{d_2}; \ldots \boldsymbol{d_{ND}}\}$, the alignment between the model and the dataset was achieved by minimizing the following function representing the norm of the difference, $Eo(\boldsymbol{\theta i})$, between $\boldsymbol{M_o^i}$ and its closest points of the dataset, $\boldsymbol{d_s}$:

$$E_o(\boldsymbol{\theta_i}) = \sum\nolimits_{\boldsymbol{p_k} \in \boldsymbol{M_o^i}} \sum_{j=1}^{N_{p_k}} \boldsymbol{min} \,||\, \boldsymbol{G_k^W}(\boldsymbol{\theta_i})\, \boldsymbol{m_j^{p_k}} - \boldsymbol{d_s}||^2$$

Where p_k represents the $k-th$ part belonging to the other branch $\boldsymbol{M_o^i}$. The resulting $\boldsymbol{G_k^W}(\boldsymbol{\theta_i})$ represents the rigid transformation which allows the best alignment between $\boldsymbol{M_o^i}$ and $\boldsymbol{d_s}$.

By minimizing the error function, the algorithm calculates the optimal rigid transformation parameters (rotation and translation) that respect the joint constraints at j_i and align the model points to those in the dataset, updating them iteratively until the overall alignment error is minimized, and convergence is achieved.

In order to choose at which joint level to divide the structure, a selection policy must be chosen. In particular, possible options include random, distributed or cyclic subdivision of the structure.

In the proposed 3D ML protocol for gait analysis, the hip was designated as the root pr, and the world-reference system I was aligned with the hip. A cyclical joint selection was applied, with a top-down approach starting from the hip and progressing to the knee.

As mentioned above, all three hip, knee and ankle joints were considered to be spherical joints with 3 DoF.

In particular,

$$\boldsymbol{\theta_{hip}} = \{\alpha_{hip}, \beta_{hip}, \gamma_{hip}\}$$

Where $\alpha_{hip}, \beta_{hip}, \gamma_{hip}$ are flexion-extension, abduction-adduction, intra-extra rotation angles around the hip, respectively.

$$\boldsymbol{\theta_{knee}} = \{\alpha_{knee}, \beta_{knee}, \gamma_{knee}\}$$

Where $\alpha_{knee}, \beta_{knee}, \gamma_{knee}$ are flexion-extension, abduction-adduction, intra-extra rotation angles around the knee, respectively.

$$\boldsymbol{\theta_{ankle}} = \{\alpha_{ankle}, \beta_{ankle}, \gamma_{ankle}\}$$

Where $\alpha_{ankle}, \beta_{ankle}, \gamma_{ankle}$ are plantar dorsi-flexion, intra-extra rotation, prono-supination angles around the ankle, respectively.

Finally,

$$\boldsymbol{\theta_{ss}} = \{\boldsymbol{\theta}_{hip}, \boldsymbol{\theta}_{knee}, \boldsymbol{\theta}_{ankle}\}$$

Following the cyclic selection policy, $\boldsymbol{T}$ was split from the hip joint and then the closest points of the 3D static posture to $\boldsymbol{T}$ were identified in order to apply a rigid alignment between them around the selected joint. Since the hip is the root of the structure, the rigid transformation was applied to the entire model, $\boldsymbol{M_o^{hip}}$, comprising thigh, shank and foot. The rigid alignment was computed by minimizing the norm of the difference, $\boldsymbol{E_o(\theta_{hip})}$, between $\boldsymbol{M_o^{hip}}$ and its closest points of the 3D static posture, $\boldsymbol{d_s}$:

$$E_o(\boldsymbol{\theta_{hip}}) = \sum\nolimits_{\boldsymbol{p_k} \in \boldsymbol{M_o^{hip}}} \sum_{j=1}^{N_{p_k}} \boldsymbol{min} \,||\, \boldsymbol{G_k^W(\theta_{hip})\, m_j^{p_k} - d_s}||^2$$

Where p_k represents the $k-th$ part belonging to the other branch $\boldsymbol{M_o^{hip}}$

Subsequently, the knee joint was selected to split $\boldsymbol{T}$ and a rigid transformation was applied to the outer branch ($\boldsymbol{M_o^{knee}}$, containing shank and foot) around this joint. As before, the rigid alignment was computed by minimizing the norm of the difference, $\boldsymbol{E_o(\theta_{knee})}$, between $\boldsymbol{M_o^{knee}}$ and its closest points of the 3D static posture, $\boldsymbol{d_s}$:

$$E_o(\boldsymbol{\theta_{knee}}) = \sum_{\boldsymbol{p_k} \in \boldsymbol{M_o^{knee}}} \sum_{j=1}^{N_{p_k}} \boldsymbol{min} \,||\, \boldsymbol{G_k^W}(\boldsymbol{\theta_{knee}})\, \boldsymbol{m_j^{p_k}} - \boldsymbol{d_s}||^2$$

Where p_k represents the $k-th$ part belonging to the other branch $\boldsymbol{M_o^{knee}}$

By iterating this procedure until convergence is reached, the rigid transformation parameters for the hip and knee, $\boldsymbol{\theta_{hip}}$ and $\boldsymbol{\theta_{knee}}$, were obtained, allowing the optimal alignment between $\boldsymbol{T}$ and the 3D static posture. Conditions on the termination of iterations were set both in terms of a maximum number of iterations fixed at 4 and in terms of no significant improvement in hip and knee alignment between the $\boldsymbol{SMPL_{subject-specific}}$ model and the 3D static posture in consecutive iterations (discrepancy between errors < 0.8%).

In addition, an optimization constraint based on the position of the knee was introduced to interrupt the hip angle adjustment process. Specifically, 3D points included between $\boldsymbol{^IKJC}$ *and* ($\boldsymbol{^IKJC} \pm \varepsilon$), where $\varepsilon = 20$ pixels, were isolated in both the $\boldsymbol{SMPL_{subject-specific}}$ model and the 3D static posture. If the minimum z-coordinate of the $\boldsymbol{SMPL_{subject-specific}}$ model points fell below the minimum z-coordinate observed in statics, the optimization process for the hip was interrupted. This measure was taken to prevent the knee of the $\boldsymbol{SMPL_{subject-specific}}$ model from bending excessively beyond the limits established by the 3D static posture.

The search limits for the implementation of the AICP were set based on physiological limits to ensure a permissible configuration. Specifically, in both the sagittal and frontal planes, the lower boundary was set at -18° and the upper boundary at 18°. The range is quite limited, and this is because in the static context, the joints are expected to maintain stable positions so there are no great variations in joint angles.

In the implementation of the algorithm, the internal and external rotation angles of the hip and knee, γ_{hip} and γ_{knee} were not optimized as they proved to be negligible during static acquisitions.

As for the ankle pose, $\boldsymbol{\theta_{ankle}}$, it was not calculated using the AICP algorithm. In particular, plantar-dorsiflexion, α_{ankle}, and internal-external rotation of the foot, β_{ankle}, were obtained using the previously calculated coordinates of the heel and toe, applying simple trigonometric formulas.

At this stage, the AICP implementation provided an initial estimate of the subject's pose, $\boldsymbol{\theta_{ss}}$, which was refined through the calculation of the shape factor, $\boldsymbol{\beta_{ss}}$, by minimizing the residual discrepancy between the mean shape, $\boldsymbol{T}$, and the subject's 3D static posture, as follows:

$$\boldsymbol{\beta_{ss}} = inv(\boldsymbol{S}) * (\ \boldsymbol{3D\ static\ posture}\ -\ (\boldsymbol{W} * \boldsymbol{G}(\boldsymbol{\theta_{ss}}) * (c_{ss} * \boldsymbol{T} + \boldsymbol{R}(\boldsymbol{\theta_{ss}}) * \boldsymbol{P}))$$

3.5. Joint centers trajectories estimation

The joint centers trajectories were tracked by fitting the $\boldsymbol{SMPL_{subject-specific}}$ on each dynamic point cloud of the gait cycle using AICP algorithm.

The point cloud relating to the lower limbs was extracted from each dynamic frame in the same way as for static acquisitions. Then, Otsu's algorithm (Otsu, 1979) was applied to separate the foreground limb from the background limb and to eliminate the hand during arm oscillation, so that only the foreground limb, $\boldsymbol{Dyn}$, could be isolated.

The most critical aspect in this step is the presence of the hand during arm oscillation, which can partially overlap the thigh during the gait cycle. This overlap could introduce challenges in aligning the $\boldsymbol{SMPL_{subject-specific}}$ model with $\boldsymbol{Dyn}$.Specifically, the presence of the hand may cause issues in the search for the closest points, leading the AICP optimization process to potentially get stuck at a local minimum.

Three different approaches were implemented to address this problem in frames where the hand was not distinguishable as a separate pixel region from the lower limbs. These approaches relied on the relative position of the hand and thigh, as follows:

- When the hand was completely external to the thigh (modulus of the difference between the x-coordinate values of the hand and thigh centroids greater than 47 pixels) and the frame under consideration fell within the first ten of the gait cycle, it was directly removed.
- When the hand partially overlapped the thigh (modulus of the difference between the x-coordinate values of the hand and thigh centroids between 20 and 47 pixels), a segment was fitted to the contour of the thigh so as to remove only the portion of the hand that went beyond the thigh.

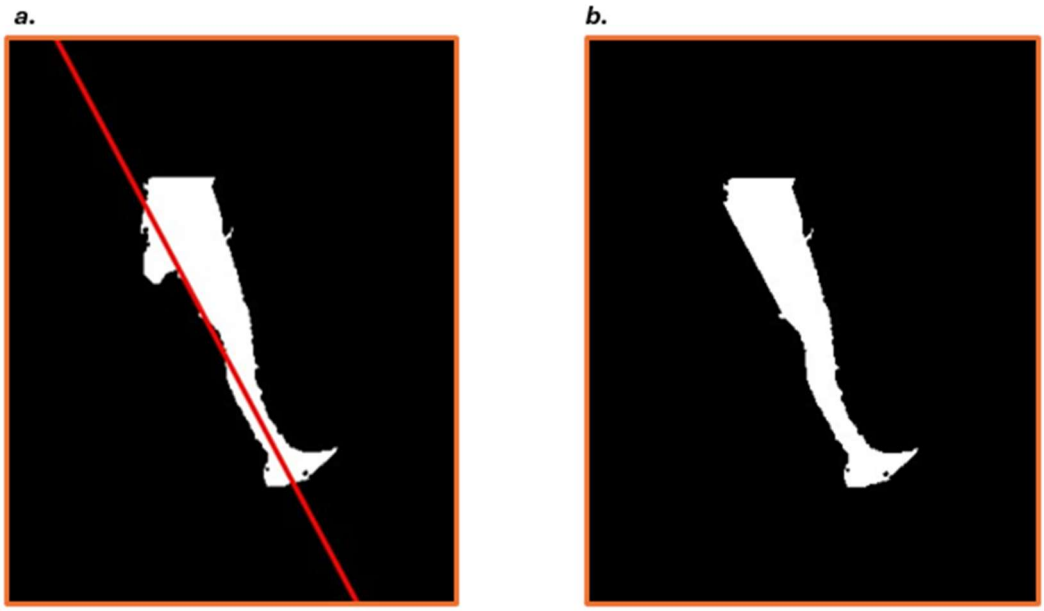

***Figure 17** : Hand oscillation during the gait cycle. a) Lower-limb depth image with the hand partially overlapped with the thigh. b) Lower-limb depth image after the hand removal.*

- When the hand was completely overlapping the thigh (modulus of the difference between the x-coordinate values of the hand and thigh centroids smaller than 20), it was left in its original position.

Once the foreground lower limb was isolated, the coordinates of the toe and heel points were identified for each dynamic frame using a specific algorithm for the gait cycle identification. The method proposed for identification of initial foot contacts was conceived to consider diverse types of foot contact with the ground which could occur in subjects with CP. According to [10], these patients may exhibit different types of gait patterns:

- **Equinus gait**: This is characterized by toe-walking and ankle plantarflexion, often caused by tightness in the calf muscles. Treatment options include botox injections to reduce spasticity, hamstring lengthening surgery, and/or the use of ankle-foot orthosis. In this gait pattern, both initial and final contacts are typically made with the toes.
- **Crouch gait**: Individuals with crouch gait show significant bending at the hips and knees and typically make contact with the ground using a flat foot.
- **Normal gait**: For patients with a low GMFCS level, initial contact is made with the heel, and final contact occurs with the toe

First, for each recorded frame, a binary segmentation mask ${}^{I}\boldsymbol{M}_{\boldsymbol{foot}}$, expressed in the image coordinate system I , was obtained for each foot using the color filter segmentation technique

[49]. Each mask was fitted within an ellipse, and a new reference system, f, integral with the foot was established with its axes aligned to the principal axes of the fitted ellipse and its origin positioned at the centroid. It subsequently became possible to convert the binary masks from the image coordinate system, I, to the foot reference system, f, by applying the transformation matrix $^{f}\boldsymbol{T}_{I}$ derived using simple trigonometric formulas. Obtained for each frame $^{f}\boldsymbol{M}_{foot}$, the position of the forefoot (FF) was derived in f, ($^{f}\boldsymbol{FF} = [FF_{xf}, FF_{yf}]$), as the point with the maximum x-coordinate (**Errore. L'origine riferimento non è stata trovata.**). As for the mid-rear foot (MRF) position in f, ($^{f}\boldsymbol{MRF} = [MRF_{xf}, MRF_{yf}]$), this was obtained instead from the intersection of two lines representing the foot sole and the foot posterior edge. Once the point with the maximum y-coordinate ($^{f}\boldsymbol{Q} = [Q_{xf}, Q_{yf}]$)was identified on the binary mask $^{f}\boldsymbol{M}_{foot}$, the foot sole was defined by isolating the points between Q_{yf} and ($Q_{yf} - \varepsilon$, where ε = 20 pixels) and fitting a straight line through this region. The line fitting the foot posterior edge was similarly obtained starting from the identification of the point with the smallest x-coordinate within the foot mask $^{f}\boldsymbol{M}_{foot}$ (**Errore. L'origine riferimento non è stata trovata.**). The $^{f}\boldsymbol{FF}$ and $^{f}\boldsymbol{MRF}$ were then referred to I by applying $^{I}\boldsymbol{T}_{f}$.

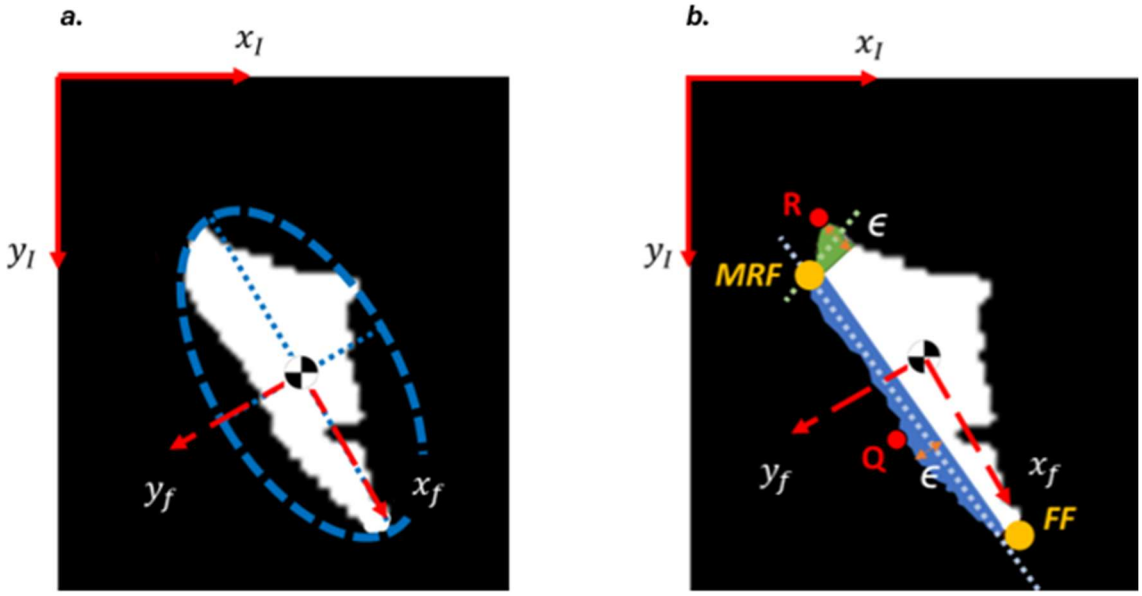

***Figure 18**: MRF and FF identification: a) An ellipse was drawn around each mask of the foot; the center and principal axes (xf, yf) were determined. b) The area where the foot sole (light blue) intersects with the posterior foot region (light green) was defined as the mid-rear foot (MRF), and the tip of the foot along the x-axis was defined as the forefoot (FF).*

After computing the toe and heel coordinates in $\boldsymbol{Dyn}$, they served as key reference points for subsequent phases. In particular, the next step involved estimating the lower limb length, by leveraging the toe coordinates to scale the $\boldsymbol{SMPL}_{subject-specific}$ model. This is particularly

important, as throughout the captured gait cycle, the dimensions and shape of the lower limb segments can vary due to factors such as soft tissue movement [50] or changes in the subject's distance from the camera. These fluctuations may lead to inaccuracies in the matching procedure, as the model size may not align with the actual size of the lower limb during dynamic movement.

Consequently , after placing the $\boldsymbol{SMPL_{subject-specific}}$ model in a neutral position (heel and toe aligned, hip rotated to align frontally with toe), the $\boldsymbol{SMPL_{subject-specific}}$ model was scaled for each dynamic frame, on the vertical axis using the ratio of the length of the lower limb in $\boldsymbol{Dyn}$ on the current frame to the length of the lower limb of the $\boldsymbol{SMPL_{subject-specific}}$ model fitted to the previous dynamic frame. These lengths were calculated as 3D Euclidean distance between the most anterior and upper point and the point representing the toe in both the $\boldsymbol{SMPL_{subject-specific}}$ model and $\boldsymbol{Dyn}$. Exceptions were applied in cases where the scaling factor of the current frame differs from the average of the scaling factors calculated on the previous frames by more than 4%. These included cases during the swing phase, where the leg and foot appear thinner in the depth image and the toe of the foot is estimated to be higher along the vertical axis (lower y-coordinate) as shown in **Figure 19**, as well as situations where the overlap of the arm on the thigh caused the most upper and anterior point to be estimated as being further forward along the anteroposterior axis (lower z-coordinate), as shown in **Figure 20**. In such instances, the scaling factor of the current frame was reset equal to that of the previous frame.

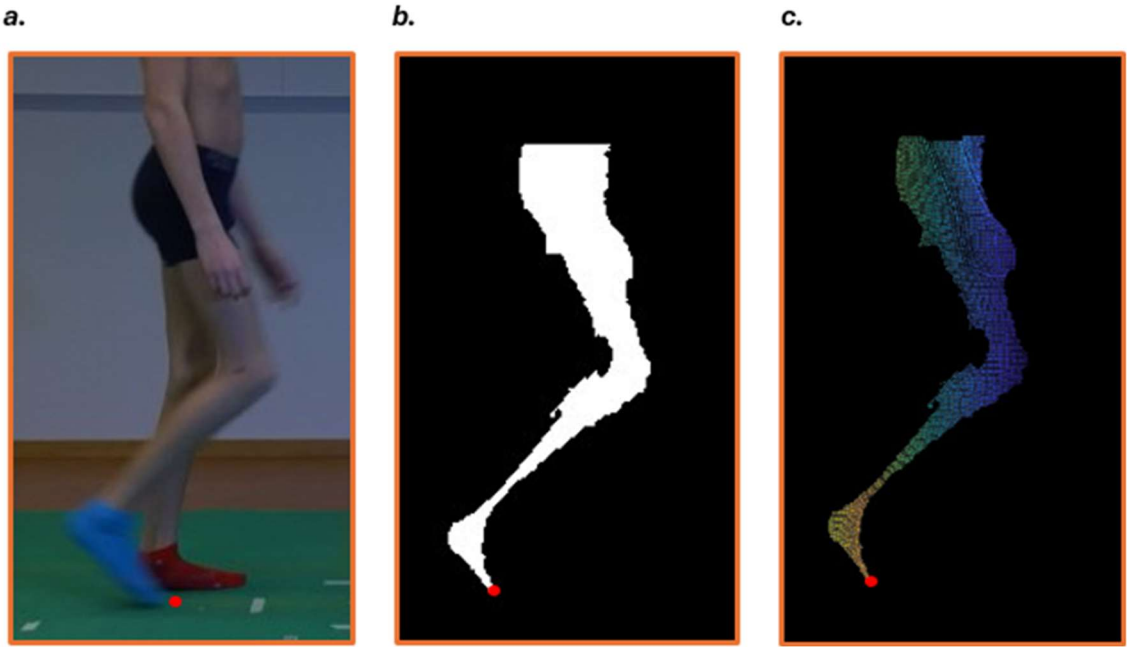

***Figure 19**: Example of a frame during the swing phase, illustrating the effects of depth image artifacts on the accuracy of toe position estimation. a) RGB image with the correct toe position highlighted in red. b) Depth image showing noticeable thinning of the leg and foot, resulting in an inaccurate estimation of the toe's vertical position and leading to an underestimation of the scaling factor. c) Point-cloud image (**$\boldsymbol{Dyn}$**) where the thinning effect and the consequent misestimation of the toe's position remain visible.*

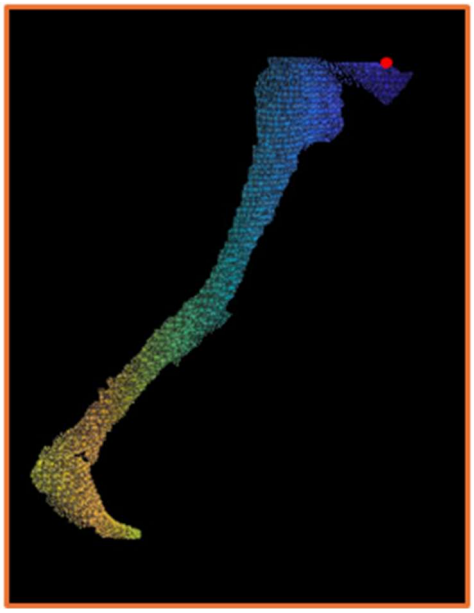

***Figure 20:** Point-cloud image (**$\boldsymbol{Dyn}$**) in which the overlapping of the arm on the thigh causes the most upper and anterior point (in red) to be estimated as further forward along the anteroposterior axis and leads to an overestimation of the scaling factor.*

The $\boldsymbol{SMPL_{subject-specific}}$ model was also scaled along the anteroposterior axis using a scaling factor calculated as the average of the ratios between dynamic and model data of pelvis widths, foot lengths, and thigh widths.

After the scaling process, the $\boldsymbol{SMPL_{subject-specific}}$ model was translated along the horizontal axis to align its external profile with the dynamics, $\boldsymbol{Dyn}$. Additionally, adjustments were made along the vertical and anteroposterior axes to ensure that the most anterior and upper point of the $\boldsymbol{SMPL_{subject-specific}}$ model matched that of the dynamics, $\boldsymbol{Dyn}$, in each frame.

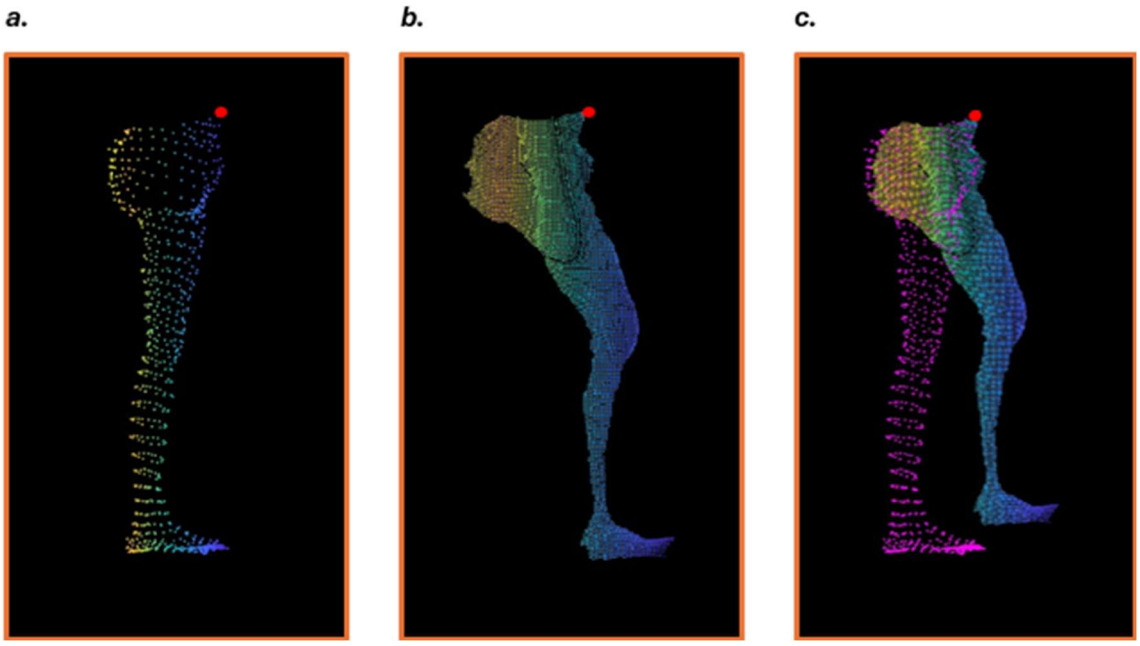

Figure 21: *a) Representation of* $\boldsymbol{Dyn}$ *and in red the most upper and anterior point. b) Representation of* $\boldsymbol{SMPL_{subject-specific}}$ *model and in red the most upper and anterior point. c)* $\boldsymbol{SMPL_{subject-specific}}$ *model (in magenta) translated so that its most upper and anterior point correspond to that of* $\boldsymbol{Dyn}$.

When the hand was positioned anteriorly with respect to the thigh, the $\boldsymbol{SMPL_{subject-specific}}$ model was translated to align the z-coordinate of its most upper and anterior point with the z-coordinate of a reference point identified in $\boldsymbol{Dyn}$ as the intersection of two lines, one fitting the profile of the thigh and the other fitting the profile of the pelvis, as shown in **Figure 22** and **Figure 23**.

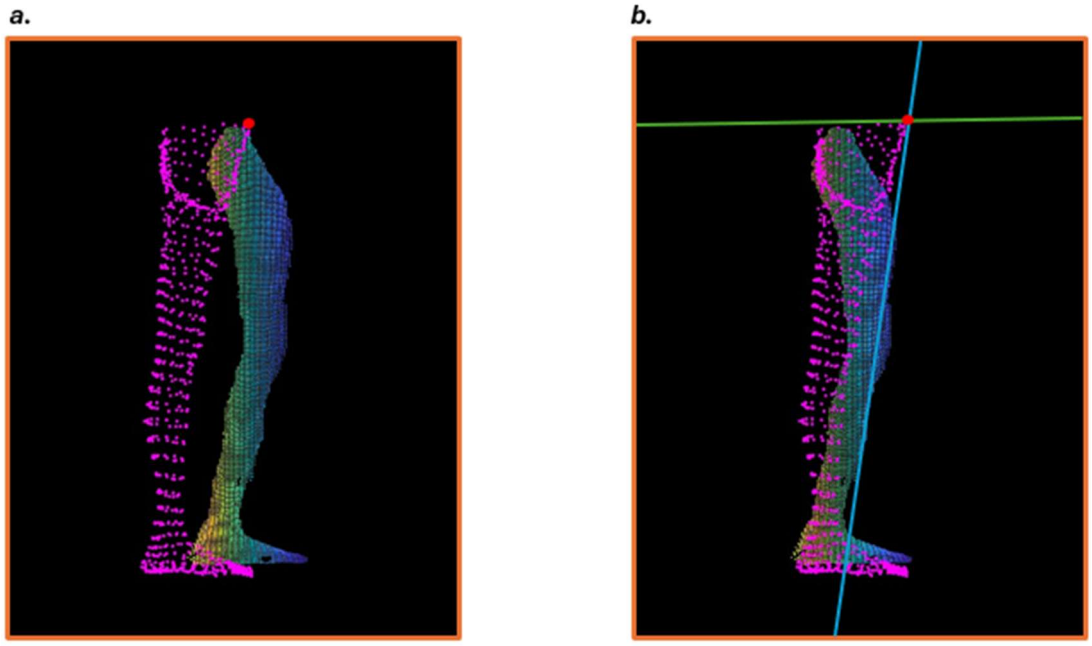

***Figure 22**: Correction of the $\boldsymbol{SMPL_{subject-specific}}$ model translation through the identification of a reference point, in a frame in which the hand, positioned anteriorly, is removed as it is recognised as a distinct region. a) Misidentification of the most upper and anterior point point in $\boldsymbol{Dyn}$, resulting in incorrect translation of the $\boldsymbol{SMPL_{subject-specific}}$ model (in magenta). b) Translation correction by identifying a point defined by the intersection of a line fitting the profile of the thigh (in blue) and one fitting the profile of the pelvis (in green) in $\boldsymbol{Dyn}$.*

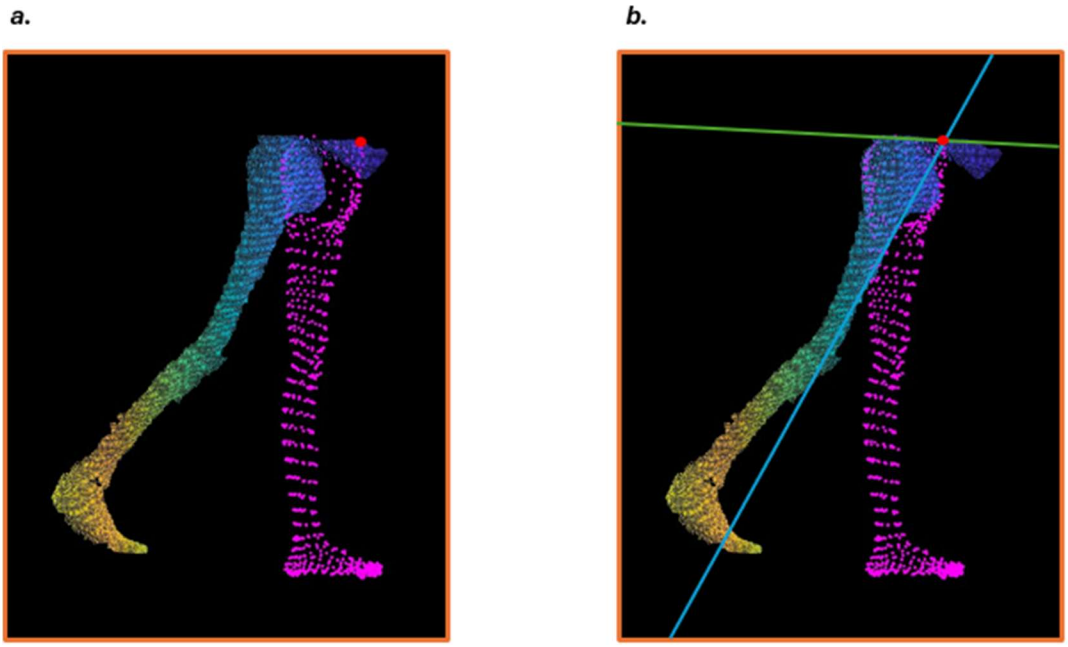

***Figure 23**: Correction of the $\boldsymbol{SMPL_{subject-specific}}$ model translation through the identification of a reference point, in a frame in which the hand is positioned notably ahead of the thigh. a) Misidentification of the most upper and anterior point point in $\boldsymbol{Dyn}$, resulting in incorrect translation of the $\boldsymbol{SMPL_{subject-specific}}$ model (in magenta). b) Translation correction by identifying a point defined by the intersection of a line fitting the profile of the thigh (in blue) and one fitting the profile of the pelvis (in green) in $\boldsymbol{Dyn}$.*

At this point, the AICP algorithm was applied to the $\boldsymbol{SMPL_{subject-specific}}$ model for each dynamic frame.

In the first frame of the gait cycle, the initial conditions of the $\boldsymbol{SMPL_{subject-specific}}$ model pose parameters were set to zero for both hip and knee, $\boldsymbol{\theta_{hip}}$ and $\boldsymbol{\theta_{knee}}$, thus starting from a neutral position.

For all other frames of the gait cycle, to speed up the process, the pose from the preceding frame was used as the initial condition for the subsequent frame.

The distinction of the first frame of the step cycle was also made in setting the values of the AICP search limits.

Specifically, in the first frame of the gait cycle the search range was selected wide as the model did not yet have a 'history' of motion to learn from, so the model was allowed to explore a wider range of possible configurations. In all other frames, the search limits were restricted around the pose parameters obtained from the previous frame. In both cases, the boundary values were set according to physiological limits to ensure permissible configurations, as shown in

Table 1.

Joint	***Angle***	***Upper limit***		***Lower limit***	
		First frame:	***Other frames:***	***First frame:***	***Other frames:***
Hip	Flexion-extension	$\frac{\pi}{4}$	$\frac{\pi}{7}$	$-\frac{\pi}{4}$	$-\frac{\pi}{7}$
	Abduction-adduction	0	0	0	0
	Intra-extra rotation	0	0	0	0
Knee	Flexion-extension	$\frac{\pi}{4}$	$\frac{\pi}{5}$	$-\frac{\pi}{4}$	$-\frac{\pi}{5}$
	Abduction-adduction	0	0	0	0
	Intra-extra rotation	0	0	0	0

***Table 1**: Physiological joint limits for the implementation of AICP algorithm*

Having acquired a sagittal view, the optimization process using the AICP algorithm was restricted to only the hip and knee flexion-extension angles. This decision was justified by the fact that, although the RGB-D camera provides 3D coordinates, its view is limited to the sagittal plane, resulting in insufficient and incomplete information regarding external rotation and

abduction-adduction movements. Additionally, it is important to note that the fitting process occurs between a 3D subject-specific model and a 2D+ dynamic point cloud, thereby confining the selection of the closest points to a single sagittal view.

Also, for the dynamic processing, a top-bottom approach was implemented starting from the hip to the knee. First of all, the hip joint was selected and the closest points of the dynamic point cloud, $\boldsymbol{Dyn}$, to the $\boldsymbol{SMPL_{subject-specific}}$ were identified in order to apply a rigid alignment between them around it. After that, the knee joint was selected and the $\boldsymbol{SMPL_{subject-specific}}$ model was split around it and the rigid transformation was applied.

For the implementation of the AICP algorithm proposed by Pellegrini et al, it is required to set a maximum number of iterations to be performed and a threshold on the value of the relative variation of the error between current and previous iteration in the alignment between the $\boldsymbol{SMPL_{subject-specific}}$ model and $\boldsymbol{Dyn}$.

Specifically, the maximum number of iterations was set to 8. The condition on the relative variation in hip and knee error between consecutive iterations was left unchanged, i.e. convergence was achieved when falling below 0.08%.

However, by applying only these two conditions, the $\boldsymbol{SMPL_{subject-specific}}$ model tended to be excessively flexed with respect to the dynamic configuration. In addition, it is important to consider that the algorithm proposed by Pellegrini et al was designed to align 3D models to 3D shapes, whereas this dynamic case is more challenging due to the need to align a 3D model with a 2D+ point cloud. To address these challenges and ensure that the $\boldsymbol{SMPL_{subject-specific}}$ model remained aligned within the limits defined by the dynamic configuration, it became necessary to introduce additional stop conditions in the AICP optimization of the hip and knee flexion angles, β_{hi} and β_{knee}. These conditions were determined based on the position of the knee joint center (for the hip angle) and the ankle joint center (for the knee angle) with respect to $\boldsymbol{Dyn}$.

This necessity arose from the difficulty of defining a maximum number of iterations and a threshold for the error discrepancy between consecutive iterations that could be generalized across all frames. Consequently, the introduction of these additional conditions was essential for the dynamic case.

Furthermore, these stop conditions allowed the process to be speeded up by preventing the exploration of all possible configurations and avoiding the optimization from getting stuck at local minima, i.e., when the model is already too flexed.

Regarding the stop condition on the optimization of the hip flexion angle, β_{hip}, the steps followed to correctly align the position of the knee by adjusting the hip angle were as follows:

3D points included between $\boldsymbol{^{I}KJC}$ *and* $(\boldsymbol{^{I}KJC} \pm \varepsilon)$,where $\varepsilon = 20$ pixels, were isolated in both $\boldsymbol{SMPL_{subject-specific}}$ and $\boldsymbol{Dyn}$. If the minimum z-coordinate of the $\boldsymbol{SMPL_{subject-specific}}$ model points fell below the minimum z-coordinate of points in $\boldsymbol{Dyn}$, implying excessive knee bending in the model, the optimization process related to the hip angle was interrupted. In order to refine the alignment between the $\boldsymbol{SMPL_{subject-specific}}$ model and the dynamic point cloud, the outer profiles of the knee were determined in both $\boldsymbol{SMPL_{subject-specific}}$ model and $\boldsymbol{Dyn}$ and from these, the boundary points (points with minimum and maximum y-coordinates) were selected to fit a line on the thigh of both $\boldsymbol{SMPL_{subject-specific}}$ and $\boldsymbol{Dyn}$. At this point, the knee flexion angle, β_{hip}, was decreased iteratively (thus extending the hip and bringing the model knee position backwards with respect to $\boldsymbol{Dyn}$) to guarantee that the number of points on the knee portion in the $\boldsymbol{SMPL_{subject-specific}}$ model that went beyond the fitted line on the outer knee profile in $\boldsymbol{Dyn}$ was less than 5%. .

In this way, the algorithm ensures that the knee center in the $\boldsymbol{SMPL_{subject-specific}}$ model does not exceed the maximum position established by the dynamic configuration data, progressively adjusting the hip angle to achieve better alignment.

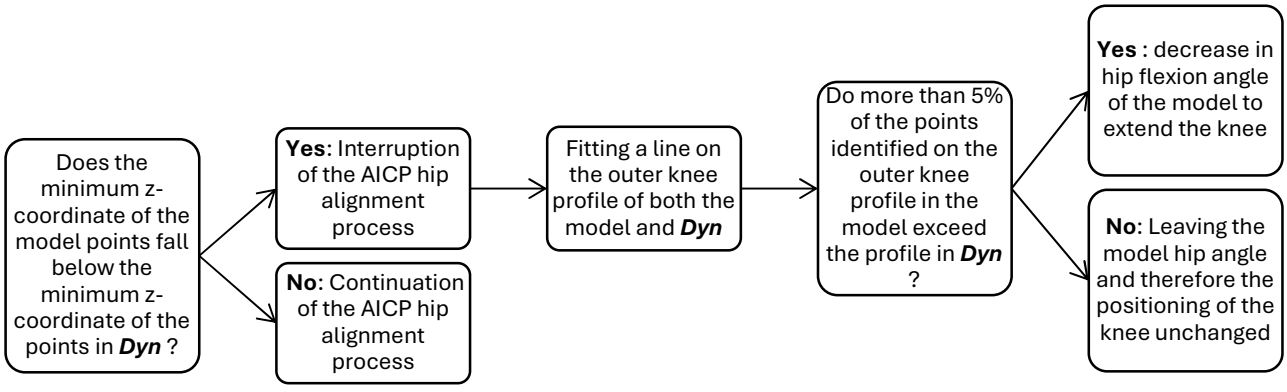

***Table 2**: Flow chart of the hip angle optimization process based on knee joint position in the $\boldsymbol{SMPL_{subject-specific}}$ model*

Moving on to the optimization of the knee angle, 3D points included between $^{I}\boldsymbol{AJC}$ and ($^{I}\boldsymbol{AJC} \pm \varepsilon$), where $\varepsilon = 20$ pixels, were isolated in both $\boldsymbol{SMPL_{subject-specific}}$ and $\boldsymbol{Dyn}$. If the maximum z-coordinate of the model points was greater than the maximum z-coordinate of points in $\boldsymbol{Dyn}$, indicating that in the sagittal plane the model ankle position was further back than in $\boldsymbol{Dyn}$, the optimization process of the knee angle was interrupted. At this point, the knee flexion angle in the $\boldsymbol{SMPL_{subject-specific}}$ model, β_{knee} ,was decreased iteratively (thus extending the knee and bringing the model ankle position forward with respect to $\boldsymbol{Dyn}$) to guarantee that the number of points on the ankle portion in the $\boldsymbol{SMPL_{subject-specific}}$ model that were behind those of $\boldsymbol{Dyn}$ (model points with a z-coordinate greater than the maximum z-coordinate of the $\boldsymbol{Dyn}$ points) was less than 5%. .

Similarly to the hip adjustment process, this iterative correction ensures that the model ankle position in the sagittal plane remains within the limits defined by the ankle position in the dynamic configuration, progressively refining the knee angle to achieve ankle alignment.

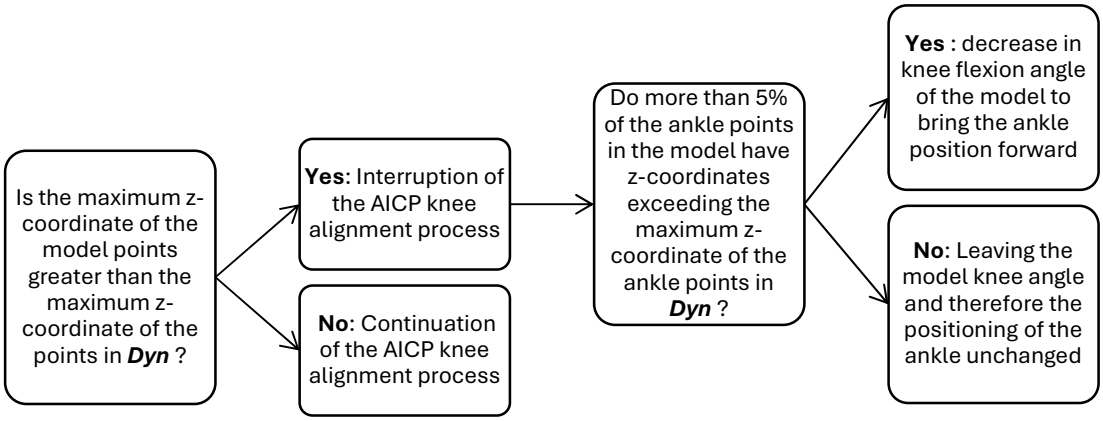

Table 3: *Flow chart of the knee angle optimization process based on ankle joint position in the* $\boldsymbol{SMPL_{subject-specific}}$ *model*

Once the AICP optimization cycle was complete, the $\boldsymbol{SMPL_{subject-specific}}$ model was scaled on the vertical axis again.

Under the assumption that the position of the joint center extracted in the $\boldsymbol{SMPL_{subject-specific}}$ model remains unchanged during walking, an additional refinement was introduced by considering the relative position of the knee and ankle center, $^{I}\boldsymbol{KJC}$ and $^{I}\boldsymbol{AJC}$ with respect to the body's surface of the $\boldsymbol{SMPL_{subject-specific}}$ model. More in detail, 3D points included between $^{I}\boldsymbol{KJC}$ and ($^{I}\boldsymbol{KJC} \pm \varepsilon$), where $\varepsilon = 20$ pixels, were isolated in the

$\boldsymbol{SMPL_{subject-specific}}$ model and its relative position of the $\boldsymbol{^{I}KJC}$ with respect to the body's surface on the sagittal axis was calculated.

Specifically, the hip angle, β_{hip}, was iteratively adjusted (flexed or extended) to ensure that the relative position of the $\boldsymbol{^{I}KJC}$ with respect to the body's surface in $\boldsymbol{Dyn}$ matched the one calculated in the $\boldsymbol{SMPL_{subject-specific}}$ model.

The same procedure was followed in refining the positioning of the ankle by acting on the knee angle in the $\boldsymbol{SMPL_{subject-specific}}$ model. Under the same assumption as before, 3D points included between $\boldsymbol{^{I}AJC}$ and $(\boldsymbol{^{I}AJC} \pm \varepsilon)$, where $\varepsilon = 10$ pixels, were isolated in the $\boldsymbol{SMPL_{subject-specific}}$ model and its relative position of the $\boldsymbol{^{I}AJC}$ with respect to the body's surface on the sagittal axis was calculated.

Specifically, the knee angle, β_{knee}, was iteratively adjusted (flexed or extended) to ensure that the relative position of the $\boldsymbol{^{I}AJC}$ with respect to the body's surface in $\boldsymbol{Dyn}$ matched the one calculated in the $\boldsymbol{SMPL_{subject-specific}}$ model.

In this case, with respect to the process of refining the knee position, additional control was provided to ensure better foot alignment using the heel point and knee joint center as references. The angle between the vectors connecting the knee joint center to the heel in the $\boldsymbol{SMPL_{subject-specific}}$ model and in $\boldsymbol{Dyn}$ was calculated. Based on the sign of the cross product between the two vectors, the knee angle, β_{knee}, was flexed or extended iteratively to guarantee that the angle between the two vectors was less than a defined threshold of 0.5 degrees.

As in static case, plantar-dorsiflexion, α_{ankle}, and internal-external rotation of the foot, β_{ankle}, were not obtained from the AICP algorithm but were instead calculated using the previously determined coordinates of the heel and toe and applied to the $\boldsymbol{SMPL_{subject-specific}}$ model. After aligning the $\boldsymbol{SMPL_{subject-specific}}$ to $\boldsymbol{Dyn}$ for each frame, the $\boldsymbol{SMPL_{subject-specific}}$ model was rescaled on the vertical axis to refine its length and as a final step , the joint centers $\boldsymbol{^{I}AJC}, \boldsymbol{^{I}KJC}, \boldsymbol{^{I}HJC}$ were estimated by applying the joint regression matrix $\boldsymbol{J}$ to the aligned $\boldsymbol{SMPL_{subject-specific}}$.

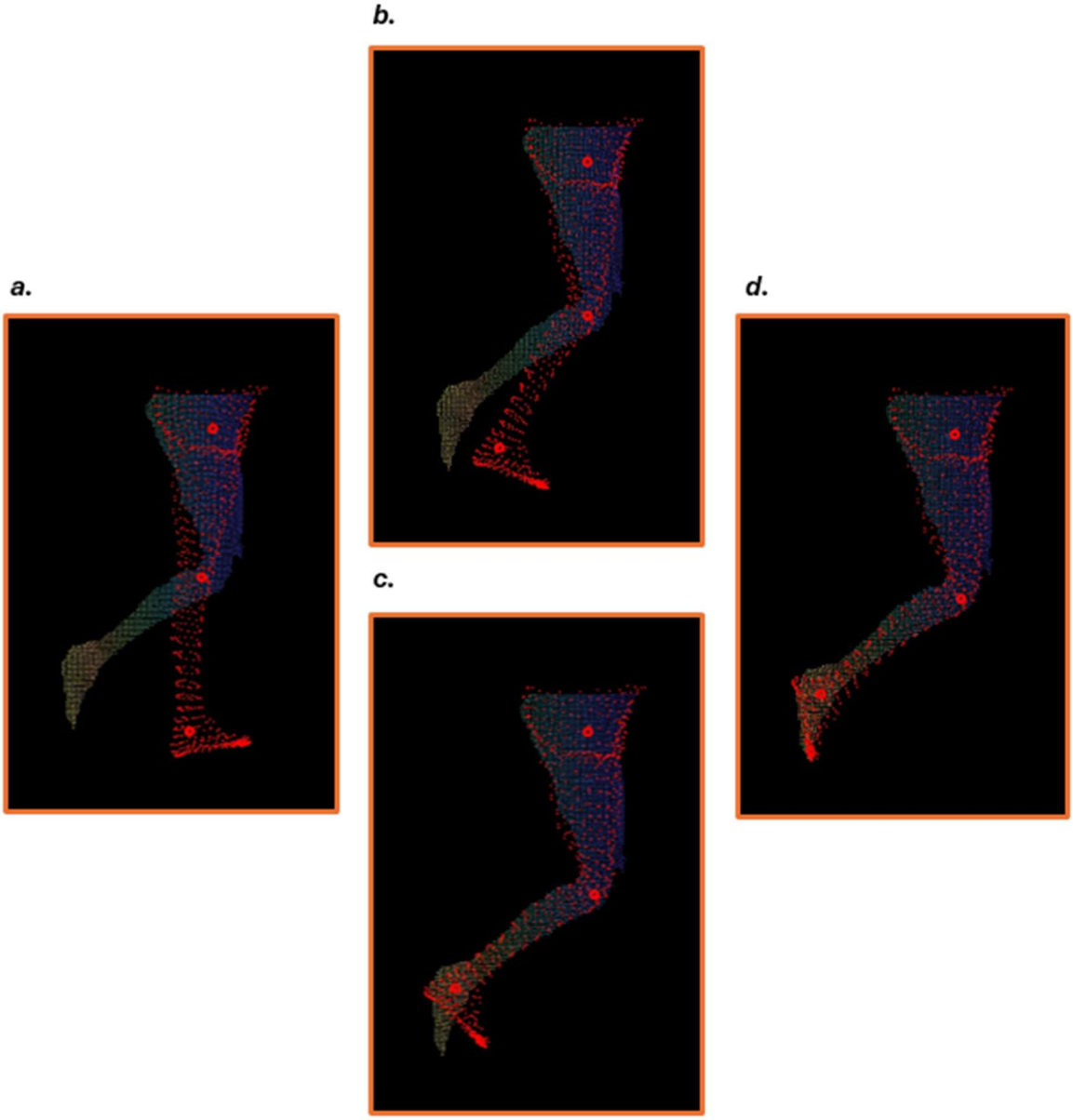

Figure 24: *AICP implementation following a top-down approach from hip to ankle and subsequent foot alignment. a) subject specific model (in red) and a dynamic point cloud (initial condition), b) Rigid alignment around the hip joint, c) Rigid alignment around the knee joint, d) The final alignment between subject specific model and a dynamic point cloud, including the foot.*

3.5.1. Estimation of joint kinematics

The joint kinematics in the sagittal plane were extracted by calculating the angles formed between the lines connecting the joint centers. The plantar-dorsi flexion angle of the ankle was calculated as the angle between the segment connecting the heel and toe points of the foot and the $^{I}\boldsymbol{AJC} - {}^{I}\boldsymbol{KJC}$ vector. The knee joint's flexion-extension angle was identified as the angle between the $^{I}\boldsymbol{AJC} - {}^{I}\boldsymbol{KJC}$ and $^{I}\boldsymbol{KJC} - {}^{I}\boldsymbol{HJC}$ vectors. For the hip joint, the flexion-extension angle was calculated as the angle between the $^{I}\boldsymbol{KJC} - {}^{I}\boldsymbol{HJC}$ vector and the horizontal axis.

Chapter 4

4. Results and Discussion

In this chapter, the results obtained are presented, with a comparison between the proposed 3D ML approach and a 12-camera stereophotogrammetric system (Oqus 400 Qualisys medical AB, Gothenburg, Sweden), which represent the gold standard for gait analysis.

The sagittal lower limb kinematic curves extracted from both ML and MB systems were mean-centered, then filtered with a fourth-order low-pass Butterworth filter with a cutoff frequency of 7 Hz and time-normalized to the gait cycle (1-100%).

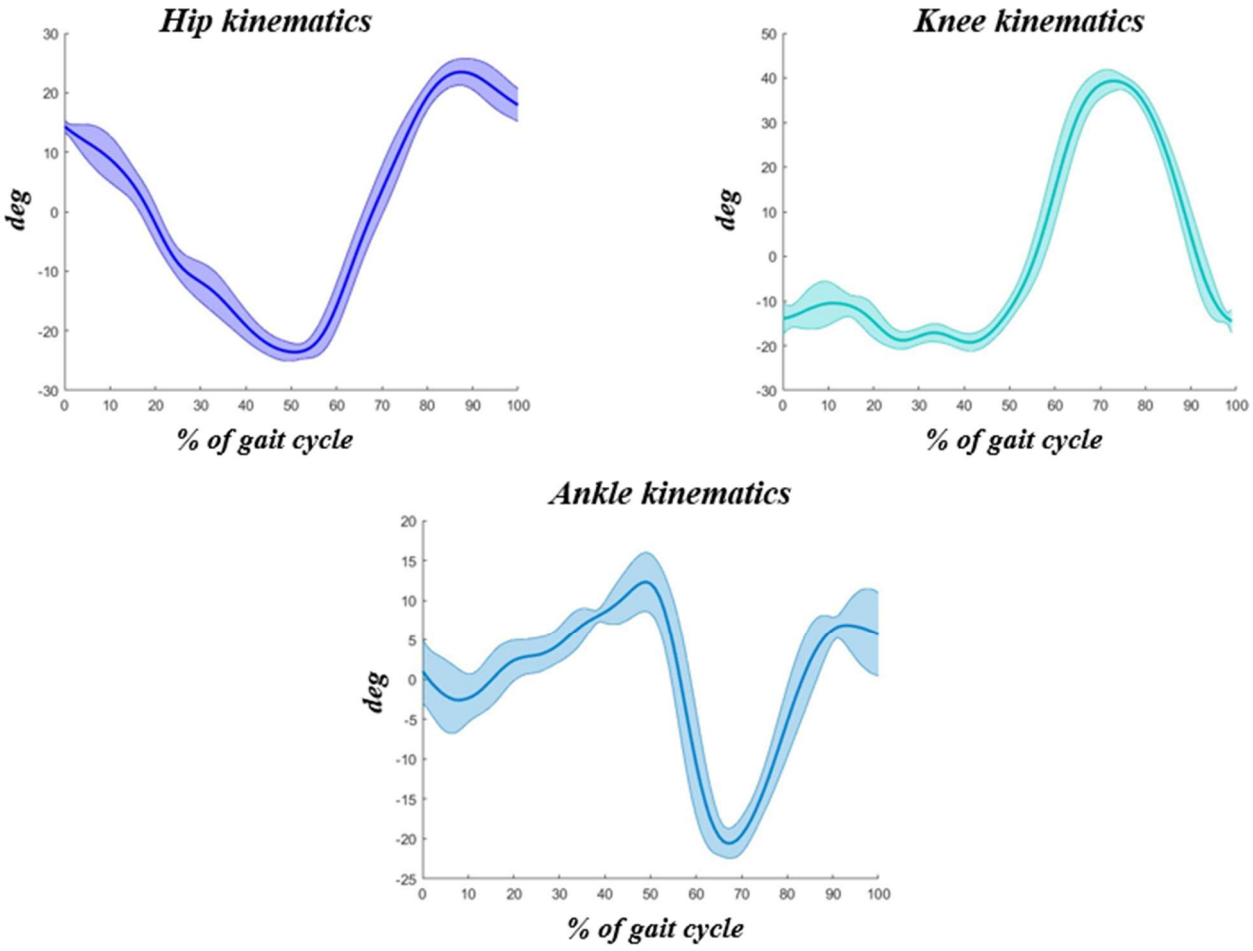

***Figure 25**: Sagittal lower limb joint kinematics of one subject extracted from the proposed ML protocol (hip, knee and ankle) averaged trials (average: dashed lines; standard deviation: shaded area).*

From the processed kinematic curves, seven clinically relevant key gait features were extracted (**Figure 26**), four related to the knee, two to the ankle, and one to the hip, as follows:

- K1: Knee flexion which occurs at 0% of the gait cycle during the initial contact.
- K2: Maximum knee flexion which occurs from 0% to 40% of the gait cycle during the loading response.
- K3: Maximum knee extension which occurs from 25% to 75% of the gait cycle during the stance phase.
- K5: Maximum knee flexion which occurs from 50% to 100% of the gait cycle during the swing phase.
- A3: Maximum ankle dorsiflexion which occurs from 25% to 75% of the gait cycle during the stance phase.
- A5: Maximum ankle plantarflexion which occurs from 50% to 100% of the gait cycle during the swing phase.
- H3: Maximum hip extension which occurs from 25% to 75% of the gait cycle during the stance phase.

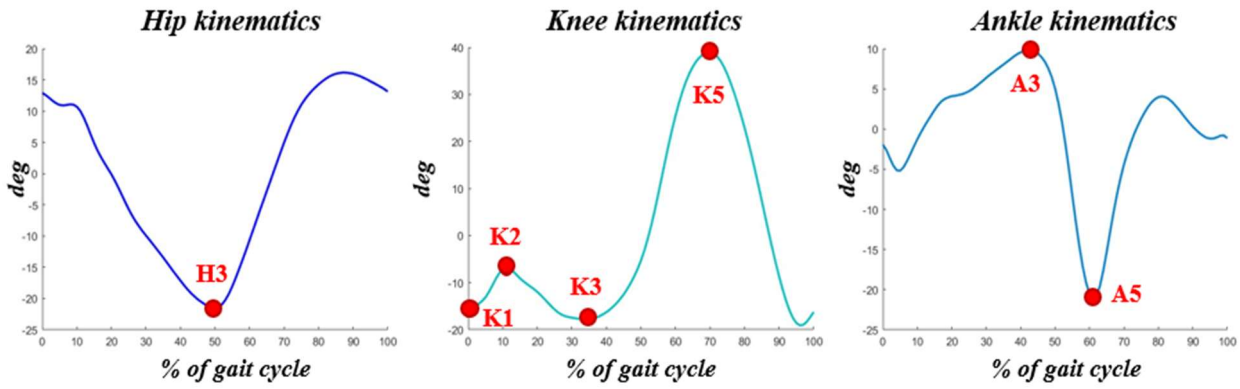

***Figure 26** : Illustration of the seven clinically relevant key gait features extracted from the joint kinematic curves*

The performance of the proposed 3D ML method was assessed using statistical parameters computed for each extracted gait feature. Specifically, the accuracy of the ML estimates was assessed using the mean absolute difference (MAD) calculated with respect to the MB system values, along with 95% confidence interval (95% CI), according to the following formula:

$$MAD_{k,s,t} = |MB_{k,s,t} - ML_{k,s,t}|$$

Where $MB_{k,s,t}$ and $ML_{k,s,t}$ represent the values obtained from the MB and ML systems respectively, corresponding to the gait feature k, for the gait trial t and relative to subject s. The obtained $MAD_{k,s,t}$ values were then averaged across gait trials and subjects, as follows:

$$MAD_k = \frac{1}{N_S}\sum_{s=1}^{N_S}\frac{1}{N_T}\sum_{t=1}^{N_T} MAD_{k,s,t}$$

Where N_S represents the total number of subjects (14) and N_T the number of recorded gait trials per subject (6). The results in terms of MAD values are shown in **Table 4**.

	Gait Variables (deg)	*MAD (95% CI)*
Knee	***Initial contact (K1)***	4.7 [4.1, 5.3]
	Load (K2)	3.2 [2.7, 3.7]
	Stance (K3)	2.6 [2.3, 2.9]
	Swing (K5)	2.8 [2.3, 3.1]
Ankle	***Stance (A3)***	4.1 [3.6, 4.7]
	Swing (A5)	4.5 [3.8, 5.1]
Hip	***Stance (H3)***	5.4 [4.8, 5.9]

Table 4 *: Mean absolute difference (MAD) of the key gait features along with 95 % CI: 95% of confidence interval averaged over six trials per fourteen subjects with respect to MB protocol.*

The lowest MAD values were obtained for the knee, except for that corresponding to K1. Specifically, knee MAD values range between 2.6° in the stance phase and 4.7° during initial contact. For the ankle, the MAD ranges between 4.1° in the stance phase and 4.5° in the swing phase. For the hip, the reported MAD is the higher with a value of 5.4°.

To assess reliability, intraclass correlation coefficients (ICC) based on absolute agreement and two-way random effects were computed in both ML and MB protocols for each gait feature and then compared. The ICC(2, k) values were calculated using data collected from 14 subjects across three gait cycles for each side, thus distinguishing right and left trials and subsequently computing an overall average. Results for ICC(2, k) for both ML and MB protocols, are reported in

Table 5.

	Gait Variables (deg)	*ICC ML*	*ICC MB*
Knee	***Initial contact (K1)***	0.90	0.95
	Load (K2)	0.79	0.82
	Stance (K3)	0.82	0.90
	Swing (K5)	0.82	0.96
Ankle	***Stance (A3)***	0.91	0.86
	Swing (A5)	0.87	0.90
Hip	***Stance (H3)***	0.90	0.95

***Table 5**: Reliability of the markerless (ML) and marker-based system (MB) methods in terms of ICC, computed for each gait variable.*

According to [51], ICC values above 0.90 indicate excellent reliability, values between 0.75 and 0.90 denote good reliability, values between 0.5 and 0.75 reflect moderate reliability, and finally, any value below 0.5 indicates poor reliability. Based on the reported values, the ML protocol demonstrates high reliability for all gait features. Specifically, the ICC value for the knee during initial contact was 0.90 for the ML system and 0.95 for the MB system. These values indicate excellent reliability in both methods. For the knee during the loading phase, the ICC value was 0.79 (ML) and 0.82 (MB), indicating good reliability in both methods. During the stance phase, knee ICC value was 0.82 (ML) and 0.90 (MB). In this case, the ML protocol demonstrated good reliability, whereas the MB protocol exhibited excellent reliability. Similarly, during the swing phase, knee ICC value was 0.82 (ML) and 0.96 (MB), confirming the same reliability classification as in the stance phase. Proceeding with the analysis of the ankle results, during the stance phase the ICC value was 0.91 (ML) and 0.86 (MB). In contrast to the previous knee results, the ML method demonstrated excellent reliability, while the MB method showed good reliability. During the swing phase, ankle ICC value was 0.87 (ML) and 0.90 (MB), thus reflecting good reliability for both methods. Finally, for the hip during the stance phase ICC value was 0.90 (ML) and 0.95 (MB). The ML method showed good reliability for the hip in this phase, while the MB method exhibited excellent reliability.

The final statistical parameter analyzed for both protocols was the Gait Variability Standard Deviation (GVSD), an index of intra-trial variability. For each joint angle k of each subject s this value was computed as follows:

$$GVSD_{k.s} = \sqrt{\frac{\sum_{j=1}^{T}\sum_{i=1}^{N}(X_{k,ij} - X_{k,j})^2}{T(N-1)}}$$

Where T is the number of time samples (100), N is the number of trials per side (3), $X_{k,ij}$ the kinematic waveform associated with the joint angle k, from the stride i defined over j time samples and $X_{k,j}$ the mean of the N waveforms defined for each time instant j [52]. Right and left trials were treated separately and then an overall average was calculated across subjects and sides.

Results for GVSD for both ML and MB protocols, are reported in

Table 6.

Joint	*GVSD ML (deg)*	*GVSD MB (deg)*
Knee	4.6 [3.9, 5.3]	3.6 [2.8, 4.3]
Ankle	3.4 [3.0, 3.8]	1.9 [1.5, 2.2]
Hip	3.2 [2.7, 3.7]	2.3 [1.7, 2.9]

***Table 6**: Gait variability standard deviation (GVSD) of the markerless (ML) and marker-based system (MB) methods for each joint angle along with 95 % CI, averaged over six trials per fourteen subjects.*

GVSD values for each joint angle are comparable, with an average difference of 1°, proving ML protocol to have similar variability to MB. The greatest difference between the two methods is reported for the ankle (1.5°).

It is important to note that the differences observed between the ML and the MB protocols are, in part, attributable to different methods by which the two systems calculate joint angles. The ML protocol calculates joint angles using simple trigonometric formulas by connecting the joint centers, whereas the MB system calculates joint kinematics through the decomposition of Euler angles. In addition, further differences arise from the fact that acquisitions with the ML and MB systems were not conducted simultaneously to avoid infrared interference between IR sensors.

Clinically, accuracy errors in terms of MAD value between 2° and 5° are typically considered acceptable but they need careful interpretation [53]. Errors in the accuracy of the ML method compared to the MB method result in part from a different definition of the anatomical axes in the two protocols. In fact, for the hip, where the highest mean absolute difference (5.4°) is reported, the proposed ML protocol does not consider the variable inclination of the pelvis during gait but calculates the hip angle as the angle between the segment connecting the hip joint center to the knee joint center and the horizontal axis. For the calculation of the knee angle, the two protocols involve a different identification of the hip joint center. The ML method identifies it as coincident with the greater trochanter, while the MB method locates it at the geometric center of the acetabulum, derived using anthropometric regression equations. For the ankle, accuracy errors arise from the fact that the ML protocol identifies the foot's antero-posterior axis as the principal axis of the best-fitting inertial ellipsoid, whereas the MB protocol derives it from the marker positions on the second metatarsal joint and calcaneus. Residual error in terms of MAD value for the ankle are also due to inaccuracies in estimating the ankle joint coordinate caused by an improper reconstruction of foot and distal part of the shank of the 3D subject-specific model, as further explained in the paragraph "Factors influencing the accuracy of 3D lower-limb model creation". As for the knee, the highest MAD value is obtained during the initial contact (0% of gait cycle). This error is related to the first knee flexion angle value recorded in the kinematic curve and can be attributed to an imprecise identification of the initial frame of the gait cycle, which may also be influenced by the lower frame rate of the ML protocol (30 Hz) compared to that of the MB protocol (100 Hz).

Despite the mentioned issues, the ICC values of the ML protocol are comparable with those obtained for the gold standard (MB) and range from good to excellent (ICC > 0.75), demonstrating consistent estimates. The largest discrepancy in terms of ICC values is observed in the knee kinematics during the swing phase (0.14) due to technological limitations attributable to the depth sensor, which fails to reconstruct depth values at high speeds. This issue will be detailed in the paragraph "Factors influencing the accuracy of joint kinematic estimation".

4.1. Factors influencing the accuracy of 3D lower-limb model creation

This paragraph highlights the need for accurate positioning of the subject to be acquired for the proper reconstruction of the 3D lower-limb model. This is a critical step as even small variations

in subject's positioning across the three static views (frontal, sagittal, posterior) can lead to inaccuracies in the reconstruction of the model, especially for the distal part of the shank and for the foot. Consequently, it is essential that the subject during static acquisitions maintains the same position to be able to correctly identify the key points to be matched among the different views. However, the subject tends to unintentionally alter the orientation of feet and legs, and this leads to a distorted reconstruction of the 3D lower-limb model, despite the keypoints being correctly identified, since the subject's pose has not been consistently replicated in the views.

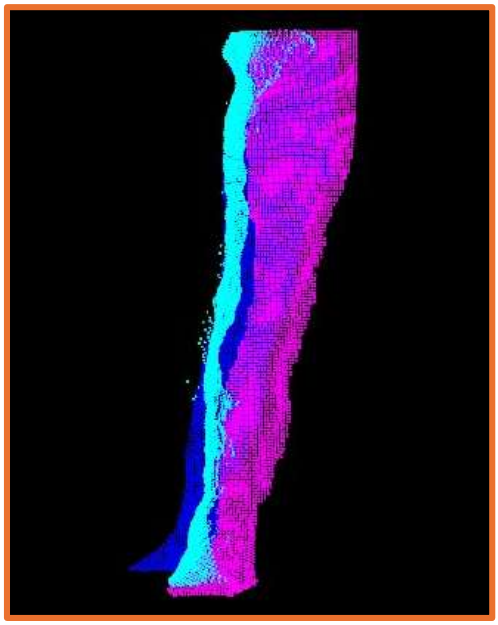

***Figure 27** : An example of incorrect 3D subject-specific reconstruction caused by a different foot position assumed in the posterior view (blue point cloud), where the foot appears to be spread apart with respect to the frontal (magenta point cloud) and sagittal (cyan point cloud) views.*

A possible solution to mitigate this problem could be to integrate on the green carpet, on which the acquisitions are made, indications of where to place the feet with reference points representing the indicative toe and heel positions to be assumed, as shown in **Figure 28**. The instructions should be positioned to ensure that the subject's feet are parallel and slightly apart, allowing for an accurate reconstruction of the foot in a neutral pose. It should still be considered that the subjects acquired are patients with pathological conditions, who often experience difficulties in maintaining a stable pose due to pain, balance impairments or foot deformities.

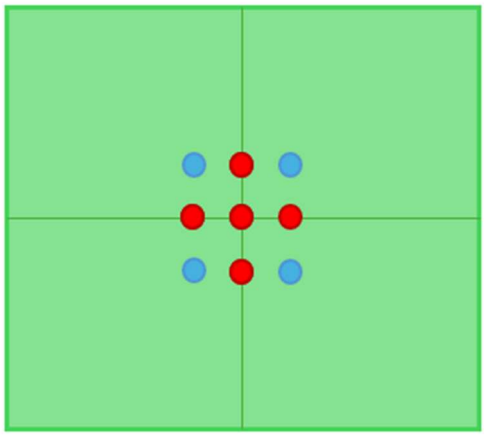

***Figure 28**: Schematic illustration of the indications to be placed on the green carpet to aid patient in correct positioning during static acquisitions. The red dots indicate the central references of the carpet, while the blue dots are for toe and heel positioning.*

4.2. Factors influencing the accuracy of joint kinematic estimation

As mentioned above, one of the factors that must be considered when comparing the performance of the two methods ML and MB, is the non-simultaneity of acquisitions to avoid IR interference, which primarily affects the results in terms of MAD value. Another primary factor in determining inaccuracies in the estimation of joint kinematics is related to the limitations of the camera depth sensor. Specifically, due to the limited exposure time of the RGB-D camera, capturing movements at high speed may cause motion blur in RGB images, which can result in artifacts such as holes or fake boundaries in the corresponding depth images [54]. This issue leads to an improper alignment between the depth image and the RGB image specifically at the shank and foot, where artifacts in depth data cause a narrowing effect that mainly affects the computation of ankle and knee kinematics. To partially mitigate this issue and improve the fitting of the subject-specific model to the dynamic point cloud, in the swing phase frames where this narrowing occurs (**Figure 19**), the scaling factor is computed using the value obtained from the previous frame rather than relying on the toe coordinates, as explained in the paragraph "**Joint centers trajectories estimation**". Despite this, in such cases the toe coordinates are often incorrectly estimated, leading to a miscalculation of the foot inclination. and consequently, improper mesh deformation and ankle joint position computation. This results in improper mesh deformation and an incorrect estimation of the ankle joint position.

Chapter 5

5. Conclusion

Optoelectronic stereophotogrammetry represents the gold standard for instrumental gait analysis because of its submillimeter spatial accuracy and high temporal resolution down to the millisecond. Nevertheless, its clinical application has some practical disadvantages mainly related to extensive equipment requirements, high cost and long setup time. In ambulatory settings where these resources are limited, markerless gait analysis is emerging as a promising clinical alternative especially in long-term monitoring programs such as CPUP. The recent development of low-cost RGB-D cameras has led to great advances for video-based markerless technology, offering opportunities for gait analysis with single-camera setups. Based on this, this study proposes a 3D ML model-based approach using a single RGB-D camera and the 3D Skinned Multi Person Linear model (SMPL) to estimate sagittal lower-limb joint kinematics. The innovative aspect of this ML approach lies in its ability to reconstruct a 3D model of the subject's lower limb using three static acquisitions from different views (frontal, sagittal, and posterior), still relying on a single RGB-D camera, unlike traditional 3D approaches that require multiple camera set-up, which is impractical for clinical applications. By extending the 2D ML protocol proposed in [25], this approach introduces a 3D version that compensates for out-of-plane movements and positional changes during gait, enhancing the robustness and accuracy of lower limb joint kinematics in the sagittal plane. Furthermore, the 3D approach eliminates the need for manual identification of anatomical landmarks during model calibration, thereby automating the process and reducing variability and operator dependency. Another key advantage of this method is its ability to integrate a deterministic model-based approach for estimating human joint kinematics, ensuring repeatability and consistency in pose tracking which are key features for the clinical applicability of the protocol. Simultaneously, the SMPL model incorporates a data-driven learning mechanism for pose and shape estimation, leveraging a large dataset of pre-registered human body scans. This aspect aligns with AI-based methodologies, which employ machine learning techniques during the training phase to automatically learn patterns and features from data. As a result, the proposed approach combines the advantages of deterministic human pose estimation with the flexibility of data-driven training by automatically estimating parameters from the aligned scans.

To validate this approach, five CPUP patients with cerebral palsy and nine with clubfoot were recruited to perform six self-selected gait trials (three per side) on a 5-meter-long walkway. For each patient, four static acquisitions were made from different camera views (frontal, sagittal and posterior) in order to obtain a static posture by point cloud merging, aligning specific keypoints across the views. This static posture was used to calibrate a 3D lower-limb SMPL model, in which the pelvis, thigh, shank, and foot segments are interconnected through the hip, knee, and ankle joints. The obtained subject-specific SMPL was then aligned to each dynamic frame of the gait cycle using AICP, to estimate 3D joint kinematics. From these, seven significant gait characteristics were extracted and analyzed to validate the proposed ML protocol against a 3D MB clinical gait analysis protocol. The accuracy, reliability, and variability of the ML protocol were assessed in terms of Mean Absolute Differences (MAD) relative to the MB system, Intraclass Correlation Coefficients (ICC), and Gait Variability Standard Deviations (GVSD), respectively. It is important to highlight that, under the assumption of movement repeatability, acquisitions were not performed synchronously by the MB and MS systems to avoid interference, as the wavelength of the Azure Kinect IR sensor is the same as the Qualisys system. This aspect, combined with the fact that the ML and MB systems adopt a different definition of the anatomical axes, results in a discrepancy in measurement accuracy compared to the MB system, with MAD values ranging between 2.6° and 5.4°. Despite this difference, these values are still within a clinically acceptable range. Moreover, the proposed ML protocol exhibited high reliability (ICC>0.75) for each extracted gait feature, with the largest discrepancy to MB values obtained for the knee swing due to the inability of the depth sensor to reconstruct depth values at high speeds. Finally, ML protocol proved to have similar variability to MB system, with an average difference in GVSD value of 1°. These results are promising and demonstrate how the proposed 3D ML protocol, based on a single RGB-D camera, can be a portable and low-cost clinical alternative for gait analysis, making it a viable solution especially for follow-up and long-term monitoring programs. In the future, thanks to the ability to reconstruct a subject-specific 3D model, this method could be enriched by including the estimation of volumetric parameters to assess and monitor asymmetries, going beyond traditional gait analysis.

References

[1] A. V. Ruescas Nicolau, H. De Rosario, F. Basso Della-Vedova, E. Parrilla Bernabé, M.-C. Juan and J. López-Pascual, "Accuracy of a 3D temporal scanning system for gait analysis: Comparative with a marker-based photogrammetry system," *Gait & Posture, Volume 97,* pp. 28-34, 2022.

[2] S. L. Colyer, M. Evans, D. P. Cosker and A. I. T. Salo, "A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System," *Colyer et al. Sports Medicine- Open,* pp. 4-24, 2018.

[3] A. Harveya and J. W. Gorterb, "Video gait analysis for ambulatory children with cerebral palsy: Why, when, where and how!," *Gait & Posture 33,* pp. 501-503, 2011.

[4] E. Pantzar-Castilla, D. Balta, U. Della Croce, A. Cereatti and J. Riad, "Feasibility and usefulness of video-based markerless two-dimensional automated gait analysis, in providing objective quantification of gait and complementing the evaluation of gait in children with cerebral palsy," *BMC Musculoskeletal Disorders,* vol. 25, no. 747, 2024.

[5] W. W. T. Lam, Y. M. Tang and K. N. K. Fong, "A systematic review of the applications of markerless motion capture (MMC) technology for clinical measurement in rehabilitation," *Journal of NeuroEngineering and Rehabilitation,* vol. 20, no. 57, 2023.

[6] K. Vitrikas, H. Dalton and D. Grant, "Cerebral Palsy: An Overview," *American family physician,* vol. 101, no. 4, pp. 213-220, 2020.

[7] F. M. Chang, J. T. Rhodes, K. M. Flynn and a. J. Carollo, "The Role of Gait Analysis in Treating Gait Abnormalities in Cerebral Palsy," *Orthopedic Clinics,* vol. 41, no. 4, pp. 489-506, 2010.

[8] S. Armand, G. Decoulon and A. Bonnefoy-Mazure, "Gait analysis in children with cerebral palsy," *EFORT Open Reviews,* vol. 1, no. 12, pp. 448-460, 2016.

[9] T. F. Winters, J. R. Gage and R. Hicks, "Gait patterns in spastic hemiplegia in children and young adults," *J Bone Joint Surg Am,* vol. 69, no. 3, pp. 437-441, 1987.

[10] J. Rodda and H. K. Graham, "Classification of gait patterns in spastic hemiplegia and spastic diplegia: a basis for a management algorithm," *european journal of neurology,* vol. 8, no. 5, pp. 98-108, 2001.

[11] S. Basita and K. Khoshhal, "Genetics of clubfoot; recent progress and future perspectives," *European Journal of Medical Genetics ,* vol. 61 , no. 2, pp. 107-113, 2017.

[12] C. A. Gurnett, F. Alaee, L. M. Kruse, D. M. Desruisseau, J. T. Hecht, C. A. Wise, A. M. Bowcock and M. B. Dobbs, "Asymmetric Lower-Limb Malformations in Individuals," *The American Journal of Human Genetics,* vol. 83, no. 5, pp. 616-622, 2008.

[13] M. B. Dobbs and C. A. Gurnett, "The 2017 ABJS Nicolas Andry Award: Advancing Personalized Medicine for Clubfoot Through Translational Research," *Clinical Orthopaedics Related Research,* vol. 475, no. 6, pp. 1716-1725, 2017.

[14] L. A. Karol, K. Jeans and R. ElHawary, "Gait Analysis after Initial Nonoperative Treatment for Clubfeet: Intermediate Term Followup at Age 5," *Clin Orthop Relat Res.,* vol. 467, no. 5, pp. 1206-1213, 2009.

[15] A. Campos da Paz, A. Ramalho, A. Momura, L. Braga and M. Almeida, "Gait Analysis in Clubfeet: An Experimental Study," *Simons, G.W. (eds) The Clubfoot. Springer, New York, NY,* 1994.

[16] F. Roggio, S. Ravalli, G. Maugeri, A. Bianco, A. Palma, M. Di Rosa and G. Musumeci, "Technological advancements in the analysis of human motion and posture management through digital devices," *World Journal of Orthopedics,* vol. 12, no. 7, pp. 467-484, 2021.

[17] E. K. A. Surer, "Methods and Technologies for Gait Analysis," in *Computer Analysis of Human Behavior*, Salah, A., Gevers, T., 2011, pp. 105-123.

[18] F. Stief, "Variations of Marker Sets and Models," in *Handbook of Human Motion*, Müller, B.; Wolf, S.I., 2016, pp. 1-18.

[19] W. Flores-Fuentes, G. Trujillo-Hernándeza, I. Y. Alba-Corpusa, J. C. Rodríguez-Quiñoneza, J. E. Mirada-Vegad, D. Hernández-Balbuenaa, F. N. Murrieta-Ricoc and O. Sergiyenkob, "3D spatial measurement for model reconstruction: A review," *Measurement,* vol. 207, no. 112321, 2023.

[20] M. Zollhöfer, "Commodity RGB-D Sensors: Data Acquisition," in *RGB-D Image Analysis and Processing. Advances in Computer Vision and Pattern Recognition.* , Rosin, P. L., Lai, Y.-K., Shao, L., Liu, Y., 2019, pp. 3-13.

[21] A. Fusiello, E. Trucco and A. Verri, "A compact algorithm for rectification of stereo pairs," *Machine Vision and Applications,* vol. 12, no. 1, pp. 16-22, 2000.

[22] H. Sarbolandi, D. Lefloch and A. Kolb, "Kinect range sensing: Structured-light versus Time-of-Flight Kinect," *Computer Vision and Image Understanding,* vol. 139, pp. 1-20, 2015.

[23] G. Kurillo, E. Hemingway, M.-L. Cheng and L. Cheng, "Evaluating the Accuracy of the Azure Kinect and Kinect v2," *Sensors,* vol. 22, no. 2469, 2022.

[24] J. A. Albert, V. Owolabi, A. Gebel, C. M. Brahms, U. Granacher and B. Arnrich, "Evaluation of the Pose Tracking Performance of the Azure Kinect and Kinect v2 for Gait Analysis in Comparison with a Gold Standard: A Pilot Study," *Sensors,* vol. 20, no. 5104, 2020.

[25] D. Balta, G. Figari, G. Paolini, E. Pantzar-Castilla, J. Riad, U. Della Croce and A. Cereatti, "A model-based markerless protocol for clinical gait analysis based on a single RGB-depth camera: concurrent validation on patients with cerebral palsy," *IEEE Access,* vol. 11, pp. 144377- 144393., 2023.

[26] L. Mündermann, S. Corazza and T. Andriacchi, "The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications," *Journal of NeuroEngineering and Rehabilitation,* vol. 3, 2006.

[27] N. Sarafianos, B. Boteanu, B. Ionescu and I. A. Kakadiaris, "3D Human pose estimation: A review of the literature and analysis of covariates," *Computer Vision and Image Understanding,* vol. 152, pp. 1-20, 2016.

[28] E. Ceseracciu, Z. Sawacha and C. Cobelli, "Comparison of Markerless and Marker-Based Motion Capture Technologies through Simultaneous Data Collection during Gait: Proof of Concept," *PLOS ONE,* vol. 9, no. 3, p. e87640, 2014.

[29] S. S. K. N. D. Pellegrini, "A Generalisation of the ICP Algorithm for," in *British Machine Vision Conference*, 2008.

[30] A. Castelli, G. Paolini, A. Cereatti and U. Della Croce, "A 2D markerless gait analysis methodology: Validation on healthy subjects," *Computational and Mathematical Methods,* vol. 2015, 2015.

[31] E. Pantzar-Castilla, A. Cereatti, G. Figari, N. Valeri, G. Paolini, U. Della Croce, A. Magnuson and J. Riad, "Knee joint sagittal plane movement in cerebral palsy: a comparative study of 2-dimensional markerless video and 3-dimensional gait analysis," *Acta Orthopaedica,* vol. 89, no. 6, pp. 656-661, 2018.

[32] Z. Liu, "3D Skeletal Tracking on Azure Kinect," in *ICIP2019 Microsoft Industry Workshop – Machine Learning and Computer Vision Applications*, 2019.

[33] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis and M. Bethge, "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning," *Nature Neuroscience,* vol. 21, no. 9, pp. 1281-89, 2018.

[34] D. Pavllo, C. Feichtenhofer, D. Grangier and M. Auli, "3D human pose estimation in video with temporal convolutions and semi-supervised training," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* , 2018.

[35] M. Omran, C. Lassner, G. Pons-Moll, P. V. Gehler and B. Schiele, "Neural Body Fitting: Unifying Deep Learning and Model-Based Human Pose and Shape Estimation," in *2018 International Conference on 3D Vision (3DV)*, 2018.

[36] N. Magnenat-Thalmann, R. Laperrière and D. Thalmann, "Joint-dependent local deformations for hand animation and object grasping," in *Proceedings on Graphics Interface '88*, Toronto,Ont., Canada, 1988.

[37] N. Abu Rumman and M. Fratarcangeli, "State of the Art in Skinning Techniques for Articulated Deformable Characters," *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - GRAPP,* pp. 200-212, 2016.

[38] L. Kavan, "Part I: Direct Skinning Methods and Deformation Primitives," *SIGGRAPH Course 2014 — Skinning: Real-time Shape Deformation,* pp. 1-11, 2014.

[39] L. Kavan, S. Collins, J. Žára and C. O'Sullivan, "Skinning with dual quaternions," *I3D '07: Proceedings of the 2007 symposium on Interactive 3D graphics and games,* pp. 39-46, 2007.

[40] L. Kavan and O. Sorkine, "Elasticity-Inspired Deformers for Character Articulation," in *ACM SIGGRAPH Asia*, Singapore, 2012.

[41] J. P. Lewis, M. Cordner and N. Fong, "Pose space deformation: A unified approach to shape interpolation and skeleton-driven deformation," in *SIGGRAPH '00: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, New York,NY,USA, 2000.

[42] T. Kurihara and N. Miyata, "Modeling Deformable Human Hands from Medical Images," in *Proceedings of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, Aire-la-Ville, Switzerland, 2004.

[43] B. Allen, B. Curless and Z. Popovic ´, "Articulated body deformation from range scan data," in *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 2002.

[44] P. G. Kry, D. L. James and D. K. Pai, "EigenSkin: Real time large deformation character skinning in hardware," in *SCA '02: Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*, New York, NY, USA, 2002.

[45] D. Anguelov, Srinivasan, K. D. Praveen, S. Thrun and J. Rodgers, "SCAPE: Shape Completion and Animation of Peopl," *ACM Transactions on Graphics (TOG), Volume 24, Issue 3,* pp. 408-416, 2005.

[46] D. A. Hirshberg, M. Loper, E. Rachlin and M. J. Black, "Coregistration: Simultaneous Alignment and Modeling of Articulated 3D Shape," *Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds) Computer Vision – ECCV 2012. ECCV 2012. Lecture Notes in Computer Science, vol 7577. Springer, Berlin, Heidelberg.,* p. 242–255.

[47] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll and M. J. Black, "SMPL: a skinned multi-person linear model," *ACM Transactions on Graphics (TOG), Volume 34, Issue 6, Article No.: 248,* pp. 1-16, 2015.

[48] M. P. Kadaba, H. K. Ramakrishnan and M. E. Wootten, "Measurement of Lower Extremity Kinematics During Level Walking," *Journal of Orthopaedic Research,* vol. 8383, 1990.

[49] H. D. Cheng, X. H. Jiang, Y. Sun and J. Wang, "Color image segmentation: advances and prospects," *Pattern Recognition,* vol. 34, no. 12, pp. 2259-2281, 2001.

[50] A. Cereatti, T. Bonci, M. Akbarshahi, K. Aminian, A. Barré, M. Begon, D. L. Benoit, C. Charbonnier, F. Dal Maso, S. Fantozzi, C.-C. Lin, T.-W. Lu, M. G. Pandy, R. Stagni, A. J. van den Bogert and V. Camomilla, "Standardization proposal of soft tissue artefact description for data sharing in human motion measurements," *Journal of Biomechanics,* vol. 62, pp. 5-13, 2017.

[51] K. T.K. and L. M.Y., "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research," *Journal of Chiropractic Medicine,* vol. 15, no. 2, pp. 155-163, 2016.

[52] M. Sangeux, E. Passmore, H. K. Graham and O. Tirosh, "The gait standard deviation,," *Gait and Posture,* vol. 46, pp. 194-200, 2016.

[53] J. L. McGinley, R. Baker, R. Wolfe and M. E. Morris, "The reliability of threedimensional kinematic gait measurements: A systematic review.," *Gait and Posture,* vol. 29, no. 3, pp. 360-369, 2009.

[54] Y. Gao, Y. Yang, Y. Zhen and Q. Dai, "Depth error elimination for RGB-D cameras," *ACM Transactions on Intelligent Systems and Technology,* vol. 6, no. 2, 2015.