Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Bioengineering and Medical Surgical Sciences (36th Cycle)

Markerless movement analysis based on RGB and Depth sensor Technology for clinical applications

By

Diletta Balta

Supervisors:

Prof. Filippo Molinari, Supervisor
Prof. Andrea. Cereatti, Co-Supervisor
Prof. Ugo Della Croce, Co-Supervisor

Reviewers:

Prof. Helios De Rosario Martinez, Referee, Instituto de Biomecánica de Valencia
Prof. Morgan Sanguex, Referee, University Children's Hospital Basel

Politecnico di Torino
2024

Declaration

I hereby declare that the contents and organization of this dissertation constitute my own original work and do not compromise in any way the rights of third parties, including those relating to the security of personal data.

Diletta Balta

2024

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Abstract

Instrumental movement analysis is a powerful tool that offers valuable insights into human movement. It is applicable in numerous fields, including screening, treatment planning, and predicting neurological disorders. Traditional marker-based systems for movement analysis, although precise, are often costly and require extensive setup and subject preparation. Advancements in computer vision technique has led to the development of more accessible and user-friendly markerless systems. There are some applications such as screening and treatment planning which would benefit from portable, affordable, and easy-to-use systems, ideally with single-camera setups. Inexpensive tracking systems combining RGB with infrared depth sensors (RGB-D) have emerged, enabling depth color image generation without the need for multiple cameras. Human motion estimation algorithms from a single camera can be categorized into deterministic and AI-based approaches. Deterministic approaches rely on formulas, clear anatomical rules and mathematical principles for defining joint centers and could require a predefined kinematic model (model-based approach) or could extract features directly from video data using human anatomical proportions (model-free approach). AI-based methods rely on data-driven motion characteristics enhanced by deep learning thus taking advantage of large datasets and could also include predetermined model to refine the estimates obtained from convolutional neural networks.

The first part of this thesis focuses on gait analysis which is a useful tool for follow-up and screening purposes. In this context, many of single camera methods, belonging to both categories, lack validation against clinical standards, particularly for pathological populations, or, if validated, only a single joint was tested, or their clinical applicability is limited by uniform backgrounds and color filters. This thesis aimed to fill these gaps in clinical gait analysis by proposing and validating against a marker-based system (*1*) original deterministic markerless protocols based on a single RGB-D camera in patients with cerebral palsy and foot deformities and (*2*) by exploring the clinical validity of AI-based algorithms on healthy subjects. Regarding deterministic approaches, a 2D model-based protocol was proposed (*1a*) and validated on 18 patients with CP. Accuracy and reliability of spatial-temporal parameters and sagittal lower limb joint kinematics were assessed by comparing them to a 3D marker-based system in terms of offset and waveform similarity. The main findings revealed that stride duration had the lowest mean absolute percentage error at 2%, followed by step length at 2.2%, stride length at 2.5%, and gait speed at 3.1%. The angular offsets were 8 deg for the ankle, 6 deg for the knee, and 7 deg for the hip joint. Additionally, the root mean square error values were 3.2 deg for the knee, 3.5 deg for the hip, and 4.5 deg for the ankle. Despite good accuracy, this method requires the calibration of the 2D kinematic model in static, loading, and swing phases to partially compensate for movements outside the sagittal plane and changes in position between the subject and the camera, including also the manual identification of specific anatomical landmarks on the three images. To address these limitations, a 3D version was developed (*1b*) benefiting from a 3D statistical lower-limb model and applied to patients with CP and foot deformities. The innovative aspect of this work lies in the reconstruction of a 3D subject-specific model from three static recordings still maintaining a portable system with an RGB-D camera, unlike traditional methods which create 3D models from multiple cameras or 3D scanners unsuitable for ambulatory settings. The accuracy and reliability of sagittal lower limb joint kinematics (hip, knee and ankle joints) were evaluated against a 3D marker-based system in terms of mean absolute error on specific gait features derived from sagittal angles. The 3D method demonstrates comparable performance in terms of mean absolute error to the 2D protocol for gait features related to the hip (4.2 deg vs 3.7 deg), knee (4.0 deg vs 4.3 deg) and ankle (3.8 deg vs 3.5 deg). Moreover, this protocol demonstrated a good reliability (ICC>0.75) for every gait feature which is comparable to the marker-based system. The main finding is that this protocol is fully automatic and effectively compensates for movements outside the sagittal plane without requiring multiple 2D models and manual identification of various anatomical points during calibration making the 3D protocol more robust and efficient. Additionally, having a 3D model, volumetric parameters to evaluate asymmetries can be extracted, going beyond traditional gait analysis. Finally, the study of sagittal ankle and metatarsophalangeal kinematics using a single camera was also explored (*1c*) since the majority of markerless methods allow modeling the foot as a single segment without articulating the metatarsophalangeal joint, which is crucial for effective foot loading and correct progression. For this reason, a 3D markerless model-based method was designed using a two-segment foot model composed by mid-rear and forefoot foot connected by metatarsophalangeal joint. Validation against manually labeled measures on 10 children with foot deformities showed averaged root mean square errors of 5 deg for the metatarsophalangeal joint and 4.8 deg for the ankle. The second aim (*2*) focused on the investigation of the clinical applicability of the AI-based Azure Kinect body tracking software development kit (SDK) evaluated against the marker-based system on five healthy subjects during straight walking. The accuracy and reliability of sagittal lower limb joint kinematics (hip, knee and ankle joints) were evaluated against a 3D marker-based system in terms of mean absolute error on specific gait features derived from lower-limb sagittal angles. Results indicated Azure Kinect body tracking SDK can introduce errors of about 8 deg for the hip, 2 deg for the knee and 33 deg for the ankle demonstrating that its main limitation is in ankle angle computation, which was estimated using the inclination of the segment from the ankle joint to the toe which is not representative of the actual foot inclination. In addition, from a visual inspection, it was noticed that when legs overlap during the gait cycle Azure Kinect body tracking SDK suffers from an unpredictable left-right confusion issue making this method unsuitable for clinical gait analysis.

The second part of the thesis regards upper-limb movement analysis for early detection of movement disorders in preterm infants. General Movement Assessment, proposed by Heinz Prechtl, is the gold standard but requires extensive training and time being based on visual assessment. 3D marker-based analysis could be accurate but interferes with infants' natural movements. Thus, many studies focusing on 2D markerless video analysis have been proposed. However, 3D analysis using a single RGB-D camera could offer more accurate insights due to the 3D nature of movement. The purpose of this thesis is to propose a novel markerless protocol for infants' upper body movements analysis based on a single RGB-D camera that features a simplified instrumental setup, suitable for home use, to a purposely developed algorithm for 3D pose estimation and general movements metrics extraction. Open-source methods, such as DeepLabCut, have proven useful for being adapted to this challenging scenario. RGB videos were processed using DeepLabCut, which was previously trained, to estimate 2D pixel locations of left and right shoulders, elbows and wrists. A specifically designed method enabled the reconstruction of their 3D trajectories using data recorded with a depth sensor, handling body occlusions and accidental movements of the seat or camera. Proper training set construction was proposed to reduce computational time for manually labeling points of interest, including biomechanical domain knowledge. This approach allows the extraction of metrics from 3D point of interest trajectories capable of describing infants' spontaneous movements. This method was tested on eight infants aged 3 to 5 months and on a pair of twins with divergent health profiles. The main findings indicated that general movement metrics could be effectively calculated and may serve as valuable tools for the early detection of movement disorders, although with some limitations due to environmental factors in uncontrolled home scenarios.

In conclusion, this thesis demonstrates the increasing viability of markerless approaches in clinical movement analysis due to advancements in technology, computer vision, and machine learning. It is worth to notice that the clinical applicability of the gait analysis markerless methods developed in this thesis is currently being tested both at Skaraborg Hospital in Skövde (Sweden) and in the outpatient clinics of ASL TO5 in Turin (Italy). The ultimate goal is to lay the foundation for establishing an Italian Cerebral Palsy Follow-up Program Registry with the same characteristics of the Sweden's one and finally introduce the markerless gait analysis into routine clinical practice.

Contents

List of acronyms

Artificial Intelligence	AI
Articulated Iterative Closest Point	AICP
Ankle joint center	AJC
At-Risk child	AR
Anterior Supine Iliac Spine	ASIS
Belly Button	B
Convolutional Neural Network	CNN
Cerebral Palsy	CP
Cerebral Palsy Follow-Up Program	CPUP
DeepLabCut	DLC
Dynamic Time Warping	DTW
Degree of Freedom	DoF
Elbow Angle	EA
Forearm	FA
Fore-foot	FF
Foot Coordinate System	f
Gait variability standard deviation	GVSD

General Movement	GM
General Movement Assessment	GMA
Gross Motor Function Classification System	GMFCS
Great trochanter	GT
Hip joint center	HJC
Image Coordinate System	*I*
Initial Contact	IC
Intraclass Correlation Coefficient	ICC
Iterative Closest Point	ICP
Infra-Red	IR
Knee joint center	KJC
Left (Right) Elbow	L(R) E
Left (Right) Shoulder	L(R) S
Left (Right) Wrist	L(R) W
Lateral epicondyle	LE
Lateral malleolus	LM
Mean Absolute Difference	MAD

Mean Absolute Error	MAE
Mean Absolute Percentage Error	MAPE
Marker-based	MB
Mean Difference	MD
Mean Error	ME
Model-based method	MLM
Mid-Rear-Foot	MRF
Marker-less	MS
Metatarso-phalangeal joint	MTP
Fifth Metatarso-phalangeal joint	MTP5
Posterior Supine Iliac Spine	PSIS
Point of Interest	PoI
Red-Green-Blu	RGB
Red-Green-Blu - Depth	RGB-D
Root Mean Square Error	RMSE
Range of Motion	RoM
Shoulder Angle	SA
Software Development Kit	SDK

Skinned Multi-Person Linear model	SMPL
Standard Deviation	STD
Typically Developed child	TD
Template	TMP
Time of Flight	ToF
Upper arm	UA
Shank Coordinate System	*s*
Thigh Coordinate System	*t*

Chapter 1

1. Clinical relevance and general introduction

1.1 The importance of the human motion monitoring

Instrumental movement analysis is a comprehensive tool that provides valuable insights into human movement, applicable in various fields such as screening, treatment planning, and predicting neurological disorders. Currently, optical stereophotogrammetry is the gold standard for instrumented movement analysis due to its submillimeter accuracy in tracking the position of markers attached to the subject’s skin and its high temporal resolution, down to milliseconds. However, despite its accuracy, this technology is very costly, requires expert operators, specialized laboratories, and lengthy patient preparation times. Thus, making it unsuitable for ambulatory setting and for conducting analyses directly at the patient's home. Recently, video-based markerless systems have emerged as a promising alternative to marker-based systems due to their cost-effectiveness, easy setup, and the elimination of the need for markers on the skin of the subject. At present, the precision and validity of markerless systems for biomechanical and clinical applications remain an open question (Lam et al., 2023; Wade et al., 2023), limiting their use in clinical settings. However, when portability and ease of use is a priority (e.g. screening, identifying gait patterns, monitoring progress over time, and evaluating treatment), methods based on a single camera with minimal setup time should be preferred (Harvey & Gorter, 2011). Recently, various manufacturers have introduced affordable tracking systems priced between 200-400 €/$, which incorporate an RGB camera with an infrared depth sensor (RGB-D). These systems merge RGB image data with depth information to create enhanced depth color images (2D+), eliminating the need for multiple cameras.

Several algorithms have been developed for estimating human motion from a single video data, using techniques that extract features from 2D images. These techniques are typically divided into two categories: deterministic and AI-based approaches.

Deterministic methods could be categorized into model-based and model-free approaches (Mündermann et al., 2006). Model-based ones use an *a-priori* model of the subject, such as stick figures, cylinders, or CAD models, to track or match against the video data. Among kinematic models, both 2D and 3D models were proposed. 2D models are derived from 2D images or video recordings and are capable of analyzing movement within a single plane. On the other hand, 3D models are generally reconstructed using multiple cameras or 3D scanners, allowing for a detailed reconstruction of movements in all three dimensions. Model-free approaches do not rely on pre-existing models but rather identify motion characteristics directly from the data using pre-defined human anatomical proportions.

AI-based approaches, powered by deep learning, excel at automatically learning from large datasets, reconstructing and interpreting complex motion patterns. They are typically model-free, allowing them to estimate joint positions directly from data without needing a predefined model. However, in the literature, AI model-based algorithms have been introduced to fine-tune joint centers estimates through model fitting algorithms (Romeo et al., 2021), ensuring more precise and reliable results.

Several studies, belonging to both categories, have been proposed regarding both markerless gait analysis and upper-limb movement analysis.

Clinical gait analysis is essential for understanding and interpreting the physio-pathological characteristics of human locomotion, and its diagnostic value is well-established. In this context, many of the proposed methods based on a single camera, belonging to both the categories, have not been validated against clinically accepted standards on pathological populations (Amprimo et al., 2021; Balta et al., 2020; Castelli et al., 2015; Ferraris, Amprimo, Masi, et al., 2022; Latorre et al., 2018, 2019; Leu et al., 2011; Goffredo Michela and Carter, 2009) or the validation has been conducted only on a single joint (Leu et al., 2011; Pantzar-Castilla et al., 2018; Surer et al., 2011) creating uncertainties about their accuracy in measuring joint movements for clinical applications. Moreover, some markerless approaches often focus only on the validation of the joint centers' positions (Hesse et al., 2023) or prioritize classifying motor activities and detecting gait abnormalities (Chen et al., 2011; Clark et al., 2015; Ferraris, Amprimo, Pettiti, et al., 2022; Kojovic et al., 2021; Li et al., 2018; Stricker et al., 2021) rather than investigate the validation of kinematic and spatial-temporal parameters. The practical use of some markerless studies in clinical settings is also constrained by methodological limitations such as the dependence on color filters and uniform backgrounds (Castelli et al., 2015; Pantzar-Castilla et al., 2018) for facilitating subject segmentation algorithms at the expense of the simplicity of the experimental setup. Moreover, there is a general lack of rigorous technical validation of these methods in populations with gait impairments, which is crucial for their potential application in clinical diagnostics.

Upper-limb movement analysis is particularly useful for the early detection of movement disorders in preterm infants. The gold standard for the early identification of motor disorders is the General Movement Assessment, which requires a visual and qualitative video analysis conducted by a clinician, as well as extensive training and long execution time. 3D marker-based analysis is highly accurate but not particularly suitable because the markers attached to the infant's skin could interfere with their natural movements. For this reason, numerous studies have focused on 2D markerless video analysis (Adde et al., 2010; Ihlen et al., 2020; Moro et al., 2022; Stagni et al., 2023). However, a 3D markerless analysis using a single RGB-D camera could be more useful and accurate, considering the inherently 3D nature of movement.

To summarize, the research questions addressed in this thesis are as follows:

1 Development of a 2D markerless algorithm based on a single RGB-D camera and validation against clinically accepted standards, particularly on pathological populations.

2 Development of an automatic segmentation algorithm to avoid dependence on color filters and uniform backgrounds in MS gait analysis systems, thereby simplifying experimental setups without compromising accuracy.

3 Going beyond traditional MS gait analysis techniques, that normally consider the foot as a rigid segment, by including a 3D multi-segments foot model

4 Assessment of the clinical applicability of Azure Kinect's Body Tracking SDK for clinical gait analysis by comparing it with stereophotogrammetric systems.

5 Moving beyond 2D data analysis by including a 3D subject-specific lower-limb model, leveraging 3D data for more comprehensive and clinically relevant assessments.

6 Development of a MS protocol to study general movements in preterm infants, utilizing 3D data from a single RGB-D camera directly at the patient's home, unlike existing studies in the literature that conduct a 2D analysis exclusively in clinical settings

7 How effective are the metrics proposed in the literature for studying general movements in distinguishing infants with different health profiles?

1.2 Aim of the thesis

This thesis aims to address the abovementioned gaps in clinical gait analysis using a single RGB-D camera by proposing innovative deterministic markerless protocols specifically designed for patients with CP and children with foot deformities and validated against the marker-based system. Moreover, the clinical applicability and validity of AI-based markerless algorithms for gait analysis was investigated by comparing them against traditional systems.

In addition, this thesis aims to fill the gaps in the general movement analysis on preterm infants using an RGB-D camera by (*i*) developing a AI-based method specifically designed to reconstruct the 3D coordinates of upper limb joint centers directly at home and (*ii*) by exploring its clinical applicability in extracting general movements metrics as reliable indicators of movement disorders.

In particular, the manuscript organization is detailed below:

1. In the second chapter, deterministic model-based approaches are presented. Initially, a 2D model-based markerless protocol for clinical gait analysis

using a single RGB-D camera was proposed and validated on 18 patients with cerebral palsy. This innovative protocol employed a 2D kinematic lower limb model to estimate sagittal lower-limb joint kinematics using the Iterative Closest Point algorithm. The proposed method requires the calibration of the 2D kinematic model in static, loading, and swing phases to partially compensate for movements outside the sagittal plane and changes in position between the subject and the camera including also the manual identification of specific anatomical landmarks on the three images. Despite this precaution, the main limitation is that the projection of human 3D body motion to a 2D space necessarily leads to errors and ambiguities. To overcome these limitations, the chapter's second section introduces an extended 3D version of the aforementioned 2D markerless protocol. This approach benefited from a generic statistical 3D Skinned Multi-Person Linear model, which includes the foot, shank, thigh, and pelvis, interconnected by joints at the ankle, knee, and hip. The innovation of this work lies in the reconstruction of a 3D subject-specific model from three static recordings still maintaining a portable system with an RGB-D camera, unlike traditional methods that create 3D models from multiple cameras or 3D scanners. Lower limb joint angles were estimated by matching the 3D model to each 3D point cloud of the gait cycle through the articulated iterative closest point, unlike the previous 2D protocol where the kinematic model was fitted to 2D RGB images. This method was validated on 10 patients (6 individuals with cerebral palsy, and 4 children with foot deformities).

Finally, a 3D markerless method for estimating sagittal foot kinematic was designed by using a two-segment 3D foot model composed by mid-rear and forefoot foot connected by metatarsophalangeal joint. Its clinical applicability on children with foot deformities was explored.

2. The third chapter explores the clinical applicability of AI-based markerless methods in both clinical gait analysis and upper-limb movement analysis. Initially, the performance of the Azure Kinect body tracking software development kit was assessed and compared to the 2D deterministic model-based approach, previously proposed, on five healthy participants during straight walking. Subsequently, the majority of this chapter is devoted to proposing a markerless protocol based on a single RGB-Depth camera for the study of the general movements on preterm infants within a home setting. This method involves using an open-source deep-learning algorithm, DeepLabCut (Mathis et al., 2018), and a purposely algorithm for reconstructing 3D

coordinates of each joint center by using depth images. This method was preliminary validated on a physical model and on real babies against manual measurements. Proper metrics, selected from the literature (Meinecke et al., 2006), were computed for quantifying general movements and were tested on 8 infants from 3 to 5 months. Then, a refined version of the aforementioned method was introduced for compensating for accidental movements of the camera or seat that could occur in a home environment. Additional metrics from the literature (Kanemaru et al., 2013; Karch et al., 2012) were investigated to appreciate their potential as reliable indicators of developmental abnormalities on a pair of twins, one of them typically developed and the other at risk to develop cerebral palsy, proving a unique case scenario.

1.3 Working principles of RGB-D technology

This paragraph outlines the working principles of the main RGB-D cameras available on the market, exploring the primary methods for reconstructing depth images and 3D point clouds.

1.3.1 Depth image reconstruction

The evolution of Microsoft's Kinect sensor technology, from its inception with Kinect v1 for the Xbox 360 in 2010, through to Kinect v2 for the Xbox One in 2013, and culminating with the Azure Kinect announced in 2019, illustrates a remarkable journey of technological advancement and expanding application horizons. This progression from gaming-centric devices to tools with broad commercial and research applications mirrors the advancements in underlying depth sensing technologies: from Fixed Structured Light to Time of Flight (ToF), each leap forward brought about significant improvements in accuracy, versatility, and application scope.

1.3.1.1 Fixed Structured Light

Kinect v1 introduced the possibility of controller-free gaming and basic gesture recognition, utilizing Structured Light technology. This approach involved projecting a pattern of infrared dots into a room and analyzing the distortions in this pattern caused by objects and people, to map out the depth of the scene as shown in **Figure 1**. Despite its innovative nature, Kinect v1 faced limitations, particularly with accuracy and depth resolution over distances and in challenging lighting conditions, such as direct sunlight.

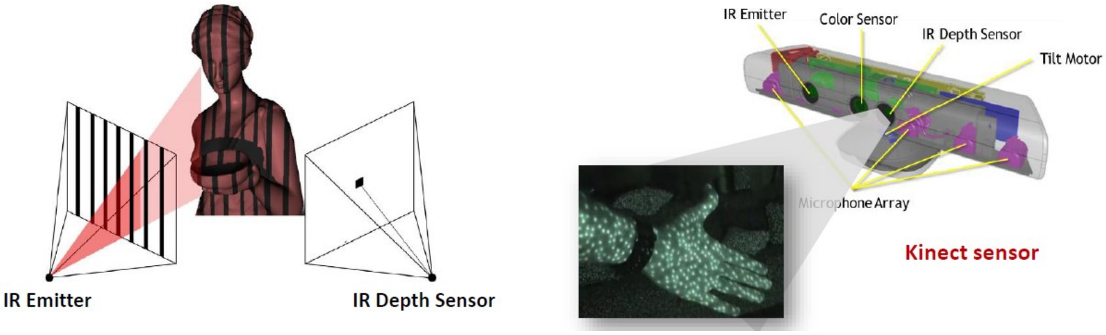


Figure 1. Illustration of fixed structed light technique. Left: IR emitter and IR depth sensor mechanism. Right: Internal components of a Kinect sensor, showing how it captures RGB and depth data for 3D point cloud reconstruction.

1.3.1.2 Time of Flight (ToF)

In contrast, Kinect v2 marked a significant leap in depth sensing technology by adopting ToF technology. ToF sensors work by emitting infrared light pulses and measuring the time it takes for these pulses to be reflected from objects. This method offers more detailed and accurate depth information, enabling Kinect v2 to track up to 25 joints per person for up to six people. It showcased improved skeletal tracking, a higher fidelity RGB camera, and a wider field of view, making it more adept at detecting user positions and movements.

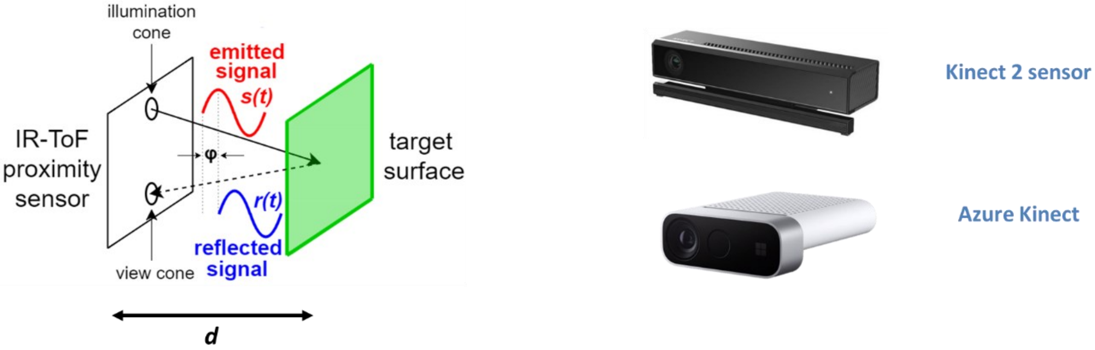


Figure 2. ToF Sensor Mechanism and Kinect Devices. Left: IR-ToF sensor mechanism with emitted and reflected signals. Right: Kinect 2 and Azure Kinect devices using ToF technology for depth sensing.

The advent of Azure Kinect further expanded the versatility and application of Kinect technology. By using a next-generation ToF sensor, Azure Kinect provides high-quality depth data with improved spatial resolution and accuracy, suitable for a range of environments and lighting conditions. Unlike its predecessors, Azure Kinect targets not just gaming but also sectors like healthcare, retail, and industrial applications, equipped with a 7-microphone array, a 12 Mega Pixel RGB camera, and a 1 Mega Pixel depth camera with a wide field of view. This technological evolution from Structured Light to ToF underlines a shift towards greater robustness and versatility. ToF technology is favored for its performance across various environmental conditions, simplicity, speed, and ability to operate effectively outdoors. Unlike Structured Light, which can be hampered by ambient light and requires complex computation to analyze pattern deformations, ToF's straightforward mechanism of measuring light reflection times enables rapid and accurate depth capture, even in real-time scenarios. Moreover, the range and accuracy of ToF sensors surpass those of Structured Light systems, offering consistent performance over more extended distances and in diverse conditions.

1.3.1.3 Stereovision technique

Stereovision, also known as stereo vision or stereo imaging, is a technique used in computer vision to reconstruct a depth image of a scene from images taken from two slightly different viewpoints, mimicking human binocular vision. This modality was principally implemented in the Intel RealSense cameras **Figure 3**.

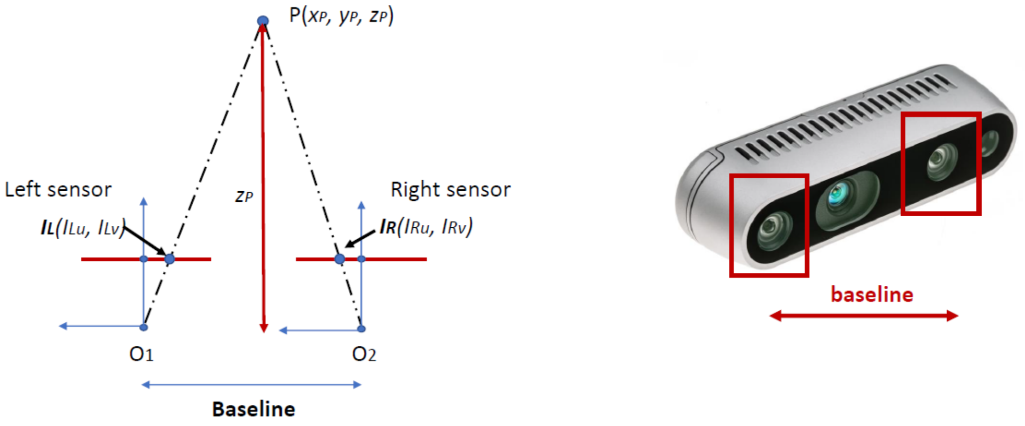


Figure 3. Stereo-vision Mechanism in Intel-RealSense cameras. Left: Stereo-vision Mechanism with left and right sensor to reconstruct a depth image through triangulation algorithm. Right: Intel RealSense D435 using stereo-vision technology for depth sensing.

Stereovision algorithms reconstruct depth information from a scene by utilizing two cameras placed at a distance like the spacing between human eyes. These cameras capture the same scene from different perspectives, creating slight disparities in the images obtained. First, the cameras are calibrated to establish their internal parameters such as focal length and optical centers, as well as their relative positions. This calibration corrects lens distortions and aligns the images in a process called rectification, making it easier to compare them as corresponding points line up horizontally.

Next, the algorithm employs either feature matching, which identifies unique features in each image to find matches, or block matching, which compares larger blocks of pixels across images. The differences in position of these matched features or blocks, known as disparities, are crucial for estimating depth; greater disparities suggest nearer objects, while smaller ones indicate objects are further away.

Finally, using the disparities and known camera parameters, the algorithm calculates the depth for each point by triangulation, which involves using the baseline distance between the cameras and their focal lengths. This calculation produces a depth map, providing a three-dimensional representation of the scene, essential for applications such as 3D modeling, autonomous driving, and robotics.

Following, a summery comprehensive of all the features of each camera was reported in **Table 1**.

Table 1. Comparison of Depth-Sensing Cameras.

Specification	*Kinect v1*	*Kinect v2*	*Azure Kinect*	*Intel RealSense D435*
Launch Year	2010	2013	2019	2018
Depth Technology	Structured Light	Time of Flight (ToF)	Time of Flight (ToF)	Stereo Vision
Field of View	Horizontal: 57°, Vertical: 43°	Horizontal: 70°, Vertical: 60°	Horizontal: 75°, Vertical: 65°	Horizontal: 85°, Vertical: 58°
Depth Resolution	640x480	512x424	Up to 640x576	Up to 1280x720
RGB Resolution	640x480	1920x1080	3840x2160 (4K)	1920x1080
Max Frame Rate	30 fps	30 fps	30 fps	90 fps (depth), 60 fps (color)
Operating Range	0.8 to 4 m	0.5 to 4.5 m	0.5 to 5.46 m	0.2 to 10 m
Interfaces	USB 2.0	USB 3.0	USB-C 3.1	USB-C 3.0
Applications	Gaming, Basic Motion Capture	Improved Motion Capture, Gaming	Advanced Motion Capture, Research, Healthcare, Retail	3D Scanning, Robotics, VR/AR

1.3.2 3D point cloud reconstruction

RGB-D cameras provide the possibility to reconstruct the 3D point clouds of the objects inside the field of view of the camera. The process of generating a point cloud begins with the camera calibration which provides intrinsic camera parameters such as the focal length and the position of the principal point. Using these parameters, a point cloud was generated as follows:

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

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

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

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.

References

Adde, L., Helbostad, J. L., Jensenius, A. R., Taraldsen, G., Grunewaldt, K. H., & StØen, R. (2010). Early prediction of cerebral palsy by computer-based video analysis of general movements: A feasibility study. *Developmental Medicine and Child Neurology*, *52*(8), 773–778. https://doi.org/10.1111/j.1469-8749.2010.03629.x

Amprimo, G., Pettiti, G., Priano, L., Mauro, A., & Ferraris, C. (2021). *Kinect-based solution for the home monitoring of gait and balance in elderly people with and without neurological diseases*.

Balta, D., Salvi, M., Molinari, F., Figari, G., Paolini, G., Croce, U. Della, & Cereatti, A. (2020). A two-dimensional clinical gait analysis protocol based on markerless recordings from a single RGB-Depth camera. *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 1–6. https://doi.org/10.1109/MeMeA49120.2020.9137183

Castelli, A., Paolini, G., Cereatti, A., & Della Croce, U. (2015). A 2D markerless gait analysis methodology: Validation on healthy subjects. *Computational and Mathematical Methods in Medicine*, *2015*. https://doi.org/10.1155/2015/186780

Chen, S. W., Lin, S. H., Liao, L. De, Lai, H. Y., Pei, Y. C., Kuo, T. S., Lin, C. T., Chang, J. Y., Chen, Y. Y., Lo, Y. C., Chen, S. Y., Wu, R., & Tsang, S. (2011). Quantification and recognition of parkinsonian gait from monocular video imaging using kernel-based principal component analysis. *BioMedical Engineering Online*, *10*. https://doi.org/10.1186/1475-925X-10-99

Clark, R. A., Vernon, S., Mentiplay, B. F., Miller, K. J., McGinley, J. L., Pua, Y. H., Paterson, K., & Bower, K. J. (2015). Instrumenting gait assessment using the Kinect in people living with stroke: reliability and association with balance tests. *Journal of Neuroengineering and Rehabilitation*, *12*, 15. https://doi.org/10.1186/s12984-015-0006-8

Ferraris, C., Amprimo, G., Masi, G., Vismara, L., Cremascoli, R., Sinagra, S., Pettiti, G., Mauro, A., & Priano, L. (2022). Evaluation of Arm Swing Features and Asymmetry during Gait in Parkinson's Disease Using the Azure Kinect Sensor. *Sensors*, *22*(16). https://doi.org/10.3390/s22166282

Ferraris, C., Amprimo, G., Pettiti, G., Masi, G., & Priano, L. (2022). Automatic Detector of Gait Alterations using RGB-D sensor and supervised classifiers: a preliminary study. *Proceedings - IEEE Symposium on Computers and Communications*, *2022-June*. https://doi.org/10.1109/ISCC55528.2022.9912923

Goffredo Michela and Carter, J. N. and N. M. S. (2009). 2D Markerless Gait Analysis. In P. and N. M. and H. J. Vander Sloten Jos and Verdonck (Ed.), *4th European Conference of the International Federation for Medical and Biological Engineering* (pp. 67–71). Springer Berlin Heidelberg.

Harvey, A., & Gorter, J. W. (2011). Video gait analysis for ambulatory children with cerebral palsy: Why, when, where and how! *Gait and Posture*, *33*(3), 501–503. https://doi.org/10.1016/j.gaitpost.2010.11.025

Hesse, N., Baumgartner, S., Gut, A., & Van Hedel, H. J. A. (2023). Concurrent Validity of a Custom Method for Markerless 3D Full-Body Motion Tracking of Children and Young Adults based on a Single RGB-D Camera. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. https://doi.org/10.1109/TNSRE.2023.3251440

Ihlen, E. A. F., Støen, R., Boswell, L., de Regnier, R. A., Fjørtoft, T., Gaebler-Spira, D., Labori, C., Loennecken, M. C., Msall, M. E., Möinichen, U. I., Peyton, C., Schreiber, M. D., Silberg, I. E., Songstad, N. T., Vågen, R. T., Øberg, G. K., & Adde, L. (2020). Machine learning of infant spontaneous movements for the early prediction of cerebral palsy: A multi-site cohort study. *Journal of Clinical Medicine*, *9*(1). https://doi.org/10.3390/jcm9010005

Kanemaru, N., Watanabe, H., Kihara, H., Nakano, H., Takaya, R., Nakamura, T., Nakano, J., Taga, G., & Konishi, Y. (2013). Specific characteristics of spontaneous movements in preterm infants at term age are associated with developmental delays at age 3 years. *Developmental Medicine and Child Neurology*, *55*(8), 713–721. https://doi.org/10.1111/dmcn.12156

Karch, D., Kang, K. S., Wochner, K., Philippi, H., Hadders-Algra, M., Pietz, J., & Dickhaus, H. (2012). Kinematic assessment of stereotypy in spontaneous movements in infants. *Gait and Posture*, *36*(2), 307–311. https://doi.org/10.1016/j.gaitpost.2012.03.017

Kojovic, N., Natraj, S., Mohanty, S. P., Maillart, T., & Schaer, M. (2021). Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-94378-z

Lam, W. W. T., Tang, Y. M., & Fong, K. N. K. (2023). A systematic review of the applications of markerless motion capture (MMC) technology for clinical measurement in rehabilitation. In *Journal of NeuroEngineering and Rehabilitation* (Vol. 20, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12984-023-01186-9

Latorre, J., Colomer, C., Alcañiz, M., & Llorens, R. (2019). Gait analysis with the Kinect v2: Normative study with healthy individuals and comprehensive study of its sensitivity, validity, and reliability in individuals with stroke. *Journal of NeuroEngineering and Rehabilitation*, *16*(1). https://doi.org/10.1186/s12984-019-0568-y

Latorre, J., Llorens, R., Colomer, C., & Alcañiz, M. (2018). Reliability and comparison of Kinect-based methods for estimating spatiotemporal gait parameters of healthy and post-stroke individuals. *Journal of Biomechanics*, *72*, 268–273. https://doi.org/10.1016/j.jbiomech.2018.03.008

Leu, A., Ristic-Durrant, D., & Graser, A. (2011). A robust markerless vision-based human gait analysis system. *SACI 2011 - 6th IEEE International Symposium on Applied Computational Intelligence and Informatics, Proceedings*, 415–420. https://doi.org/10.1109/SACI.2011.5873039

Li, T., Chen, J., Hu, C., Ma, Y., Wu, Z., Wan, W., Huang, Y., Jia, F., Gong, C., Wan, S., & Li, L. (2018). Automatic timed up-and-go sub-task segmentation for Parkinson's disease patients using video-based activity classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(11), 2189–2199. https://doi.org/10.1109/TNSRE.2018.2875738

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

Meinecke, L., Breitbach-Faller, N., Bartz, C., Damen, R., Rau, G., & Disselhorst-Klug, C. (2006). Movement analysis in the early detection of newborns at risk for developing spasticity due to infantile cerebral palsy. *Human Movement Science*, *25*(2), 125–144. https://doi.org/10.1016/j.humov.2005.09.012

Moro, M., Pastore, V. P., Tacchino, C., Durand, P., Blanchi, I., Moretti, P., Odone, F., & Casadio, M. (2022). A markerless pipeline to analyze spontaneous movements of preterm infants. *Computer Methods and Programs in Biomedicine*, *226*. https://doi.org/10.1016/j.cmpb.2022.107119

Mündermann, L., Corazza, S., & Andriacchi, T. P. (2006). The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. In *Journal of NeuroEngineering and Rehabilitation* (Vol. 3). https://doi.org/10.1186/1743-0003-3-6

Pantzar-Castilla, E., Cereatti, A., Figari, G., Valeri, N., Paolini, G., Della Croce, U., Magnuson, A., & Riad, J. (2018). Knee joint sagittal plane movement in cerebral palsy: a comparative study of 2-dimensional markerless video and 3-dimensional gait analysis. *Acta Orthopaedica*, *89*(6), 656–661. https://doi.org/10.1080/17453674.2018.1525195

Romeo, L., Marani, R., Malosio, M., Perri, A. G., & D'Orazio, T. (2021). Performance analysis of body tracking with the microsoft azure kinect. *2021 29th Mediterranean Conference on Control and Automation, MED 2021*, 572–577. https://doi.org/10.1109/MED51440.2021.9480177

Stagni, R., Doto, T., Tomadin, A., Sansavini, A., Aceti, A., Corvaglia, L. T., & Bisi, M. C. (2023). General movements automatic assessment: Methodological issues for pose estimation. *Gait & Posture*, *106*, S195–S196. https://doi.org/https://doi.org/10.1016/j.gaitpost.2023.07.236

Stricker, M., Hinde, D., Rolland, A., Salzman, N., Watson, A., & Almonroeder, T. G. (2021). Quantifying step length using two-dimensional video in individuals with Parkinson's disease. *Physiotherapy Theory and Practice*, *37*(1), 252–255. https://doi.org/10.1080/09593985.2019.1594472

Surer, E., Cereatti, A., Grosso, E., & Croce, U. Della. (2011). A markerless estimation of the ankle-foot complex 2D kinematics during stance. *Gait and Posture*, *33*(4), 532–537. https://doi.org/10.1016/j.gaitpost.2011.01.003

Wade, L., Needham, L., Evans, M., McGuigan, P., Colyer, S., Cosker, D., & Bilzon, J. (2023). Examination of 2D frontal and sagittal markerless motion capture: Implications for markerless applications. *PLoS ONE*, *18*(11 NOVEMBER). https://doi.org/10.1371/journal.pone.0293917

Chapter 2

2. Deterministic approaches for clinical gait analysis

2.1 Introduction

The methods and the results presented in this chapter have been published in (Balta et al., 2020, 2023)

In this chapter, the first category of methods, those associated with the deterministic approach, will be detailed. The term "deterministic approach" refers to the presence of deterministic formulas, clear anatomical rules and mathematical principles for defining joint centers. Deterministic methods ensure consistent and reliable results due to their foundation in well-defined anatomical and mathematical principles. This replicability is crucial for clinical applications, as it allows for standardized assessments across different patients and time points. Moreover, the deterministic nature of these methods means that they can be applied with confidence, knowing that the outcomes will be consistent and based on established anatomical relationships. Deterministic methods maintain their efficacy without requiring extensive data unlike AI-based approaches, which typically improve with larger datasets through machine and deep learning techniques. This characteristic is particularly advantageous in clinical settings where obtaining large amounts of data can be challenging, such as when dealing with rare conditions or when the pathological population is limited.

In deterministic approaches applied to human motion analysis, we can primarily distinguish between two categories (Mündermann et al., 2006): those based on fitting a predefined model on video data (model-based) and those based on extracting gait features directly from the human silhouette based on anatomical proportions (model-free).

The term "model-based approach" refers to a methodological framework in which a predefined theoretical model is used to guide the analysis and interpretation of data. This model can be mathematical, statistical, or computational and serves as a representation of the real-world phenomena being studied. The key advantage of this approach is that it allows for a deep understanding of the dynamics and relationships within the data based on the theory or prior knowledge encapsulated in the model.

In the context of markerless (MS) gait analysis, a model-based approach involves using a predefined biomechanical model of the human body to reconstruct the motion captured by cameras without the use of physical markers on the body. This model typically represents the body as a series of linked segments and joints which are mathematically described to mimic the mechanical properties and movements of the human body. The model contains information about the number of joints and their degrees of freedom, as well as constraints on the movement of each body part to achieve smooth motion while respecting anatomical and physiological constraints. These methods are designed to be replicable and exclude machine learning approaches. Their benefit lies in not requiring a specialized training set, although they must be tailored to address specific issues.

Generally, model-based approaches use a predefined anatomical model of the human body, which may consist of stick figures, solid surfaces, or other geometric representations of body parts. The fitting process involves adjusting and optimizing the position, orientation, and sometimes the shape of these models, to match as closely as possible to the captured motion data. This adjustment is driven by direct measurements, such as those obtained from motion capture systems, or visual data like images or videos. The identification of joint centers and the positions of various model segments is determined through sophisticated algorithms that minimize the error between the model and the observed data following specific steps:

a) Model Construction: A biomechanical model of the human body is constructed using segments that represent bones and joints. This model includes parameters such as the lengths of limbs, joint constraints, and possible ranges of motion, which are crucial for accurate movement simulation.

b) Data Capture: High-resolution cameras or depth sensors capture the movement of a person walking.

c) Data Processing and Model Fitting: The model-based approach then involves fitting the biomechanical model to the observed data.

d) Motion Analysis: Once the model is fitted to the data, it can be used to calculate various biomechanical parameters such as joint angles, velocities, and accelerations. These calculations are based on the movements of the model segments as they are mapped onto the observed data.

On the other hand, model-free approaches attempt to capture skeleton features in the absence of an a priori kinematic model but having the information related to human anatomical proportions. Generally, those methods use silhouettes extracted from images. Unlike model fitting algorithms, where a pre-existing 2D or 3D model is modified, silhouette-based approaches use information directly extracted from silhouettes to determine joint center positions. Anatomical proportions of the human body are used to infer the positions of joints within the silhouette. For example, the proportion between the total length of the arm and that from the elbow to the wrist can guide the identification of the elbow joint within the silhouette. This method is particularly useful when there is a need to obtain a quick and relatively accurate representation of kinematics without using complex 3D models.

Both approaches offer specific advantages and challenges. Model-based approaches can provide greater accuracy in representing the body's kinematics but require detailed anatomical models and complex optimization algorithms. On the other hand, model-free approaches, including anatomical information, are often simpler to implement and can be effectively used even with less sophisticated equipment, but they may suffer from lower precision in determining the positions of internal joints not directly visible from the external silhouette.

The primary limitations affecting the clinical applicability of many previous studies based on single camera were the requirement for color filters and uniform backgrounds to simplify subject segmentation (Castelli et al., 2015; Pantzar-Castilla et al., 2018). Additionally, many studies focused on single joint analysis, which does not capture the complexity of gait involving multiple interacting joints (Hatamzadeh et al., 2022; Leu et al., 2011; Pantzar-Castilla et al., 2018; Surer et al., 2011). Moreover, a significant limitation has been insufficient technical validation against established benchmarks and testing on pathological populations (Balta et al., 2020; Castelli et al., 2015; Leu et al., 2011; Saboune & Charpillet, 2005; Goffredo Michela and Carter, 2009).

The application of color filters techniques and homogeneous backgrounds simplifies the subject identification process but restricts the direct applicability of such methods outside controlled research environments. Furthermore, single joint analysis, while useful for specific investigations, fails to provide a comprehensive view of the entire gait cycle, which is essential for accurate diagnosis and treatment planning.

Additionally, the lack of technical validation against gold standards means that the reliability and accuracy of these methods in clinical practice remain uncertain. Moreover, testing on healthy populations does not account for the variations and challenges presented by pathological gait, limiting the generalizability of the findings to clinical populations.

In order to fill those gaps, the aim of the present chapter can be divided into three subsections:

- To propose and validate an innovative 2D model-based MS clinical gait analysis protocol on 18 patients with CP based on the use of a single RGB-D camera. This protocol introduced a 2D lower limb model, with joint center positions determined by applying the iterative closest points algorithm between the 2D model and dynamic 2D images. The accuracy and reliability of spatial-temporal parameters and sagittal lower limb joint kinematics were evaluated against a 3D marker-based (MB) protocol for clinical gait analysis.

- To propose and validate a novel 3D model-based MS protocol for clinical gait analysis on 6 subjects with CP and 4 individuals with clubfeet based on a single RGB-D camera to enhance the accuracy of sagittal lower limb kinematics estimation with respect to a 2D analysis. This protocol incorporates a 3D lower limb model, featuring interconnected foot, shank, thigh, and pelvis segments through revolute joints at the ankle, knee, and hip. Joint center positions were determined by applying the articulated iterative closest points algorithm directly to depth data. The reliability of sagittal lower limb joint angles was evaluated against a 3D MB protocol for clinical gait analysis.

- To propose and validate a 3D model-based MS protocol based on a single RGB-Depth camera in estimating sagittal ankle and metatarsophalangeal kinematics. A two-segment 3D foot model composed of mid-rear and forefoot foot connected by fifth metatarsophalangeal joint was introduced. Its clinical applicability was explored on 10 children with clubfeet. Kinematic curves from the proposed MS protocol were validated against those obtained by marking on the image the anatomical landmarks previously identified by manual palpation.

These studies received approval from the regional ethical review board in Gothenburg, Sweden (approval number 660-15).

2.2 State of the art

Recently, multi cameras model-based approach have been introduced to perform a 3D joint kinematics analysis, some of them are based on multiple body scanners while others use multiple RGB cameras.

MOVE4D system (Parrilla et al., 2019) is an advanced dynamic body scanning setup that utilizes a network of high-speed cameras to capture high-resolution 3D models of human movement. The MOVE4D system is named to reflect its capabilities in capturing and analyzing movement in four dimensions—three spatial dimensions plus time—hence "4D". This designation highlights its ability to create dynamic 3D scans that track the complex motions of the human body over time, providing a comprehensive view of kinematics that is critical for advanced applications in biomechanics, sports science, and other fields requiring precise movement analysis. This four-dimensional scanning capability allows for a detailed and temporal mapping of human motion, essential for accurate biomechanical assessments and the development of related technologies. These cameras are strategically positioned to encompass a large scanning area, enabling them to record rapid sequences of images at rates of up to thousands of frames per second. This capability is crucial for accurately documenting intricate human motions across a variety of actions and poses without any blur, making it ideal for detailed movement analysis in fields such as sports biomechanics and clinical studies of gait and posture. In addition to its hardware capabilities, the MOVE4D system includes sophisticated software designed to process the captured images into coherent 3D models. This software aligns multiple camera feeds to construct three-dimensional volumes of the subject in motion, meticulously filling in data gaps and eliminating noise to produce seamless and detailed models. These models are enriched with up to 50,000 points per body, providing an exceptionally high-resolution mesh that can represent minute anatomical features and dynamic changes in the body's surface as it moves. Furthermore, the system's modular design allows for flexible configurations to scan various parts of the body or full bodies with texture detail. It can provide a spatial resolution down to 1mm and capture frequencies ranging from 90 to 180 frames per second, depending on the configuration. Regarding model-based approaches based on multiple RGB cameras, Mündermann and Ceseracciu (Ceseracciu et al., 2014; Mündermann et al., 2006) have proposed the visual hull as a geometric technique used to approximate the shape of the human body from images taken from multiple RGB cameras. The process involves setting up multiple cameras around the subject to capture images simultaneously from various viewpoints. Each image is then processed to separate the subject (foreground) from the background, extracting the silhouette. The silhouettes from all cameras are combined to reconstruct a three-dimensional shape of the subject. This model, known as the visual hull, represents the maximal volume consistent with the silhouettes from all angles and serves as a rough approximation of the subject's shape. Then, a subject-specific model was matched to the subject's shape in each frame using the Articulated Iterative Closest Point (Pellegrini et al., 2008) algorithm, which is specifically tailored for objects with articulated parts, such as the human body. The proposed method includes a point cloud registration, where the visual hull's surface is treated as point clouds, and the Iterative Closest Point algorithm adjusts the model's segments to reduce the distance between the model's joints and the visual hull surface. Articulation constraints are then applied; unlike standard Iterative Closest Point that assumes rigidity, Articulated Iterative Closest Point integrates joint constraints to allow realistic movement. Finally, the model's alignment with the visual hull is iteratively refined to produce a detailed sequence of the human body in motion, closely replicating the captured movements (**Figure 4**).

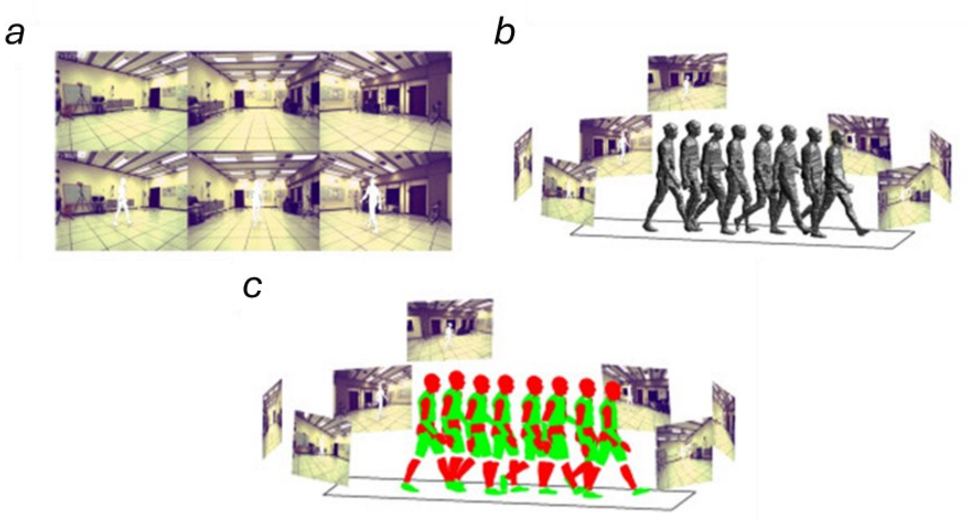


Figure 4. Model-based MS approach proposed by (Mündermann et al., 2006). a) Video sequences capturing the subject from various viewpoints. b) Selected visual hulls of the subject in motion from multiple angles. c) Articulated body matched to visual hulls for joint localization.

Leu and colleagues (Leu et al., 2011) proposed a model-free MS gait analysis system based on two RGB cameras. In order to extract joint centers, the subject in the image was extracted using a proper thresholding method in each view. Both vertical and horizontal projections for each view were defined as the number of segmented pixels in each row and as the number of white pixels in each column, respectively. As shown in **Figure 5**, the vertical projection was used to determine the location of the neck joint, in particular the y-coordinate corresponds to the minimum of the vertical projection while the x-coordinate is represented by a middle point of the segmentation in the y row of the segmentation. The horizontal projection was used in order to identify arms regions as the right and left 'hill' of the horizontal projection. Other joints are extracted by using statistical anatomical measurements of the body segments.

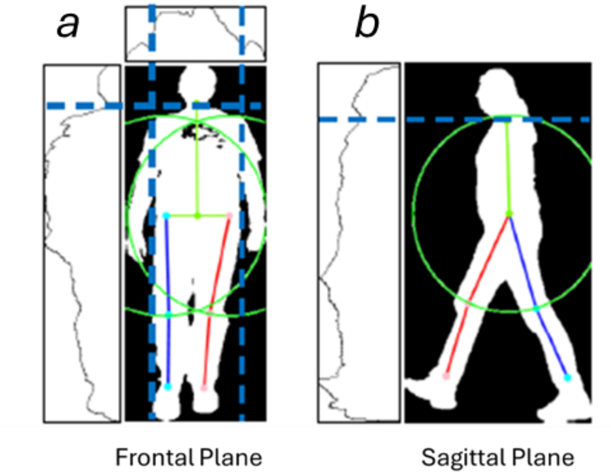


Figure 5. Process of localizing joint centers using segmented images of the human body in the a) frontal and b) sagittal planes.

However, the need for extensive installation and extrinsic camera calibration makes a multi-camera setup impractical for outpatient settings lacking dedicated laboratories. Furthermore, there are scenarios where two-dimensional (2D) joint kinematic analysis proves to be valuable for clinical applications, such as screening, identifying gait patterns, monitoring progress over time, and evaluating treatments. In these cases, portability, affordability, and ease of use are crucial. Therefore, methods utilizing a single camera with minimal setup time are preferable.

Also, Goffredo and colleagues (Goffredo Michela and Carter, 2009) proposed a model-free MS method from single RGB camera to estimate the lower limbs pose based on anatomical studies about human body anthropometric proportions (**Figure 6**). First of all, the silhouette was extracted by applying a thresholding method on the RGB image.

After that, the vertical positions of hip, knee, ankle were defined as:

$$y_{hip} = min\,(y_{sil}) + 0.5 * H$$

$$y_{knee} = min\,(y_{sil}) + 0.75 * H$$

$$y_{ankle} = min\,(y_{sil}) + 0.90 * H$$

where H is the subject’s height.

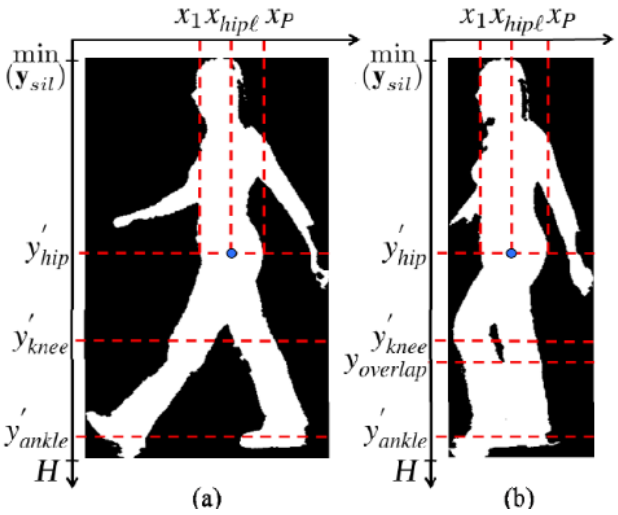


Figure 6. Process of localizing joint centers using segmented images of the human body based on anthropometric proportions.

However, these methods based on anthropometric measures could fail in correctly identifying the joint centers positions for pathological subjects with important gait and body asymmetries.

Surer and colleagues (Surer et al., 2011) developed a 2D model-based MS technique to study the sagittal kinematics of the shank-foot complex during the stance phase of the gait cycle using a single RGB camera. Their approach utilized a multi-rigid body model consisting of three segments: the shank (tibia and fibula), the rearfoot (tarsus and metatarsus), and the forefoot (phalanges). These segments were connected by cylindrical hinges, allowing for two degrees of freedom: the ankle's plantar/dorsi-flexion angle and the flexion/extension angle between the rearfoot and forefoot. Alignment between the above-mentioned foot model and the actual foot during the stance phase was achieved through cross-correlation between the two images. Although this seems a promising approach, the limit of this study is that it analyzed a single joint and a single phase of the gait cycle and that the cross-correlation technique could fail in describing segment deformity and in aligning the foot model during the swing phase which represent a challenging task.

Castelli (Castelli et al., 2015) and Panztar-Castilla (Pantzar-Castilla et al., 2018) have embraced the model-based methodology suggested by Mündermann but using a single camera approach including the alignment of a 2D multi-segment model with data acquired. Specifically, Castelli and colleagues, using a single RGB camera, have introduced a 2D multi-segment lower limb model and a singular value decomposition technique to align the reference system of the model with the one dynamically reconstructed on each frame of the gait cycle. However, this method was validated against the MB system on twenty healthy subjects and its clinical applicability on pathological populations has not been investigated.

Moreover, Panztar-Castilla (Pantzar-Castilla et al., 2018) have proposed a model-based MS approach using a single RGB-D camera. The proposed 2D lower limb model is composed by foot, shank, thigh, and pelvis segments, all interconnected by revolute joints at the ankle, knee, and hip, providing 6 degrees of freedom (DoF). The joint center trajectories were estimated by aligning a 2D lower-limb model to dynamic data. However, a significant limitation of Castilla's approach is the need for a green background to facilitate subject identification. This requirement complicates the setup process, thereby reducing the clinical applicability and portability of the system. Moreover, proper validation against the MB system has been conducted only for knee kinematics.

In conclusion, this paragraph highlights the need for further development of a markerless protocol specifically tailored for analyzing pathological data.

2.3 Concurrent validation of a 2D model-based approach on individuals with Cerebral Palsy based on a single RGB-D camera.

Clinical gait analysis is essential for understanding and interpreting the physio-pathological characteristics of human locomotion, and its diagnostic value is well-established. Specifically, it has proven effective in identifying optimal surgical procedures and guiding rehabilitation for individuals with bilateral cerebral palsy (Benedetti et al., 1998). While 3D analysis provides comprehensive data, a simpler 2D analysis on the sagittal plane can be sufficient for certain clinical applications, such as quantifying gait and addressing specific clinical questions (Harvey & Gorter, 2011).

For children with CP, 2D video analysis is particularly useful for documenting changes in gait patterns over time and for frequent monitoring during rehabilitation following interventions like multilevel surgery, orthotic adjustments, botulinum toxin injections, serial casting, and intensive therapy. A single-camera setup is adequate for capturing 2D gait analysis, simplifying the experimental setup, and reducing the need for extensive space, multiple cameras, and high costs. This approach allows for quantitative assessment of joint kinematics from video footage, enhancing patient care without additional resources.

In Sweden, there is a Cerebral Palsy Follow-Up Program (CPUP) follow-up program for children with cerebral palsy aims to monitor these children over time to detect and treat deformities early. Clinical exams are conducted biannually for children under 6 and annually for those aged 6-18. However, the program currently lacks an instrumented gait assessment component. Including such assessments would improve the detection of children at risk of crouch gait and enhance monitoring of changes over time. MB techniques, however, are impractical in this context due to the lengthy examinations, high costs, and need for dedicated space. To address these clinical needs, this study aims to propose and validate an innovative MS model-based protocol for clinical gait analysis using a single RGB-D camera on 18 patients with CP. The precision and consistency of spatio-temporal parameters and sagittal lower limb joint angles were evaluated against a 3D MB protocol for clinical gait analysis.

2.3.1 Material and methods

A. Participants: 18 patients enrolled in the Swedish CPUP program, consisting of 4 females and 14 males, aged between 6.5 and 28 years, with an average age of 15. The majority of participants had bilateral CP (11 individuals), while three had unilateral CP, three had dyskinetic CP, and one had ataxic CP. According to the Gross Motor Function Classification System (GMFCS), six participants were classified as level I, eleven as level II, and one as level III.

B. Experimental Setup: A Kinect v2 camera 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 made to coincide to the sagittal plane identified by the direction of progression and the vertical direction. Two LED lamps were also employed to ensure clarity and prevent blurring from automatic exposure.

C. Subject preparation: Subjects were instructed to wear colored socks—red for the right and blue for the left—and minimal clothing. An expert operator, after conducting a proper palpation, identified and marked five anatomical landmarks: the lateral malleolus (LM), lateral epicondyle (LE), great trochanter (GT), anterior superior iliac spine (ASIS), and posterior superior iliac spine (PSIS).

D. Data collection: Initially, two static lateral views (right and left) of each subject standing upright were recorded. After that, ten walking trials at self-selected speed were recorded for each participant (five for each side).

E. Validation: Data comparison was conducted using a 12-camera stereo-photogrammetric system (Oqus 400 Qualisys medical AB, Gothenburg, Sweden, fs = 100 Hz). Thirty-eight retro-reflective markers according to the modified Helen-Heyes model (Kadaba et al., 1990) were used and attached on the skin of the subject. Visual 3D software (C Motion Inc., USA) was used for calculating lower-limb joint angles.

Image pre-processing

The Heikkilä undistortion algorithm was employed to refine the calibration and correct the camera lens distortion (Herrera et al., 2012; Bouguet, 2022) and to finally provide intrinsic and extrinsic camera parameters. A matching operation was performed to align RGB and Depth images of the same dimensions. (N_{row} = 1080, N_{col} = 1536).

2.3.1.1 Method description

The proposed method comprised four key stages: identifying the gait cycle, segmenting the subject, calibrating subject-specific models, and estimating joint center trajectories (**Figure 7**).

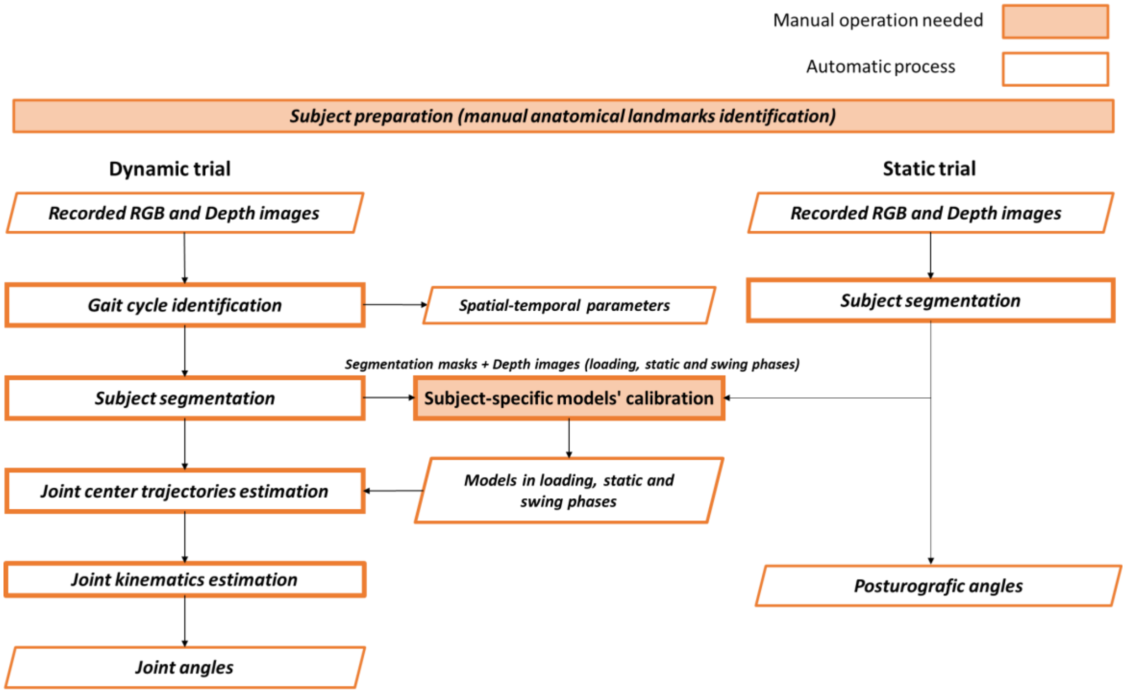

Manual operation needed
Automatic process
Subject preparation (manual anatomical landmarks identification)
Dynamic trial
Static trial
Recorded RGB and Depth images
Recorded RGB and Depth images
Gait cycle identification
Spatial-temporal parameters
Subject segmentation
Segmentation masks + Depth images (loading, static and swing phases)
Subject segmentation
Subject-specific models' calibration
Joint center trajectories estimation
Models in loading, static and swing phases
Joint kinematics estimation
Posturografic angles
Joint angles

Figure 7. Block diagram of the proposed 2D markerless protocol.

1) Gait cycle identification

The gait cycle identification allows:

- To select both RGB and Depth images representing the most central gait cycle;
- To compute the spatial-temporal parameters of interest (i.e., stride and step length, stride duration and gait speed).

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 (Rodda & Graham, 2001), these patients could present different types of gait:

- Equinus gait which is characterized by toe-walking and ankle plantarflexion. This pattern is often caused by calf muscle spasticity and can be managed by botox injections for reducing the spasticity, hamstring lengthening surgery and/or ankle-foot orthosis. In this case both initial and final contacts generally occur with the toe-ground contact (**Figure 8**a);

- Crouch gait: Individuals who walk in a crouch gait have excessive hip and knee flexion and shown a foot-flat contact (**Figure 8**b);

- Normal gait: in patients with a low GMFC system level initial contact occurs with the heel and final contacts occur with the toe (**Figure 8**c)

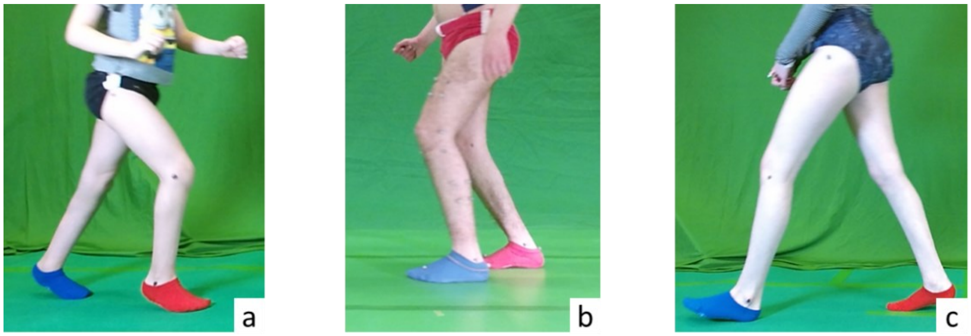


Figure 8. Different types of foot contacts a) Equinus gait, b) Crouch gait, c) Normal gait.

First, for each recorded frame, a binary segmentation mask ${}^{I}\boldsymbol{M}_{foot}$, expressed in the image coordinate system *I*, was obtained for each foot using a color filter segmentation technique (Cheng et al., 2001). Each ${}^{I}\boldsymbol{M}_{foot}$ (**Figure 9**a) was fitted within an ellipse. Then, a foot coordinate system (f) was established with axes aligned with the principal axes of the inertial ellipsoid and the origin at the centroid. The transformation matrix ${}^{I}\boldsymbol{T}_{f}$ from f to I was calculated using simple trigonometric formulas and applied to convert ${}^{I}\boldsymbol{M}_{foot}$ in the f (${}^{f}\boldsymbol{M}_{foot}$). From ${}^{f}\boldsymbol{M}_{foot}$, the point Q $\left({}^{f}\boldsymbol{Q}=\left[Q_{xf},Q_{yf}\right]\right)$ with the highest y-coordinate was tracked, the foot points included between Q_{yf} and (Q_{yf} - ϵ, where $\epsilon = 20$ pixels) were isolated and the foot sole was identified as the line fitting this region (**Figure 9**b). Similarly, the posterior foot edge was reconstructed starting from the point with the smallest x-coordinate. (**Figure 9**b).

The mid-rear foot (MRF) position in f $\left({}^{f}\boldsymbol{MRF}=\left[MRF_{xf},MRF_{yf}\right]\right)$ was determined by intersecting the foot sole with the posterior foot edge, previously identified, while the forefoot (FF) position in f $\left({}^{f}\boldsymbol{FF}=\left[FF_{xf},FF_{yf}\right]\right)$ was identified as the point with the highest x-coordinate (**Figure 9**b). The ${}^{f}\boldsymbol{MRF}$and ${}^{f}\boldsymbol{FF}$ were then referred to I by applying ${}^{I}\boldsymbol{T}_{f}$.

The foot points MRF and FF were considered to be in contact with the ground when their vertical y_I and horizontal x_I velocities feel below a specific threshold $Th = \pm 3$ pixels (**Figure 10**a). Initial contact (IC) of the foot was identified as the first instance between MRF and FF when the velocity along either their vertical and horizontal axis dropped to zero. A gait cycle was determined by two consecutive IC events of the foot in foreground, while a step was identified by the IC of foot in foreground followed by the consecutive IC of the foot in background.

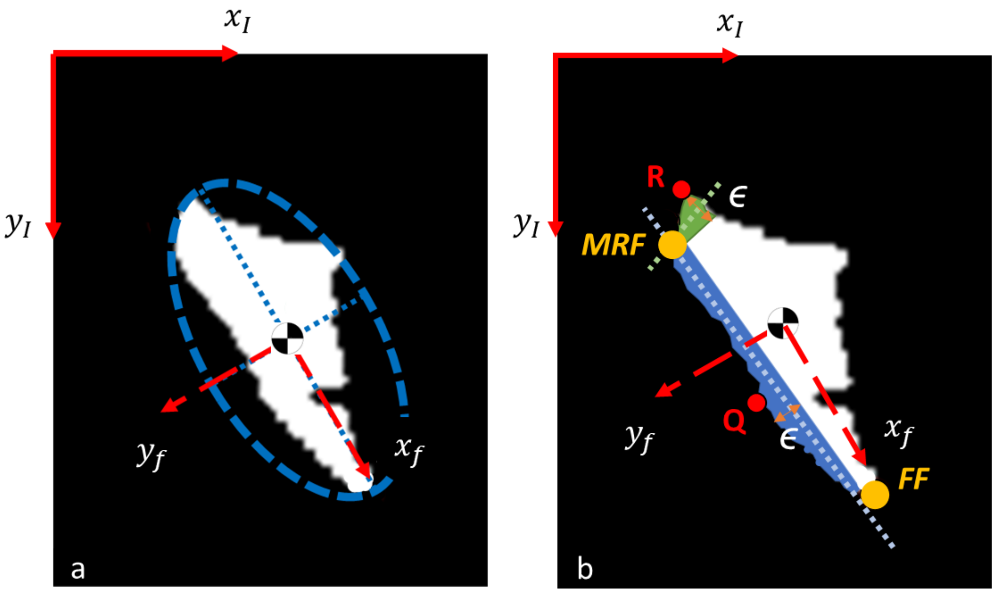


Figure 9. MRF and FF identification: a) An ellipse was drawn around each mask of the foot; the center and principal axes (x_f, y_f) were determined. b) The area where the foot sole (light blue) intersects with the posterior foot region (light green) was defined as the mid-rear foot (MRF), and the tip of the foot along the x-axis was defined as the forefoot (FF).

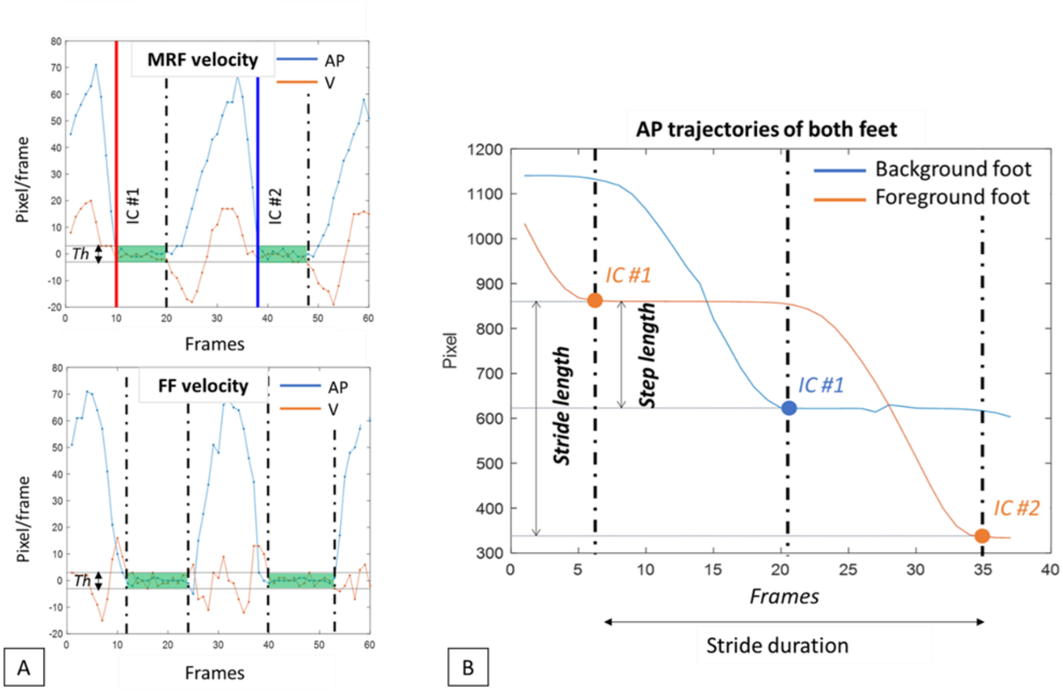


Figure 10: Stride length, stride duration, step length, and gait speed. A) Velocity of MRF and FF locations. The bold red line indicates the first initial contact (IC #1), while the blue line marks the subsequent initial contact (IC #2). The green areas show periods where MRF and FF are in contact with the ground (stationary condition). B) Stride length is measured as the distance between two successive initial contacts of the same foot (IC #1 and IC #2 in orange). Step length is the distance from the initial contact of one foot (IC #1 in orange) to the initial contact (IC #1 in blue) of the opposite foot.

Spatial-temporal parameters, such as stride length and duration, step length, and gait speed were determined by analyzing the coordinates of key foot points (MRF/FF) at the IC instants (**Figure 10**b).

2) Subject segmentation

The aim of the subject segmentation algorithm is to identify and isolate the subject in the image.

Background subtraction: for each frame, a preliminary subtraction operation between the frame containing the subject, ${}^{I}I(x,y,c)$, and the frame representing the background, ${}^{I}B(x,y,c)$, was performed as follows:

$${}^{I}D(x,y,c)=\left|{}^{I}I(x,y,c)-{}^{I}B(x,y,c)\right|$$

Where ${}^{I}D(x,y,c)$, ${}^{I}I(x,y,c)$, and ${}^{I}B(x,y,c)$ are the generic pixels expressed in I and c = $[r,g,b]$ is the color channel vector.

The resulting difference image ${}^{I}\boldsymbol{D}$ was converted to grayscale ${}^{I}\boldsymbol{D}_{gray}$ (**Figure 11**) by computing the norm of color channel of each pixel:

$${}^{I}D_{gray}(x,y)=\sqrt{{}^{I}D(x,y,r)^{2}+{}^{I}D(x,y,g)+{}^{I}D(x,y,b)^{2}}$$

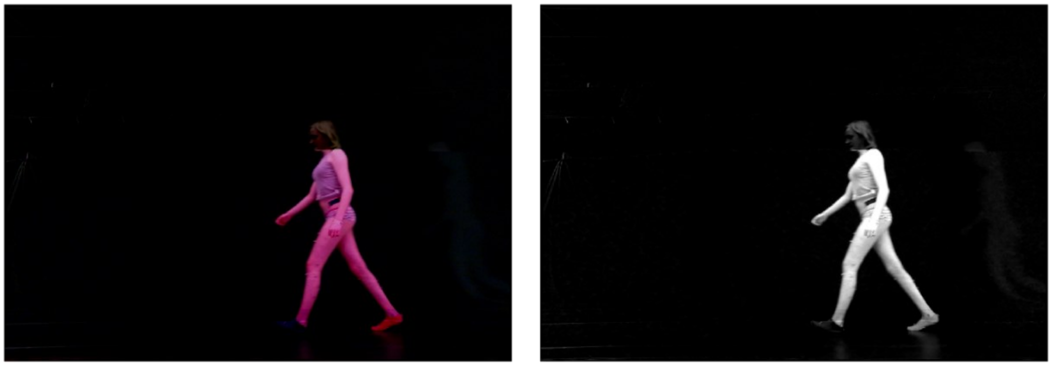


Figure 11. In the grayscale difference image *D*, background pixels have a lower intensity than the pixels representing the subject.

- Thresholding method: The subject was isolated from the image background by setting an appropriate threshold on the image pixel grey levels. This threshold was calculated by extracting the weighted mean of the grayscale histogram (Salvi & Molinari, 2018). (**Figure 12**):

$$Th = \frac{\sum_{i=0}^{255} w_i \cdot g_i}{\sum_{i=0}^{255} w_i}$$

where w_i represents the occurrence of each *i*-th grayscale level (g_i: 0,...,255).

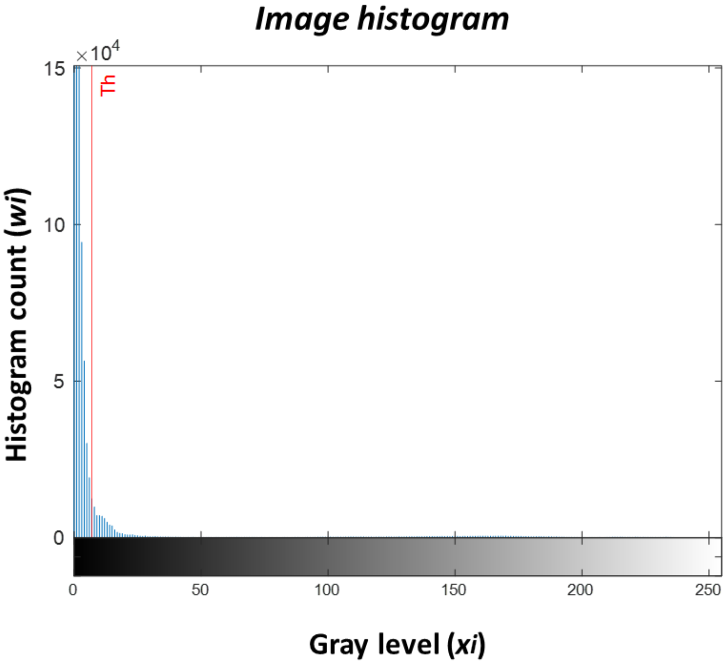


Figure 12. Image histogram of the difference image D. The red line represents the threshold Th.

The majority of pixels in image D represent the background and are therefore characterized by low-intensity values. In gait analysis, background pixels outnumber those of the subject by approximately ten to one. Using the weighted mean of the grayscale histogram ensures an inclusive threshold. This approach provides a low-intensity threshold that effectively separates the subject (with higher pixel values) from the background (with lower pixel values). This method is particularly advantageous in experimental conditions with non-uniform backgrounds, where the intensity difference between the subject and background pixels can be minimal.

The segmentation mask ${}^{I}\boldsymbol{M}_{sub}$ (**Figure 13**) was extracted from the ${}^{I}\boldsymbol{D}_{gray}$ as follows:

$$ {}^{I}M_{sub}(x,y) = \begin{cases} 1, & \left| {}^{I}D_{gray}(x,y) \geq Th \right. \\ 0, & \left| otherwise \right. \end{cases} $$

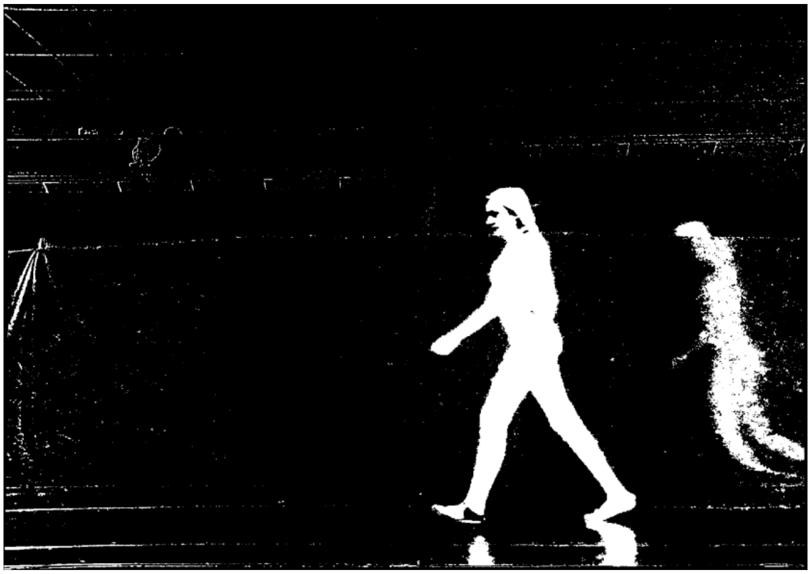


Figure 13. The final segmentation ${}^{I}\boldsymbol{M}_{sub}$ obtained by applying the threshold Th.

As shown in the **Figure 13**, the resulting image mask could show the subject segmentation but also other undesired regions due to residual noise in ${}^{I}\boldsymbol{D}$ or time-variant shadows. These undesired regions can be easily removed since the subject one is characterized by the largest connected area as shown in **Figure 14**.

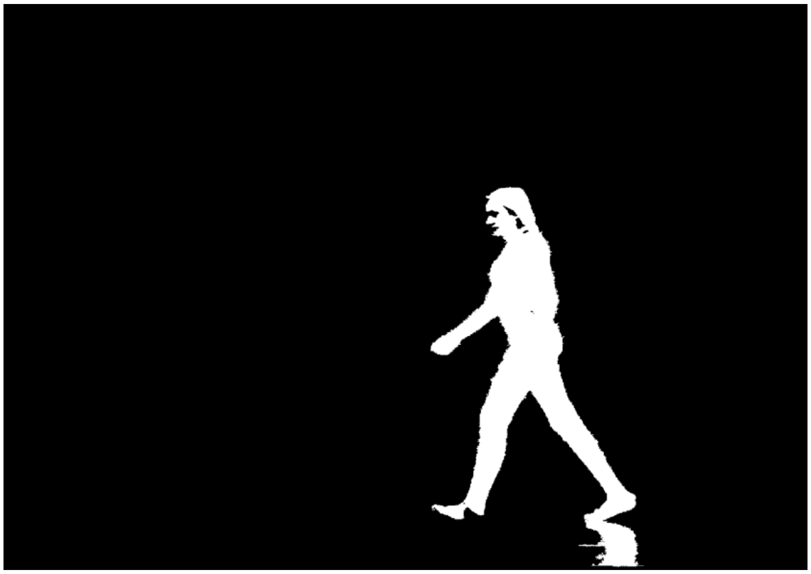


Figure 14. Segmentation mask obtained after the small areas' removal. The largest area is the subject.

After completing these steps, the feet segmentation remains suboptimal due to shadows on the walking surface. To address this, two color filters are applied to the RGB images: a red filter for the right foot and a blue filter for the left foot (**Figure 15**a). The feet were excluded from the segmentation mask, leading to the formation of small, easily removable connected regions (**Figure 15**b and **Figure 15**c).

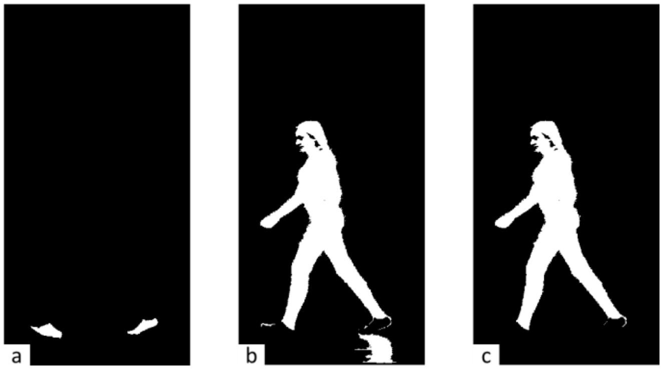


Figure 15. Identification of the feet (a), the feet are removed from the segmentation mask (b), the small, connected regions are removed (c).

The upper body segmentation (**Figure 15**c) was subsequently merged with the feet segmentation (**Figure 15**a), resulting in the final automated segmentation shown in **Figure 16**.

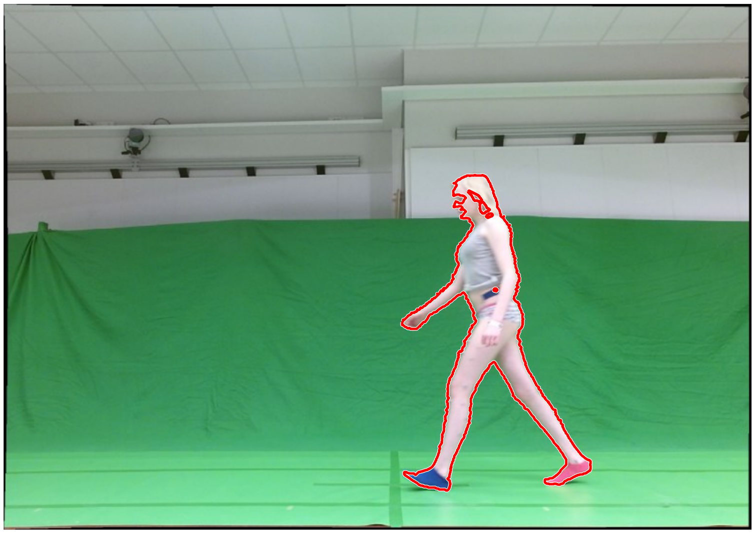


Figure 16. The final segmentation, achieved by implementing all the described steps, is represented by the red line.

3) *Subject-specific models' calibration*

To estimate lower-limb joint angles, a 2D subject-specific kinematic model of the lower limb was employed. This model consisted of four segments: the foot, shank, thigh, and pelvis. These segments were interconnected by revolute joints located at the ankle, knee, and hip, providing a total of six degrees of freedom (DoF). The foot segment, serving as the parent segment, featured two translational and one rotational DoF. The ankle joint center (AJC) was aligned with the lateral malleolus (LM), the knee joint center (KJC) corresponded to the lateral epicondyles (LE), and the hip joint center (HJC) was positioned at the greater trochanter (GT).

- *Anatomical calibration and body segment templates definition*

During an initial static acquisition while standing upright (image "0") foot, shank, thigh and pelvis templates and their corresponding coordinate systems were calibrated by manually identifying of the image the anatomical landmarks (LM, LE, GT, ASIS, PSIS), previously marked directly on the subject's body using a felt pen (diameter = 0.5

cm) by an expert operator, to determine their position vectors in I (${}^{I}\boldsymbol{LM}_0$, ${}^{I}\boldsymbol{LE}_0$, ${}^{I}\boldsymbol{GT}_0$, ${}^{I}\boldsymbol{ASIS}_0$, ${}^{I}\boldsymbol{PSIS}_0$). The identification of the point MRF and FF (${}^{I}\boldsymbol{MRF}_0$ and ${}^{I}\boldsymbol{FF}_0$) was performed as described in "Gait cycle identification". To address any potential asymmetries between the right and left sides of the body, a subject-specific model was defined for each side.

- *Foot template*

A template ${}^{I}\boldsymbol{TMP}_{foot}$ of the mid-rear foot portion was extracted from ${}^{I}\boldsymbol{M}_{foot}$,where the value of its generic pixel ${}^{I}TMP_{foot}(x,y)$ in the I was defined as:

$${}^{I}TMP_{foot}(x,y) = \begin{Bmatrix} 1, \left| {}^{I}M_{foot}(x,y) = 1 \cap MRF_{xi} < x < MRF_{xi} + 0.9l_f \right. \\ 0, \left| otherwise \right. \end{Bmatrix}$$

Where ${}^{I}M_{foot}(x,y)$ is a generic pixel of ${}^{I}\boldsymbol{M}_{foot}$ expressed in the I, l_f is the distance between ${}^{I}\boldsymbol{MRF}_0$ and ${}^{I}\boldsymbol{FF}_0$ (**Figure 17**).

In other words, pixels within the foot segmentation mask between the MRF x-coordinate and the 90% of the foot length were considered part of the foot template.

The foot coordinate system f_0 was defined as described in paragraph "Gait cycle ". The transformation matrix ${}^{I}\boldsymbol{T}_{f0}$ from f_0 to I determined and applied allowing the transformation of ${}^{I}\boldsymbol{TMP}_{foot}$ in the f_0 (${}^{f0}\boldsymbol{TMP}_{foot}$).

- *Shank template*

The central shank region was isolated using an annular area centered in ${}^{I}\boldsymbol{LM}_0$ and bounded by the radius $l_{shank25}$ and the radius $l_{shank75}$ equal to the 25% and the 75% of the distance between ${}^{I}\boldsymbol{LM}_0$ and ${}^{I}\boldsymbol{LE}_0$, respectively (**Figure 17**).

Then, the generic pixel ${}^{I}TMP_{shank}(x,y)$ of ${}^{I}\boldsymbol{TMP}_{shank}$ in *I* was obtained as:

$$ {}^{I}TMP_{shank}(x,y) = \begin{cases} 1, \left| {}^{I}M_{sub}(x,y) = 1 \cap l_{shank25} < \sqrt{x^2+y^2} < l_{shank75} \right. \\ 0, \left| otherwise \right. \end{cases} $$

In other words, pixels within the subject segmentation mask between the 25% and the 75% of the shank length were considered part of the shank template. Then, the top and bottom edges of the annular region were cut to create the final template.

${}^{I}\boldsymbol{TMP}_{shank}$ was fitted within an ellipse. Then, a shank coordinate system (s_0) was defined with the axes aligned with the principal axes of the inertia ellipsoid, and its origin set at the centroid of the ellipse. The transformation matrix ${}^{I}\boldsymbol{T}_{s0}$ from s_0 to *I* was calculated using basic geometric principles and applied to convert ${}^{I}\boldsymbol{TMP}_{shank}$ in the s_0 (${}^{s_0}\boldsymbol{TMP}_{shank}$).

- *Thigh template*

The central thigh region was isolated using an annular area centered in ${}^{I}\boldsymbol{LE}_0$ and bounded by the radii $l_{thigh25}$ and $l_{thigh75}$ equal to the 25% and the 75% of the distance between ${}^{I}\boldsymbol{LE}_0$ and ${}^{I}\boldsymbol{GT}_0$, respectively. (**Figure 17**).

Then, the generic pixel ${}^{I}TMP_{thigh}(x,y)$ of ${}^{I}\boldsymbol{TMP}_{thigh}$ in *I* was obtained as:

$$ {}^{I}TMP_{thigh}(x,y) = \begin{cases} 1, \left| {}^{I}M_{sub}(x,y) = 1 \cap l_{thigh25} < \sqrt{x^2+y^2} < l_{thigh75} \right. \\ 0, \left| otherwise \right. \end{cases} $$

In other words, pixels within the subject segmentation mask between the 25% and the 75% of the thigh length in static were considered part of the thigh template. Then, the top and bottom edges of the annular region were cut to create the final template

The thigh coordinate system t_0 and the transformation matrix ${}^{I}\boldsymbol{T}_{t_0}$.from t_0 to I were defined and applied to convert ${}^{I}\boldsymbol{TMP}_{thigh}$ in the t_0 (${}^{t_0}\boldsymbol{TMP}_{thigh}$).

- *Pelvis*

The inclination of the pelvis, relative to the x_I, was determined during the static upright standing acquisition based on the positions of the ${}^{I}\boldsymbol{ASIS}_0$ and ${}^{I}\boldsymbol{PSIS}_0$ (**Figure 17**).

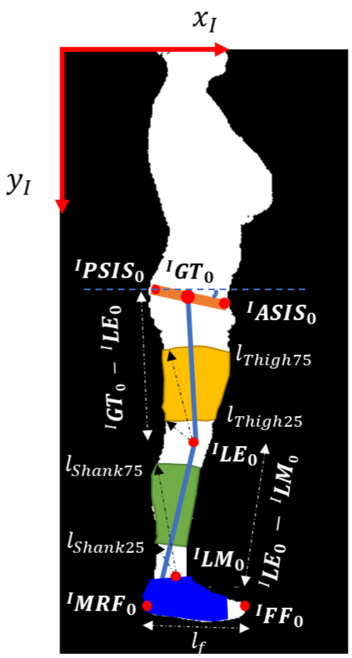


Figure 17: Body segment templates definition for the right side

4) Joint centers trajectories estimation

For each frame within the gait cycle, the positions of joint centers were determined using a bottom-up tracking approach. This method involves tracing the joint centers starting from the foot and moving upward towards the pelvis.

- *Ankle joint center (AJC) estimation*

The foreground foot was segmented from the RGB image based on color filters (${}^{I}\boldsymbol{M}_{foot}$) and the posterior foot region was extracted as proposed in the paragraph "Gait cycle ". The foot coordinate system f and its transformation matrix ${}^{I}\boldsymbol{T}_{f}$ was computed as proposed in paragraph "Gait cycle " (**Figure 18**a)

After having transformed ${}^{I}\boldsymbol{TMP}_{foot}$ (**Figure 18**b) and ${}^{I}\boldsymbol{M}_{foot}$ in *I* (**Figure 18**c), the origins of f and f_0 were initially aligned (**Figure 18**d) and then, using an iterative closest point (ICP) technique (Besl & McKay, 1992), the ${}^{f_0}\boldsymbol{TMP}_{foot}$ was aligned to the ${}^{f}\boldsymbol{M}_{foot}$ and the relevant matrix ${}^{f_0}\boldsymbol{T}_{f}$ (4x4), determined (**Figure 18**e).

The ${}^{I}\boldsymbol{AJC}$ coordinates referred in *I*, coincident to ${}^{I}\boldsymbol{LM}$, were determined for each frame by referencing the coordinates of LM in the template, ${}^{I}\boldsymbol{LM}_{0}$, through the application of the following three successive transformations:

$${}^{I}\boldsymbol{AJC} \equiv {}^{I}\boldsymbol{LM} = {}^{I}\boldsymbol{T}_{f}\,{}^{f}\boldsymbol{T}_{f_0}\,{}^{f_0}\boldsymbol{T}_{I}\,{}^{I}\boldsymbol{LM}_{0}$$

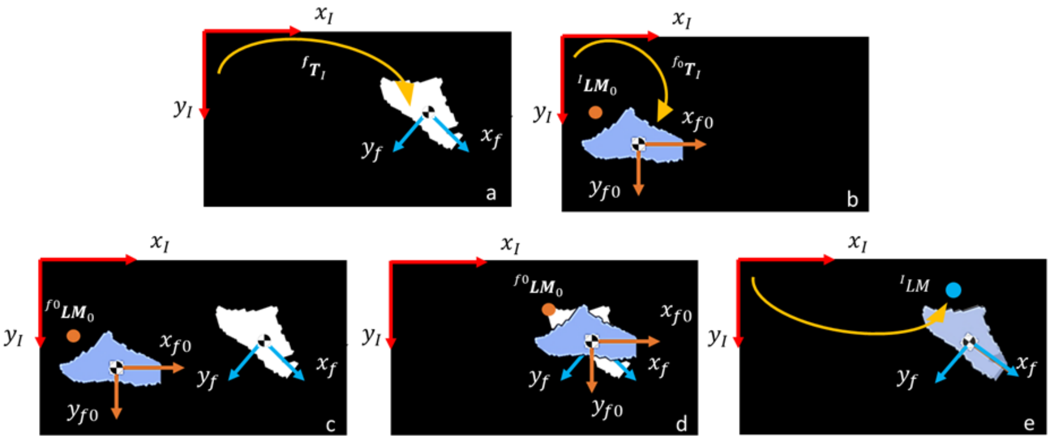


Figure 18: Ankle joint center estimation. a) ${}^{I}\boldsymbol{M}_{foot}$ and its relevant ${}^{I}\boldsymbol{T}_f$ b) ${}^{I}\boldsymbol{TMP}_{foot}$ with ${}^{I}\boldsymbol{LM}_0$ c) ${}^{I}\boldsymbol{TMP}_{foot}$ and ${}^{I}\boldsymbol{M}_{foot}$ in the common I. d) The origins of f and f_0 were made to coincide and e) the ${}^{f_0}\boldsymbol{TMP}_{foot}$ was matched with the ${}^{f}\boldsymbol{M}_{foot}$ and the relevant matrix ${}^{f_0}\boldsymbol{T}_f$ determined.

- *Knee joint center (KJC) estimation*

The separation of the foreground and background shanks was accomplished using two different strategies, depending on whether or not there was overlap between them. To distinguish between overlapping and non-overlapping conditions, a circle centered in ${}^{I}\boldsymbol{LM}$ with radius equal to the distance between ${}^{I}\boldsymbol{LM}_0$ and ${}^{I}\boldsymbol{LE}_0$ was drawn. If there was no overlap, the segmentation mask ${}^{I}\boldsymbol{M}_{sub}$ were grouped in two separated regions, and the foreground shank, being closer to the camera, coincided with the largest area (**Figure 19**a).

Conversely, in cases where overlap occurred, the shanks formed a single connected region. To separate the foreground from the background shanks under these circumstances, auxiliary depth sensor data were utilized. Specifically, a histogram of depth values within the connected region was created. Then, the Otsu method (Otsu, 1979) was applied to this histogram for binary classification. This method divides the data into two classes (class 0: foreground shank, class 1: background shank) by maximizing the variance between these classes, as shown in **Figure 19**b.

The central portion of the foreground shank (${}^{I}\boldsymbol{M}_{shank}$) was isolated using an annular area centered in ${}^{I}\boldsymbol{LM}_{0}$ and bounded by the radius $l_{shank25}$ and the radius $l_{shank75}$.

The shank coordinate system s was reconstructed with the axes coincident to the inertial ellipsoid principal axes of the ${}^{I}\boldsymbol{M}_{shank}$ and the transformation matrix ${}^{I}\boldsymbol{T}_{s}$ from s to I determined. The origin of s and s_0 were initially aligned and then the ${}^{s_0}\boldsymbol{TMP}_{shank}$ was matched with the ${}^{s}\boldsymbol{M}_{shank}$ and the relevant matrix ${}^{s_0}\boldsymbol{T}_{s}$ (4x4) was computed using an ICP technique (Besl & McKay, 1992).

The ${}^{I}\boldsymbol{KJC}$ coordinates in I, coincident to ${}^{I}\boldsymbol{LE}$, were determined for each frame by referencing the coordinates of LE in the shank template, ${}^{I}\boldsymbol{LE}_{0}$, through the application of the following three successive transformations:

$${}^{I}\boldsymbol{KJC} \equiv {}^{I}\boldsymbol{LE} = {}^{I}\boldsymbol{T}_{s}\,{}^{s}\boldsymbol{T}_{s_0}\,{}^{s_0}\boldsymbol{T}_{I}\,{}^{I}\boldsymbol{LE}_{0}$$

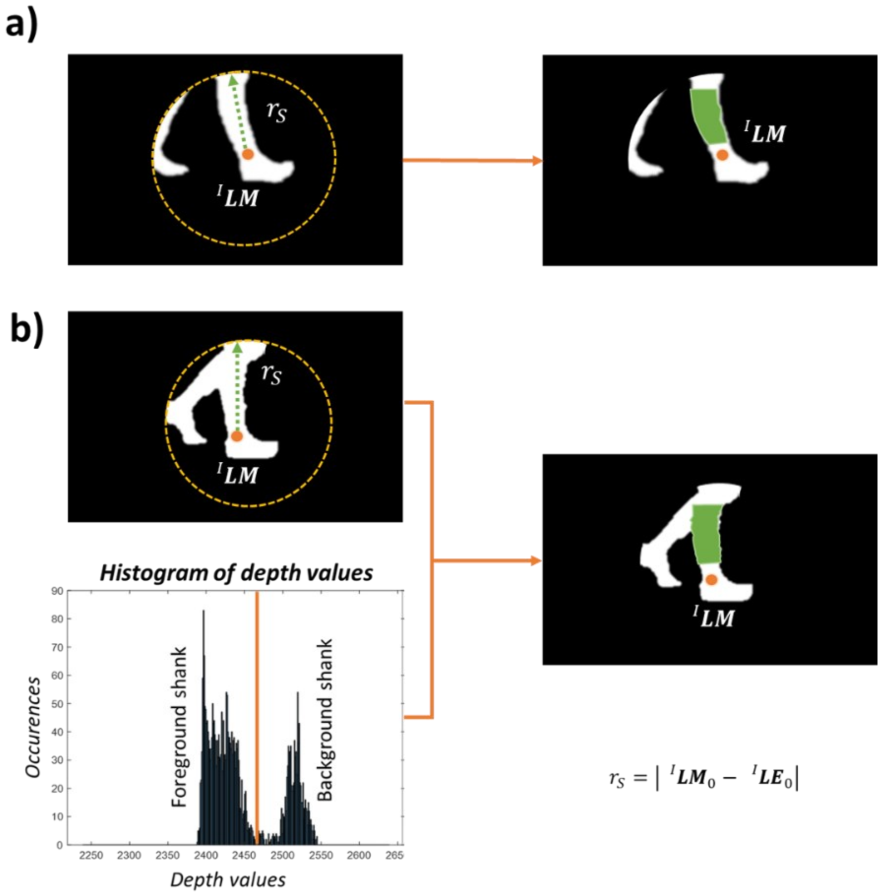


Figure 19: Separation between foreground and background shanks: A circle was centered at ^{I}LM with its radius equal to the Euclidean distance between $^{I}LM_0$ and $^{I}LE_0$. A) No overlap between the two shanks. On the left, two distinct regions were identified. On the right, the foreground shank $^{I}M_{shank}$ was recognized. B) Overlap between the two shanks. On the left, a single connected region was observed, and the histogram of depth values inside the region was computed; the Otsu method was then used to separate the two shanks. On the right, the foreground shank $^{I}M_{shank}$ was identified.

- *Hip joint center (HJC) estimation*

To distinguish the foreground thigh from the background thigh and the hand during arm oscillation, two different procedures were used based on the positioning of the foreground hand. If the foreground hand overlapped with the foreground thigh, one procedure was applied; if there was no overlap, another procedure was utilized. Preliminarily, a circle centered in ${}^{I}\boldsymbol{LE}$ with radius equal to the distance between ${}^{I}\boldsymbol{LE}_0$ and ${}^{I}\boldsymbol{GT}_0$ was drawn and the envelope of the histogram of depth values of the pixels within this circle was computed and the maxima were identified.

In scenarios where the foreground hand overlapped with the thigh, the histogram envelope revealed three distinct peaks representing the foreground hand (class 0), foreground thigh (class 1), and background thigh (class 2), as shown in **Figure 20**a. For these cases, the Otsu method (Otsu, 1979) was applied to classify these three classes. Alternatively, as depicted in **Figure 20**b, when the hand is not present in the segmentation, a binary classification was used to differentiate between the foreground thigh (class 0) and the background thigh (class 1).

The central portion of the foreground thigh (${}^{I}\boldsymbol{M}_{thigh}$) was extracted as the region included in the anulus centered in ${}^{I}\boldsymbol{LE}$ and defined by the radius $l_{thigh25}$ and the radius $l_{thigh75}$. The thigh coordinate system, t, was reconstructed with the axes coincident to the inertial ellipsoid principal axes of the ${}^{I}\boldsymbol{M}_{thigh}$ and the transformation matrix ${}^{I}\boldsymbol{T}_t$ from t to I determined. The origin of t and t_0 was initially aligned and then the ${}^{t_0}\boldsymbol{TMP}_{thigh}$ was matched with the ${}^{t}\boldsymbol{M}_{thigh}$ and the relevant matrix ${}^{t_0}\boldsymbol{T}_t$ (4x4), was computed, using an ICP technique (Besl & McKay, 1992)

The ${}^{I}\boldsymbol{HJC}$ coordinates referred to I, coincident to ${}^{I}\boldsymbol{GT}$, were determined for each frame by referencing the coordinates of GT in the thigh template, ${}^{I}\boldsymbol{GT}_0$, through the application of the following three successive transformations:

$$ {}^{I}\boldsymbol{HJC} \equiv {}^{I}\boldsymbol{GT} = {}^{I}\boldsymbol{T}_t\,{}^{t}\boldsymbol{T}_{t_0}\,{}^{t_0}\boldsymbol{T}_I\,{}^{I}\boldsymbol{GT}_0 $$

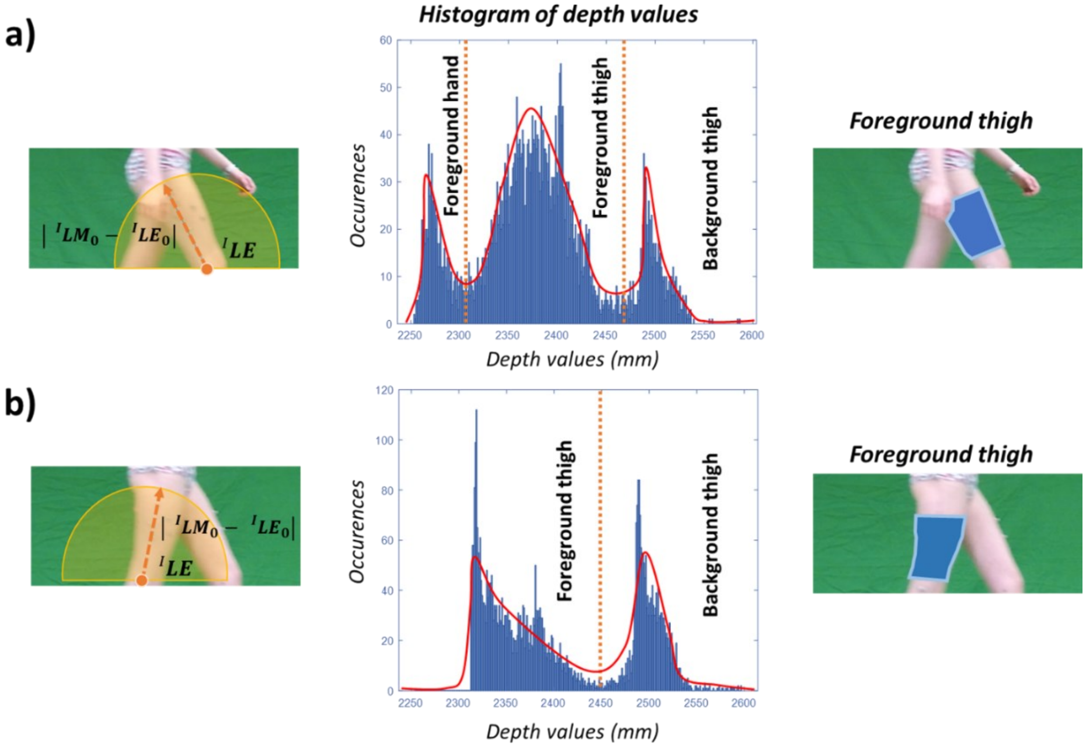


Figure 20: Separation between foreground and background thighs: A circle was drawn centered on ^{I}LE, with its radius determined by the Euclidean distance between $^{I}LE_0$ and $^{I}GT_0$. Within this area, the histogram of depth values was computed, and the envelope defined. a) With a foreground hand superimposed on the thigh, the envelope displayed three peaks, prompting the implementation of the Otsu method for three-class classification. b) Without a foreground hand in the circular area, the envelope showed only two peaks, leading to the use of the Otsu method for binary classification.

It is important to note that during gait, the size and shape of lower limb body segments are subject to variations. These changes are attributed to soft tissue deformation (Cereatti et al., 2017), alterations in the subject's position relative to the camera, and potential movements outside the sagittal plane, all of which can compromise the effectiveness of matching body segment templates to the segmented body masks. To address these issues, a multiple calibration procedure was employed (Cappello et al., 1997), utilizing three different sets of body segment templates. The first set was defined using the subject's standing posture (**Figure 21**a), while the second and third sets were derived from selected frames during the loading and swing phases of the gait cycle, respectively (**Figure 21**b and **Figure 21**c). The identification procedure for joint center trajectories in each frame of the gait cycle using the ICP algorithm, as outlined in paragraph "Joint centers trajectories estimation", was then applied using these additional templates, resulting in three distinct trajectories for each joint center.

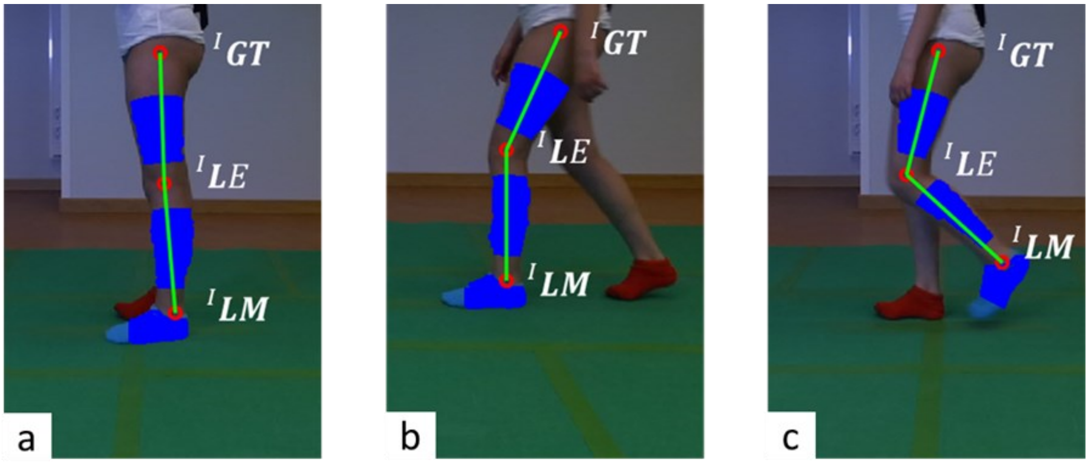


Figure 21: Set of body segment templates definition during static phase (a), loading phase (b) and swing phase (c).

5) Joint kinematics estimation

Joint angles were determined using the inclination of segments, defined connecting the joint centers. The plantar-dorsi flexion angle of the ankle was calculated as the angle between the segment which best fits the foot centered in ${}^{I}\boldsymbol{AJC}$ 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. Hip flexion-extension angle was calculated as the angle between the ${}^{I}\boldsymbol{KJC}$ - ${}^{I}\boldsymbol{HJC}$ vector and the constant direction identified by the ${}^{I}\boldsymbol{ASIS}_0$ - ${}^{I}\boldsymbol{PSIS}_0$ vector (pelvic tilt) during the static acquisition. For each joint, kinematic analysis generated three separate curves using the three sets of templates corresponding to static, loading and swing phases. These curves were subsequently merged to obtain a single comprehensive curve through the application of a nonlinear sinusoidal weighting function (Cereatti et al., 2015).

2.3.2 Performance assessment and statistical analysis

The precision of gait event identification was assessed by calculating the mean absolute error (MAE) and mean error (ME) based on the time differences between the gait events visually identified from RGB images by a team of expert clinicians and those estimated automatically by the MS method across trials and subjects. Additionally, the spatial-temporal gait parameters estimated were evaluated in terms of MAE, MAE%, ME, and ME%, compared to the results from the 3D MB protocol across trials and subjects. Prior to this comparison, both the MS and MB kinematic data were processed with a fourth-order Butterworth filter at a 7 Hz cutoff frequency and were time-normalized to the gait cycle, from 1% to 100% (Bergamini et al., 2014).

The performance of the proposed MS protocol for each subject, s, gait trial, t, and joint, j, was evaluated in terms of offset and waveform similarity (Picerno et al., 2008). The offset was the absolute difference between the mean value of the MS ($\overline{MS}$) and MB kinematic curves ($\overline{MB}$) within a gait cycle:

$$Offset_{s,t,j} = \left| \overline{MB_{s,t,j}} - \overline{MS_{s,t,j}} \right|$$

For each joint, the latter values were then averaged across trials and subjects:

$$Offset_{j} = \frac{1}{N_S} \sum_{s=1}^{N_S} \frac{1}{N_T} \sum_{t=1}^{N_T} Offset_{s,t,j}$$

Where $N_S = 18$ is the number of patients with CP and $N_T = 10$ is the number of recorded trials.

The root mean square error (RMSE) between the MS and MB joint kinematic curves was computed for each subject, gait trial, and joint to assess the waveform similarity. Prior to computing the RMSE, the mean values from each set of kinematic curves were removed (Picerno et al., 2008):

$$RMSE_{s,t,j} = RMS\left(\left(MB_{s,t,j} - \overline{MB_{s,t,j}} \right) - \left(MS_{s,t,j} - \overline{MS_{s,t,j}} \right) \right)$$

For each joint, the latter values were then averaged across trials and subjects:

$$RMSE_j = \frac{1}{N_S}\sum_{s=1}^{N_S}\frac{1}{N_T}\sum_{t=1}^{N_T} RMSE_{s,t,j}$$

Where $N_S = 18$ is the number of patients with CP and $N_T = 10$ is the number of recorded trials.

A set of key gait features, considered as clinically relevant, were extracted from both MB and MS sagittal lower limb joint angle after removing their offsets, as detailed in (Benedetti et al., 1998). These features include: (**Figure 22**):

- K1: The knee flexion at the initial contact (0% of the gait cycle);
- K2: Maximum knee flexion observed during the loading response, between 0% and 40% of the gait cycle;
- K3: Maximum knee extension occurring during the stance phase, from 25% to 75% of the gait cycle;
- K5: Maximum knee flexion during the swing phase, from 50% to 100% of the gait cycle;
- A3: Maximum ankle dorsiflexion during the stance phase, between 25% and 75% of the gait cycle;
- A5: Maximum ankle plantar-extension during the swing phase, from 50% to 100% of the gait cycle;
- H3: Maximum hip extension during the stance phase, also from 25% to 75% of the gait cycle.

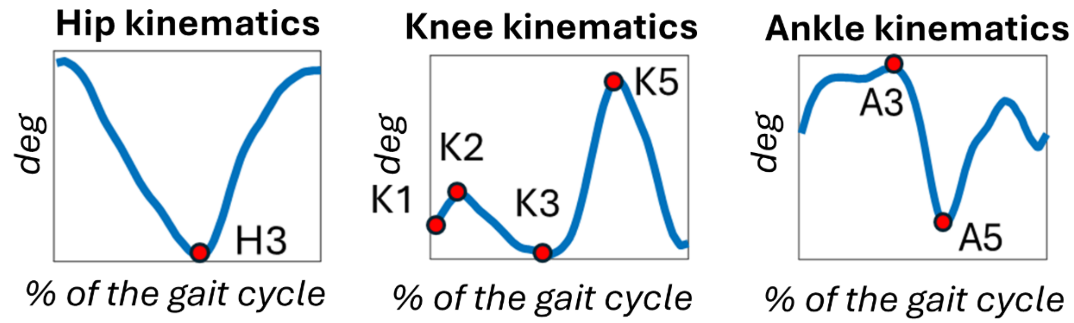


Figure 22. Seven key gait features extracted from sagittal hip, knee and ankle kinematics.

For each clinically relevant key gait feature, the MAE was calculated with respect to the estimates from the MB method. Additionally, 95% confidence intervals (95% CI) for this error were computed.

Reliability for each gait feature and each method (MS and MB) was assessed using the intraclass correlation coefficient (ICC) based on absolute agreement and a 2-way random effects model ($ICC(2,k)$). This analysis was conducted based on data collected over multiple subjects (n = 18) and different gait cycles (k = 10). According to the criteria specified in (Koo & Li, 2016), ICC values lower than 0.5 indicate poor reliability, values between 0.5 and 0.75 suggest moderate reliability, values between 0.75 and 0.9 denote good reliability, and values greater than 0.90 reflect excellent reliability.

Additionally, the differences between the MS and MB protocols were examined using Spearman's correlation coefficient (R). The correlation values are interpreted as follows: values less than 0.19 demonstrate a negligible relationship, values from 0.20 to 0.29 indicate a weak relationship, values from 0.30 to 0.39 suggest a moderate relationship, values from 0.40 to 0.69 show a strong relationship, and values above 0.70 provide a very strong relationship (Dancey & Reidy, 2007).

2.3.3 Results

Table 2 reports the results obtained for gait events identification, spatial-temporal gait parameters and lower limb joint kinematics. Results from **Table *2*** indicate the lowest MAE% for spatial-temporal gait parameters were achieved in stride duration at

2%, followed by step length at 2.2%, stride length at 2.5%, and gait speed at 3.1%. As for the lower-limb joint kinematics **Table 3**, RMSE values ranged from 3.2 deg to 4.5 deg, with the smallest RMSE observed in knee joint kinematics at 3.2 deg, hip joints at 3.5 deg, and ankle joints at 4.5 deg.

The evaluation of key gait features as presented in **Table 3** shows that MAE values ranged between 3.1 deg and 5.9 deg. The reliability assessments and correlation studies detailed in **Table 4** demonstrate that both MS and MB protocols displayed excellent reliability for K1, K2, K3, K5, and H3 with ICC values from 0.90 to 0.94. The ankle features A3 and A5 demonstrated good reliability with ICC values between 0.80 and 0.88. The correlation analysis revealed very strong relationships for knee and hip gait features with correlations of 0.85 or higher, and a strong relationship for ankle kinematics features at 0.66.

Table 2: Comparison between Visual inspection and Automated Identification of Initial Foot Contacts in terms of mean absolute error (MAE) and mean error (ME) for both feet in milliseconds. MAE and ME of the gait speed, stride length, stride duration and step length with respect to MB system. Lower limb joint kinematics. The average root-mean-square errors (RMSE) value between the MS and the MB lower limb joint kinematics calculated over the entire gait cycle, averaged across all subjects, and recorded trials.

	ME (95% CI)	***MAE (95% CI)***
IC detection (ms)	4 ± 20	20 ± 20
Spatial-temporal parameters	***ME (ME%)***	***MAE (MAE%)***
Gait speed (m/s)	-0.01 (1.5 ± 3)	0.02 (3.1 ± 1.7)
Stride length (cm)	-0.7 (0.8 ± 2.5)	2.2 (2.5 ± 0.9)
Stride duration (ms)	-6.4 (0.7 ± 2.5)	20 (2.0 ± 1.6)
Step length (cm)	-0.1 (0.3 ± 2.6)	1.2 (2.2 ± 1.1)
Joint angles	***Offset (°)***	***RMSE(°)***
Ankle	8	4.5
Knee	6	3.2
Hip	7	3.5

Table 3: Mean absolute error (MAE) of seven gait features extracted from MS and MB protocols along with 95 % CI: 95% of confidence interval.

Gait features (°)	***MAE (95% CI)***
A3	3.5 (3.1, 4.1) *
A5	3.1 (2.6, 3.6) *
K1	3.7 (3.2, 4.3) *
K2	4.8 (4.1, 5.6) *
K3	3.8 (3.3, 4.5) *
K5	5.9 (5.9, 6.8)
H3	3.7 (3.2, 4.3)

Table 4: Reliability in the estimation of each gait feature obtained from the MS and MB protocol. MS ICC (2, K): intraclass correlation for MS protocol, MB ICC (2, K): intraclass correlation for MB protocol, R: spearman's correlation coefficient.

Gait features (°)	*MS ICC (2, k)*	*MB ICC (2, k)*	*R*
A3	0.88	0.80	0.66
A5	0.80	0.81	0.66
K1	0.91	0.94	0.85
K2	0.90	0.82	0.85
K3	0.93	0.94	0.94
K5	0.93	0.94	0.90
H3	0.94	0.94	0.86

An overview of joint kinematics curves, averaged over trials and subjects, and normalized between 0 and 100% of the gait cycle is depicted in **Figure 23**.

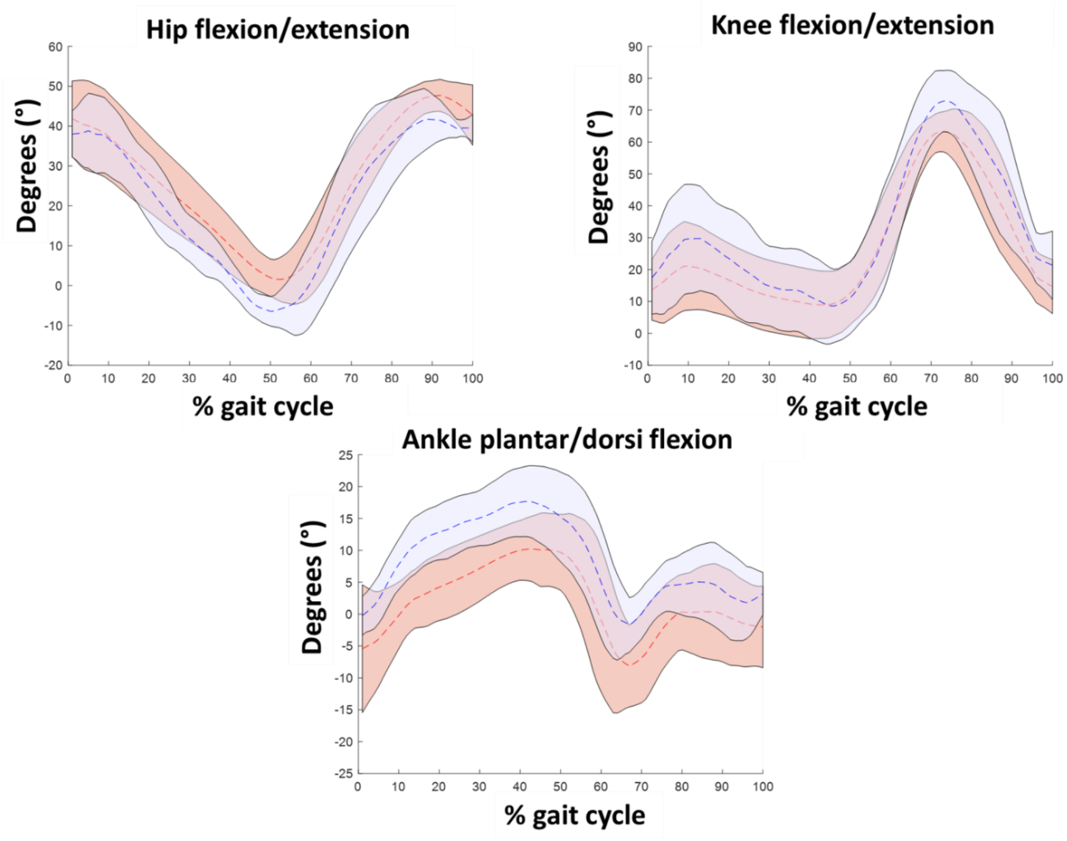


Figure 23: Average sagittal lower limb joint kinematics (hip, knee, and ankle) across subjects and trials. Dashed lines represent the mean values, and the shaded area indicates the standard deviation (SD). Red lines correspond to the MB system, while blue lines correspond to the MS system.

2.3.4 Discussions

The study was designed to assess the accuracy and reliability of a clinical MS gait analysis protocol using a single RGB-depth camera. This protocol estimates spatial-temporal parameters and sagittal lower-limb joint angle in patients with CP. This research aims to provide a quantitative tool for easy implementation in screening and monitoring the motor progression related to the disease in these patients.

The protocol achieves several key enhancements over prior research (Pantzar-Castilla et al., 2018). Firstly, it introduces an automatic thresholding segmentation algorithm that eliminates the need for a homogeneous background, enhancing the clinical usability and portability of the system. Secondly, it addresses issues such as left-right confusion (Nguyen et al., 2022) and problems with skeleton tracking when foreground and background elements overlap (Pantzar-Castilla et al., 2018), by deploying a robust separation strategy that maximizes the variance between the depth values of foreground and background lower-limbs. Thirdly, the method expands the validation of clinical concurrent validity beyond just knee kinematics (Pantzar-Castilla et al., 2018), to also include hip and ankle kinematics together with spatial-temporal parameters, resulting in approximately a 35% increase in the accuracy of knee joint gait measurements (Pantzar-Castilla et al., 2018).

Additionally, the method for detecting gait events is specially designed to consider the different types of foot contact typically seen in patients with CP—heel, flat, and toe-ground contacts. In particular, this method depends on the orientation of the foot model relative to the ground, marking a significant improvement with respect to other studies that focus solely on the 3D coordinates of the ankle joint center, thus neglecting the type of foot contact (Albert et al., 2020; Bertram et al., 2023; Castelli et al., 2015; Cimolin et al., 2022; Clark et al., 2013; Ferraris et al., 2021; Lonini et al., 2022).

A) Spatial-temporal parameters

The deterministic model-based MS protocol, proposed in this research, demonstrated very high accuracy, meeting clinical standards with MAE values of 1.2 cm for step length, 20 ms for stride duration, 2.5 cm for stride length, and 0.02 m/s for gait speed. These error values in patients with CP were comparable to those found in prior single-camera studies involving healthy subjects (Albert et al., 2020; Bertram et al., 2023; Castelli et al., 2015; Clark et al., 2013; Hatamzadeh et al., 2022; Yamamoto et al., 2021).

To the best of the authors' knowledge, this study is the first to specifically validate spatial-temporal parameters in patients with CP using stereophotogrammetric system for comparison. Previously, single-camera methods have been validated in studies focusing on post-stroke and Parkinson's patients, showing lower performance. For instance, studies by (Ferraris et al., 2021) and (Cimolin et al., 2022) assessed the accuracy of spatio-temporal parameter extracted from body tracking SDK of Kinect v2 in post-stroke and Parkinsonian subjects, respectively. They found ME values of 0.02

m/s for gait speed and 2 cm for step length, which are higher than the errors observed in our MS protocol (0.01 m/s for gait speed and 0.06 cm for step length). Lonini et al ., 2022 proposed a MS algorithm for gait analysis using DeepLabCut software on post-stroke patients with a single RGB camera and found significant error variability in gait speed (± 0.11 m/s ME).

B) Lower-limb joint kinematics

In assessing the 2D joint kinematics determined by the MS method with respect to MB protocol which provides 3D kinematics, it is crucial to distinguish between the impacts of using different angular conventions and different definition of anatomical axes from the actual estimation errors (Picerno et al., 2008). The different anatomical axis definitions primarily result in an offset between curves, whereas inaccuracies in reconstructing joint center trajectories impact waveform similarity, quantifiable by the RMSE after offset removal.

The average offset was 8 deg for the ankle joint, 6 deg for the knee joint, and 7 deg for the hip. The ankle offset might be due to the fact that the MS 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. The knee joint offset is linked to different definitions for HJC identifications between MS and MB protocols. The MS protocol identifies the HJC coinciding with the GT, while the MB protocol locates it at the geometric center of the acetabulum, identified using an anthropometric regression equations (Davis et al., 1991). The hip joint offset arises because the MS protocol assumes a constant pelvis inclination during gait, determined as the pelvic tilt during static posture.

In terms of waveform similarity, the knee joint angle curves was the most precise, with an RMSE of 3.2 deg, followed by the hip joint with an RMSE of 3.5 deg, and the ankle joint with an RMSE of 4.5 deg. Clinically, errors between 2 deg and 5 deg are typically considered acceptable but they need for careful interpretation (McGinley et al., 2009). The largest errors in ankle kinematics primarily stem from the camera's auto-exposure, which can blur images of the foot and lower shank during rapid movements like the swing phase.

In recent years, various single-camera MS methods for gait analysis have been proposed. However, often an immediate comparison with the proposed MS protocol was not feasible since: (*i*) the kinematic outputs were not validated against a clinical benchmark (Amprimo et al., 2021; Ferraris, Amprimo, Masi, et al., 2022; Latorre et al., 2018, 2019), (*ii*) the method's performance was only validated in term of precision in tracking joint centers (Hesse et al., 2023), (*iii*) the ultimate objective was to categorize motor activities or abnormalities in gait patterns (Chen et al., 2011; Clark et al., 2015; Ferraris, Amprimo, Pettiti, et al., 2022; Kojovic et al., 2021; Li et al., 2018; Stricker et al., 2021).

To the best of the authors' knowledge, the sole MS study involving children with CP was performed by (Ma et al., 2019). This research evaluated the concurrent validity of Kinect v2 body tracking SDK against a stereophotogrammetric system using a frontal view on 10 children with CP (GMFS I-II). Significant errors were reported for all joints, with RMSE values of 11.2 deg for the hip, 10.3 deg for the knee, and 7.5 deg for the ankle.

Moreover, a few previous MS studies have been validated solely on normal gait (Castelli et al., 2015; Yamamoto et al., 2021; Yeung et al., 2021). Yeung and colleagues (Yeung et al., 2021) investigated the impact of five camera viewing angles on kinematic curve estimates in healthy subjects using the body tracking SDK of Kinect v2, finding that a frontal viewing angle resulted in better performance with a RMSE of 8 deg for hip flexion/extension, 11.4 deg for the sagittal knee, and 17.4 deg for ankle plantar/dorsi flexion.

(Yamamoto et al., 2021) evaluated the performance of OpenPose on healthy individuals, finding similar reliability for knee and hip kinematics, with ICC values between 0.60 and 0.98. However, they observed significantly lower reliability for ankle angles, with an ICC of 0.1 for maximum dorsiflexion during the stance and swing phases (A3 and A5, respectively). In contrast, our MS protocol demonstrated an ICC of 0.90 for the same measurements.

Castelli (Castelli et al., 2015) reported RMSE values of 3 deg at the ankle, 3.6 deg for the knee, and 4.8 deg for the hip in 10 healthy subjects, comparable to those obtained using our method on individuals affected by CP.

It is worth to notice that our MS protocol demonstrates accuracy comparable to Salford Gait Tool (widely used observation-based clinical gait assessment tool) which involves manually identifying anatomical landmarks in each recorded image (Larsen et al., 2012). In their research, Larsen and colleagues assessed the Salford Gait Tool's accuracy against the MB protocol in 10 adult patients with CP, revealing similar errors in key gait features. This indicates that the proposed MS method achieves performance comparable to observation-based tools but with significantly less clinician effort, as manual intervention is required only for calibrating the three templates.

2.4 Reliability of a 3D model-based approach with a Single RGB-D Camera

Despite employing three models, in the above-mentioned 2D MS protocol, residual errors persist due to the limitations of 2D modeling in compensating for out-of-plane movements and position changes between the subject and the camera that could occur during gait. In addition, the 2D MS protocol requires manual identification of anatomical landmarks during models' calibration.

The aim of this work was to design an extension of the 2D MS protocol benefiting from a 3D statistical skinned multi-person linear model to estimate the 3D lower-limb joint centers trajectories and to improve the robustness of the sagittal lower-limb kinematics on individuals affected by CP and foot deformities.

2.4.1 Materials and methods

2.4.1.1 Method description

Subjects – Six participants with cerebral palsy from the Swedish CPUP and four with clubfeet were acquired.

The subject preparation is the same as described in Chapter 2 "Concurrent validation of a 2D model-based approach on individuals with Cerebral Palsy based on a single RGB-D camera.", as is the camera positioning. The only difference is that the Azure Kinect camera was used (RGB images: 720 × 1280 pixels at 30 fps, Depth images: 640 × 576 pixels at 30 fps).

Data collection – to create a 3D subject-specific lower-limb model, one frontal, two lateral (right and left side) and one posterior views of the subject while standing upright were captured. To enhance the stability of the subject while maintaining a static posture, an expert operator was needed to assist the patient in case of difficulty. Participants were then instructed to walk at a comfortable, self-selected speed along a straight 10-meter walkway. Six gait trials were recorded for each subject, capturing three complete gait cycles for both the right and left sides.

For validation purposes, the same procedure described in Chapter 2 “Concurrent validation of a 2D model-based approach on individuals with Cerebral Palsy based on a single RGB-D camera.” was used.

However, the acquisitions were not performed synchronously by the MB and MS systems as interferences in the depth map reconstruction were observed in the Azure Kinect recordings. The wavelength of the Azure Kinect IR sensor is the same as the Qualisys system (850 nm) and this resulted in very poor quality depth images with many invalidated pixels.

For this reason, the same trial was repeated twice to be acquired separately with the two systems (MS and MB) under the hypothesis of repeatability of the gesture.

1) Multi-segmental model definition

Basic Linear Blend Skinning models (Loper et al., 2015) are the most common technique for animating 3D characters and are supported by all game engines due to their rendering efficiency. This method relies on a linear combination of bone transformations that influence each vertex of the mesh. The popularity of these models stems from its computational efficiency, allowing complex animations to be rendered in real-time without compromising system performance, making it an ideal choice for interactive applications such as video games. However, these models often result in unrealistic deformations at joints, including the well-known "taffy" and "bowtie" effects (**Figure 24**). To address this problem, the Skinned Multi-Person Linear (SMPL) (Loper et al., 2015) model was introduced by the Max Planck Institute in 2014. SMPL is a realistic statistical 3D model of the human body that combines skinning and blend shapes, and it has been developed from thousands of 3D body scans.

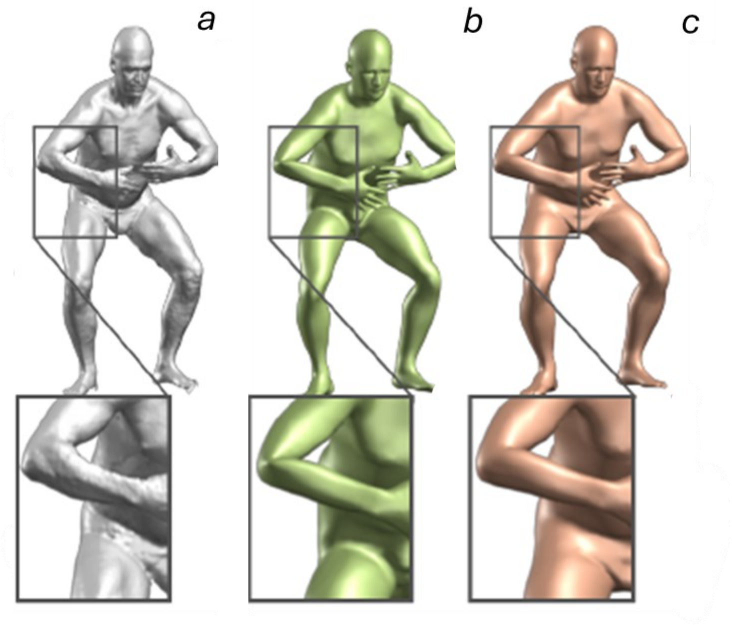


Figure 24. a) 3D scan of the human body, b) Linear Blend Skinning model results in artifacts at the right elbow, c) SMPL model is able to reconstruct more realist movements.

Moreover, Max Plank institute has provided a detailed website including a good documentation together with all the model's parameters useful to use this model on several software and for different applications which represents a great added value for open-science research purposes.

The SMPL model employs a vertex-based skinning approach that begins with a mean shape. To this shape, a vector of concatenated vertex offsets is applied to achieve specific body shapes and poses. The mesh maintains the same topology for both male and female. It is designed with a clean quad structure, segmented into different parts, equipped with initial blend weights, and supported by a skeletal rig. For the sake of simplicity of notation, meshes and shapes are vectors of vertices represented by bold capital letters (e.g. $\boldsymbol{X}$) and lowercase bold letters (e.g. $\boldsymbol{x_i}$) are vectors representing a particular vertex.

The model is defined by a mean template shape represented by a vector of N = 6890 concatenated vertices. The pose of the body is defined by a standard skeletal rig composed by 23 joints, hence a pose $\boldsymbol{\theta}$ is a vector of 72 parameters where 69 parameters (23*3) are the relative rotation of each *k-th* part with respect to its parent in the kinematic tree around the three axes and the remaining 3 parameters are the root orientation. The statistical aspect of this model lies in the fact that it uses a statistical data-driven approach to represent the variety of human forms and poses.

In particular, this 3D model is described by two types of parameters (**Figure 25**):

- **β** is a shape vector of scalar values which could contain from 10 to 300 values. This type of parameter could be interpreted as an amount of expansion or shrink of a human subject;
- **θ** is a pose matrix of 24*3 scalar values in terms of Euler angles that describes the relative rotation of each joint with respect to the rest pose.

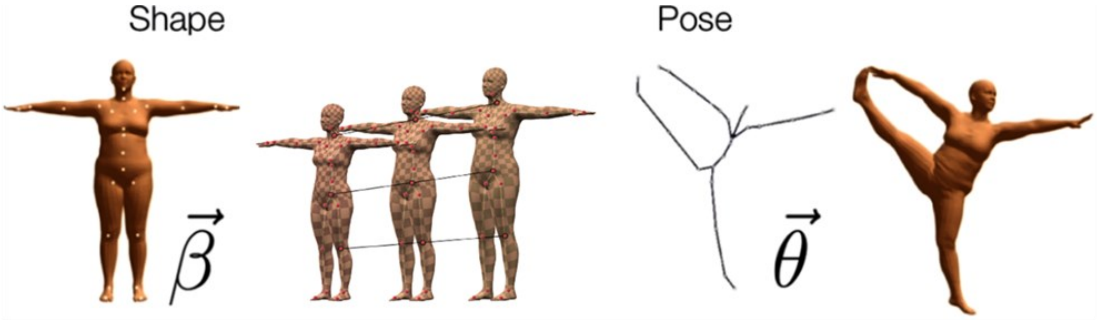


Figure 25. SMPL model. On the left, a representation of the functionality of beta parameter. On the right, an illustration of the variation of the pose parameter.

To change the shape and pose of the model using **β** and **θ**, the following functions were proposed:

1. Blend shape function: the body shapes of different people are represented by a linear function B_s

$$B_s(\boldsymbol{\beta}; \boldsymbol{S}) = \sum_{n=1}^{|\beta|} \beta_n \, \mathbf{S_n}$$

Where $|\beta|$ 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.

2. Pose blend shape function: the vertex deviations from the rest template, ***T,*** are described by the following function:

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

Where K = 23 represents the number of joints of the SMPL model. $\mathbf{P_n}$ is a matrix of vertex displacements which are applied to the vertices following the rotation of a joint, $\boldsymbol{R_n}(\boldsymbol{\theta})$ is a vector of length 23*9 = 207 representing a function which maps a pose vector $\boldsymbol{\theta}$ to a vector of concatenated part relative rotation matrices. $\boldsymbol{\theta}^*$ represents the rest pose.

The coordinates of the joint centers, J, are derived from the template $\boldsymbol{T}$ to which the function B_S has been applied. Using the stable mesh topology of the SMPL model, the position of each joint location could be estimated as a weighted average of surrounding vertices. This average is represented by a joint regression matrix, $\boldsymbol{J}$, provided by Max Plank Institute (learned from the training set) that defines a sparse set of vertex weights for each joint (**Figure 26**).

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

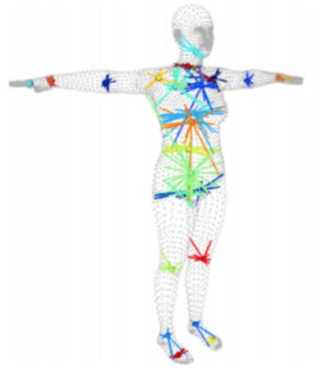


Figure 26. Estimation of joint centers positions. The left knee joint is computed as the weighted average of red vertices.

Application of transformations: During animation, transformations (such as translation, rotation, and scaling) of the skeleton segments are applied to the vertices of the SMPL model mesh. However, instead of applying the transformations directly to the segments, they are combined using blend weights. This means that the mesh vertices are influenced by multiple transformations simultaneously, with the intensity of each influence determined by the corresponding blend weights.

During the dynamic processing, the pose of the model could change accordingly:

$$\boldsymbol{t}_{\boldsymbol{i}}' = \sum_{k=1}^{K} w_{k.i}\, \boldsymbol{G}_{\boldsymbol{k}}'(\boldsymbol{\theta},\boldsymbol{J})\, \boldsymbol{t}_{\boldsymbol{i}}$$

Where $\boldsymbol{G}_{\boldsymbol{k}}'(\boldsymbol{\theta},\boldsymbol{J})$ is the local 3x3 rotation matrix around each *k-th* joint of rotation angles, $\boldsymbol{\theta}$. For each vertex of the SMPL model mesh, blend weights, $w_{k.i}$, are defined to indicate how much each segment influences the deformation of the corresponding vertex. These blend weights are assigned during the model training phase.

Finally, each vertex, $\boldsymbol{t}_{\boldsymbol{i}}'$, of the final model is transformed as:

$$\boldsymbol{t}_{\boldsymbol{i}}' = \sum_{k=1}^{K} w_{k.i}\, \boldsymbol{G}_{\boldsymbol{k}}'(\boldsymbol{\theta},\mathrm{J}(\boldsymbol{\beta};\boldsymbol{J},\boldsymbol{T},\boldsymbol{S}))\, \boldsymbol{t}_{\boldsymbol{P},\boldsymbol{i}}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{T},\boldsymbol{S},\boldsymbol{P})$$

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

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.

In other words, the blend shape function $\boldsymbol{S} * \boldsymbol{\beta}$ and the pose blend shape function $\boldsymbol{R}(\boldsymbol{\theta}) * \boldsymbol{P}$ were applied to the mean shape $\boldsymbol{T}$ to modify the shape of the model. The pose of the model was changed by applying $\boldsymbol{W} * \boldsymbol{G}(\boldsymbol{\theta})$.

The training of the model was carried out separately for the pose parameters and the shape parameters as shown in **Figure 27**. For the pose parameters, a multi-pose dataset consisting of 1786 recordings from 40 individuals was used. The parameters $\boldsymbol{W}$ and $\boldsymbol{P}$ were refined through gradient optimization by reducing the discrepancy between the 3D model and the edges of the recorded 3D scans. Conversely, the mean template shape, $\boldsymbol{T}$, the regressor matrix, $\boldsymbol{J}$, and blend weights for the shape parameters were automatically derived from the multi-shape dataset consists of registrations to the CAESAR dataset, totaling 1700 registrations for males and 2100 for females. This process utilized principal component analysis to maximize the explained variance of vertex offsets in the rest pose, using a constrained number of shape directions (**Figure 27**).

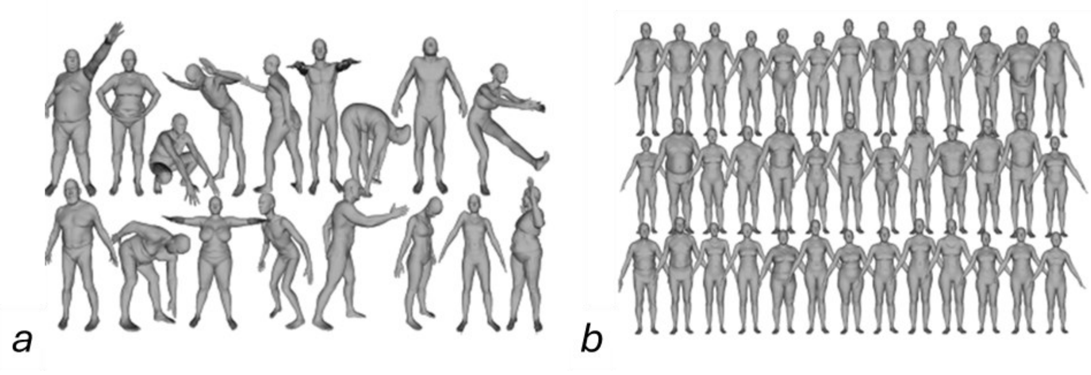


Figure 27. a) Multi-pose dataset, 1786 high-resolution 3D scans of various subjects in different poses from the CAESAR dataset, b) Multi-shape dataset, 3800 high-resolution 3D scans of various subjects in different shapes from the CAESAR dataset.

Figure 28 shows the block diagram of the 3D MS protocol.

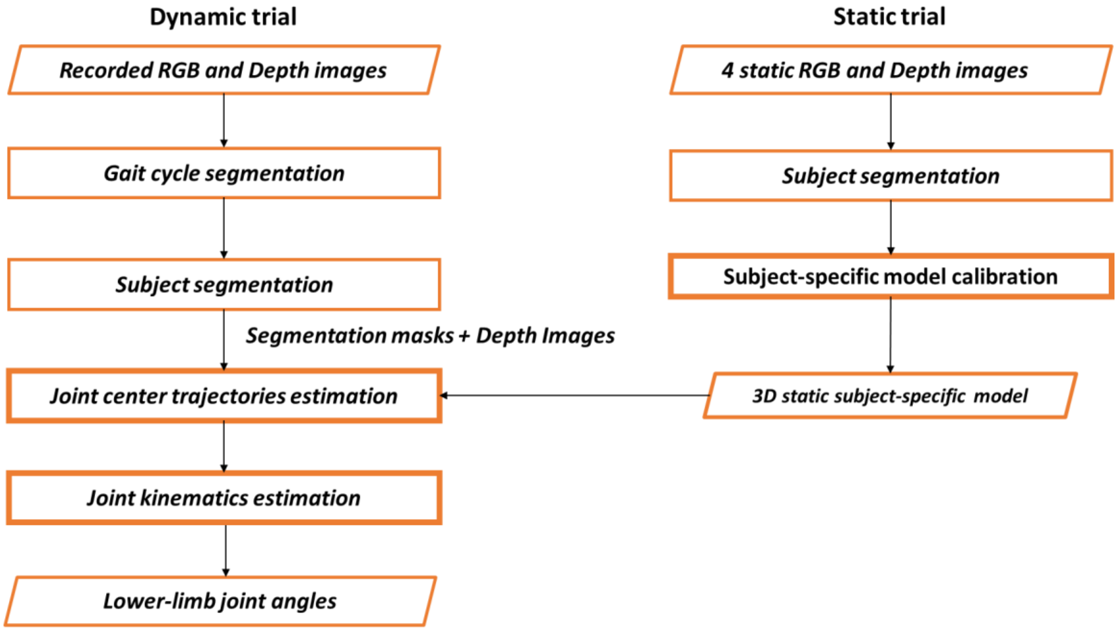

Dynamic trial
Recorded RGB and Depth images
Gait cycle segmentation
Subject segmentation
Segmentation masks + Depth Images
Joint center trajectories estimation
Joint kinematics estimation
Lower-limb joint angles
Static trial
4 static RGB and Depth images
Subject segmentation
Subject-specific model calibration
3D static subject-specific model

Figure 28. Block diagram of the proposed 3D markerless protocol.

2) Subject-specific model calibration

For gait analysis purposes, a generic SMPL lower-limb model embedding foot, shank, thigh and pelvis interconnected by joints at the ankle, ${}^{I}\boldsymbol{AJC}$, knee, ${}^{I}\boldsymbol{KJC}$, and hip, ${}^{I}\boldsymbol{HJC}$ was used. The key idea of the proposed MS protocol was to adapt the mean shape of the SMPL lower limb model, $\boldsymbol{T}$, to the 3D static posture of the subject in order to create a 3D static subject specific model, $\boldsymbol{SMPL_{subject-specific}}$, as follows:

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

Where 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 of $\boldsymbol{T}$ was adapted to the subject's pose by adjusting the pose parameters, $\boldsymbol{\theta_{ss}}$. Additionally, a displacement can be added to each point of the model to further refine its shape and alignment with the observed subject's data. By combining these steps, a static subject-specific model, $\boldsymbol{SMPL_{subject-specific}}$, based on $\boldsymbol{T}$ can be reconstructed. In addition, thanks to the fixed mesh topology of $\boldsymbol{T}$, each joint location, J, can be estimated as a weighted average of surrounding vertices.

First of all, to create the 3D static posture of the subject, four static acquisitions of the subject standing upright were conducted by using single RGB-D camera: frontal, posterior and two laterals (left and right) views of the subject as shown in **Figure 29**.

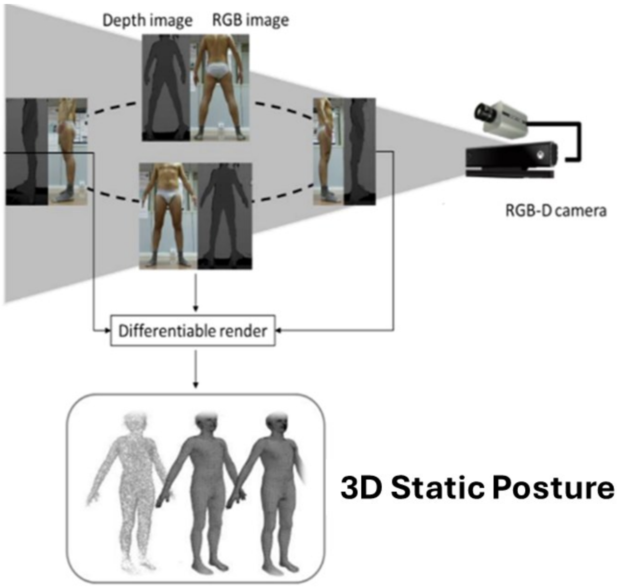


Figure 29. Four static views acquired to create a 3D static posture using a single RGB-D camera a) frontal view, b) right sagittal view, c) left sagittal view and d) posterior view of the subject.

For each RGB frame of the four static acquisitions, the subject was identified as described in (Balta et al., 2023). The subject's lower limbs were isolated from the segmentation masks. The depth values belonging to the subject's lower limbs were then converted in the 3D coordinate system in order to generate a point cloud representing the subject's lower limbs as described in Chapter 1, paragraph "3D point cloud reconstruction". All the following steps will be implemented on the reconstructed point clouds.

The identification of the feet was performed by leveraging color information on each RGB image (Cheng et al., 2001). Then, the foot segmentation mask was overlapped to the depth image to identify the corresponding depth values. The depth values belonging to the feet of the subject were then converted in the 3D coordinate system in order to generate a point cloud representing the subject's foot as described in Chapter 1, paragraph "3D point cloud reconstruction".

A 3D static posture of the subject was created by merging the four different point clouds, previously generated, by aligning four common points, for each side, automatically identified (i.e. T, H, M1, M2 in **Figure 30**) among the point clouds. M1 and M2 were identified as the intersection between the segment fitting the upper portion of the pelvis and the segment fitting the upper portion of the thigh. Heel and toe were identified as described in “Gait cycle identification.”.

The right and left side were treated separately.

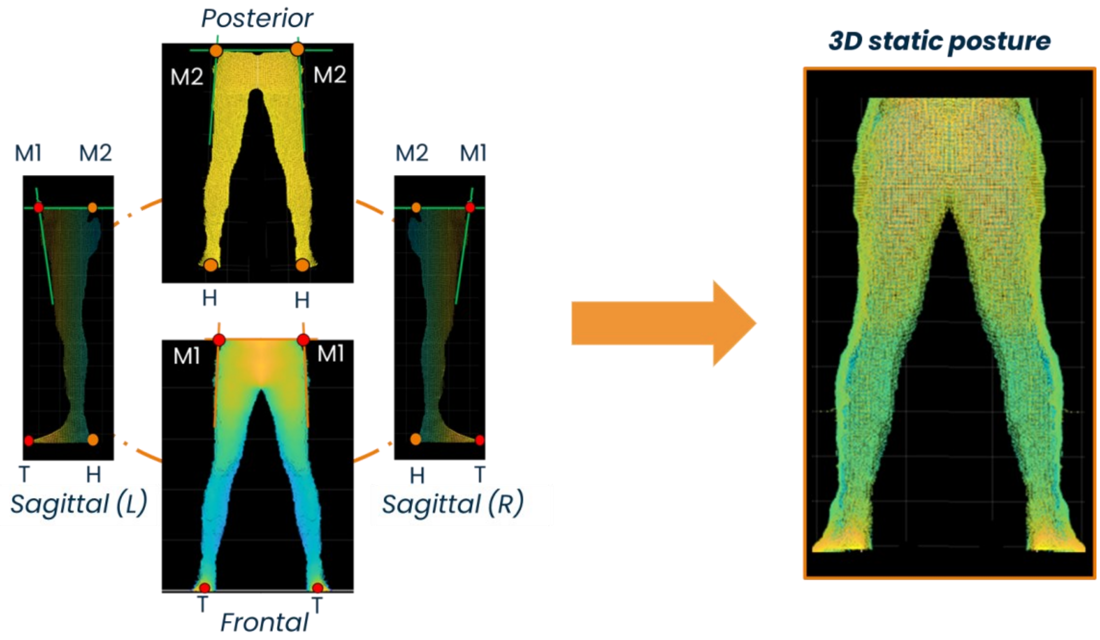


Figure 30. Creation of a 3D static posture by aligning four common points among different views.

First of all, the centroids of $\boldsymbol{T}$ and 3D static posture were estimated and aligned.

The scaling factor, c_{ss}, was computed as the ratio between the length of $\boldsymbol{T}$ and the length of the 3D static posture.

In order to compute an initial estimate of the pose of subject’s lower limb, represented by $\boldsymbol{\theta}_{ss}$, the Articulated Iterative Closest Point (AICP) was implemented.

- *Articulated iterative closest point algorithm*

AICP (Pellegrini et al., 2008) is an extension of the traditional ICP algorithm, tailored to manage articulated objects like human bodies or robotic arms. ICP is a popular algorithm for registering 3D point sets, aligning points from a source surface to a target surface by minimizing the distance between them.

In the context of articulated ICP, the algorithm adjusts to the object's articulated nature. It accounts for the constraints and relationships between the object's various segments. By integrating information about the joints, articulated ICP can precisely align the surfaces of articulated objects, even among complex deformations or movements.

An articulated body model *M* is composed of rigid parts $\{p_1; p_2; \ldots p_{NP}\}$. 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 j_r, which has no parent, is connected to the world.

A spherical generic joint ji 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:

$$\boldsymbol{G_i^W(\theta_i)} = \boldsymbol{G_r^W} * \ldots * \boldsymbol{G}\,_{\boldsymbol{i}}(\boldsymbol{\theta_i})$$

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$\{\boldsymbol{m_1^{pi}} \ldots \boldsymbol{m_{Ni}^{pi}}\}$ such as the position of a point $\boldsymbol{m_j^{pi}}$ in world coordinate system is $\boldsymbol{G_i^W}\ \boldsymbol{m_j^{pi}}$.

The core concept of AICP involves segmenting the articulated body into individual parts that can be aligned to the target surface (denoted as $\boldsymbol{D}$) in the same manner as the original ICP, but with added constraints to maintain the integrity of the articulated structure. Similar to the original ICP, for each point on the model, the closest points on the data surface are selected. This method allows for precise alignment while respecting the mechanical constraints of the articulated body.

In particular, 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 $\boldsymbol{M_b^r} = \boldsymbol{M_o^r}$, i.e., both branches correspond to the entire model.

Pellegrini and colleagues proposed that, instead of estimating all pose parameters using an iterative local optimizer, attention be restricted to 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 split into the base $\boldsymbol{M_b^i}$ and the outer branch $\boldsymbol{M_o^i}$ as defined above, and only the outer branch was aligned with a rigid transformation which respects the joint constraints at j_i. The alignment between the model and the dataset was computed by minimizing the norm of the difference, $E_o(\boldsymbol{\theta_i})$, between $\boldsymbol{M_o^i}$ and its closest points of a given dataset, $\boldsymbol{d_s}$, as follows:

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

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 $\boldsymbol{d_s}$.

As suggested by (Pellegrini et al., 2008), the choice of joint for executing the splitting can be made randomly, in a distributed manner, or cyclically. Specifically, the AICP algorithm involves a repetitive procedure where, in each cycle, it computes the rigid transformation parameters (rotation and translation) that optimally align the points between the model and the dataset surfaces, adhering to the joint constraints at j_i . This process is continued until convergence is achieved, that is, when the total alignment error is minimized.

In the proposed 3D MS protocol for gait analysis, the root p_r was the hip, the origin of the world-reference system I has been coincided with the hip and the cyclical joint selection was implemented. More in details, a top-down approach was implemented starting from the hip to the ankle. Both hip, knee and ankle were considered as 3 DoF's spherical joints.

In particular,

$$\boldsymbol{\theta}_{\boldsymbol{hip}} = \{\alpha_{hip}, \beta_{hip}, \gamma_{hip}\}$$

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

$$\boldsymbol{\theta}_{\boldsymbol{knee}} = \{\alpha_{knee}, \beta_{knee}, \gamma_{knee}\}$$

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

$$\boldsymbol{\theta}_{\boldsymbol{ankle}} = \{\alpha_{ankle}, \beta_{ankle}, \gamma_{ankle}\}$$

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

Finally,

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

A cyclical joint selection was implemented for splitting $\boldsymbol{T}$.

First of all, the hip joint was selected and the closest points of the 3D static posture to the $\boldsymbol{T}$ were identified in order to apply a rigid alignment between them around the selected joint. Since the hip corresponds to the root of $\boldsymbol{T}$, a rigid transformation was applied to the entire $\boldsymbol{T}$ (thigh and shank and foot), $\boldsymbol{M}_{\boldsymbol{o}}^{\boldsymbol{hip}}$. The rigid alignment was computed by minimizing the norm of the difference, $E_o(\boldsymbol{\theta}_{\boldsymbol{hip}})$, between $\boldsymbol{M}_{\boldsymbol{o}}^{\boldsymbol{hip}}$ and its closest points of the 3D static posture, $\boldsymbol{d}_{\boldsymbol{s}}$:

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

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

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

$$E_o(\theta_{knee}) = \sum\nolimits_{p_k \epsilon\, \boldsymbol{M_o^{knee}}} \sum_{j=1}^{N_{pk}} min \left\| \boldsymbol{G_{p_k}^W}(\boldsymbol{\theta_{knee}})\, \boldsymbol{m_j^{p_k}} - d_s \right\|^2$$

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

This process was repeated for a certain number of iterations until convergence is reached. After having reached the convergence, the resulting $\boldsymbol{G_k^W}(\boldsymbol{\theta_{hip}})$ and $\boldsymbol{G_k^W}(\boldsymbol{\theta_{knee}})$ represent the rigid transformation around the hip and the knee joints which allows the best alignment between the $\boldsymbol{T}$ and the 3D static posture.

The search limits for the AICP implementation were established based on physiological limits and to ensure an admissible configuration.

The internal-external rotation of hip (γ_{hip}) and knee (γ_{knee}) during the AICP implementation was blocked since, for the static acquisitions of the subjects, it turned out to be negligible. The subjects were asked to assume a static position with parallel feet and extended knee to the best of their ability. An expert operator was required to support the patient in case of difficulty in maintaining such a position during the acquisition. However, in the case of patients with more severe deformities, it could be necessary to estimate the angle of internal-external rotation. Moreover, at this stage, the AICP implementation allows to ensure an initial estimate of the subject pose, $\boldsymbol{\theta_{ss}}$, that will be refined through the estimation of the shape factor.

To speed up the alignment process, the foot pose, $\boldsymbol{\theta_{ankle}}$, was determined by leveraging the previously identified positions of the heel (H) and toe (T) and not using the AICP algorithm. Specifically, the plantar-dorsiflexion, α_{ankle}, and internal-external rotation of the foot, β_{ankle}, in a 3D static posture were calculated using simple trigonometric formulas.

Finally, using the above-explained procedure, the pose parameter $\boldsymbol{\theta_{ss}}$ was computed based on the 3D posture of the subject.

Then, the residual discrepancy between the mean shape, $\boldsymbol{T}$, and the 3D static posture of the subject was minimized by computing $\boldsymbol{\beta_{ss}}$ as follows:

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

Final results for the creation of a 3D $\boldsymbol{SMPL_{subject-specific}}$ are shown in **Figure 31**.

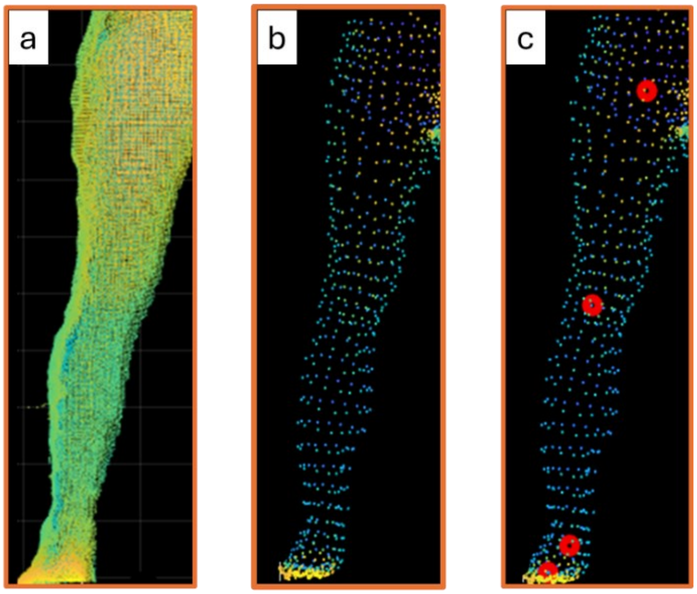


Figure 31. 3D subject-specific model creation. a) 3D static posture, b) 3D subject-specific model c) 3D subject-specific model with their relevant joint centers.

3) Joint centers trajectories estimation

As in Balta et al., 2023, the joint centers trajectories were tracked by fitting the $\boldsymbol{SMPL_{subject-specific}}$ on each dynamic point cloud of the gait cycle. Unlike Balta et al., 2023, who introduced a 2D lower-limb model and performed separate 2D fittings for the thigh, shank, and foot, the proposed 3D MS protocol utilizes a 3D articulated lower-limb model. This model was fitted on each dynamic point cloud of the gait cycle using the AICP.

First of all, for each frame of the gait cycle, the same procedure proposed for the static processing was used to create the point cloud of the lower limbs. Then, the Otsu method was applied (Otsu, 1979) to separate the foreground limb from the background one and the hand during arm oscillation, following the approach suggested by (Balta et al., 2023). Thanks to this procedure, only the foreground lower limb, $\boldsymbol{Dyn}$, was isolated and took in consideration for the next steps.

The presence of the hand during arm oscillation, when it partially overlaps with the thigh, is a critical factor that can affect the effectiveness of the AICP. Specifically, the hand's presence could lead to local minima in the performance of the AICP, if not properly addressed. Therefore, depending on its relative position to the thigh, two different strategies were implemented:

- A segment was fitted to the contour of the thigh to exclude only the portion of the hand that extended beyond the thigh **Figure 32**a.
- If the hand completely overlapped the thigh, it was left in its original position.

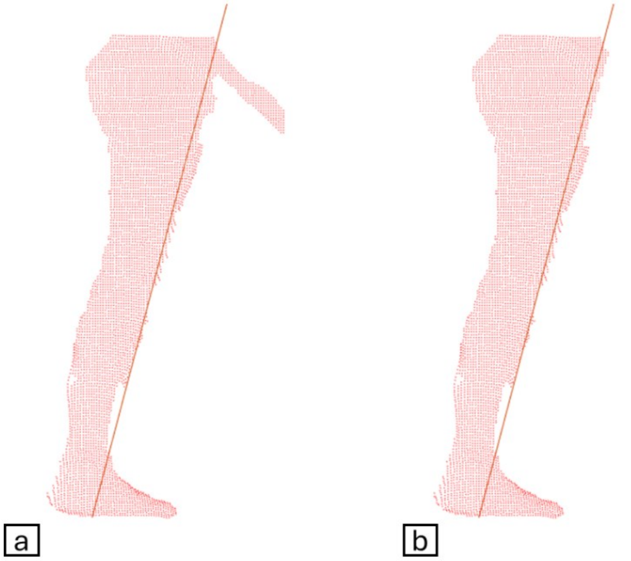


Figure 32. Hand oscillation during the gait cycle. a) Lower-limb point cloud with the hand partially overlapped with the thigh b) Lower-limb point cloud after the hand removal.

Initial conditions for the implementation of the AICP algorithm

It must be emphasized that during the recorded gait cycle, the size and shape of the lower limb segments can change due to soft tissue deformation or variations in the distance between the subject and the camera. This could introduce inaccuracies in the matching procedure due to discrepancies between the size of the model and the size of the lower limb in dynamic. To overcome that problem, the scaling factor of the $\boldsymbol{SMPL_{subject-specific}}$ was properly set according to the size of the lower limb, $\boldsymbol{Dyn}$, in each dynamic frame.

Selecting initial conditions for the AICP algorithm implementation is not a straightforward task, as these conditions significantly impact on its performance. First of all, the most anterior and upper point of $\boldsymbol{SMPL_{subject-specific}}$ was made to coincide to the one of $\boldsymbol{Dyn}$.

Also for the dynamic processing, heel and toe were identified in each RGB image of the gait cycle as in (Balta et al., 2023) and then their 2D coordinates were converted in 3D in order to obtain ${}^{I}\boldsymbol{HEEL}$ and ${}^{I}\boldsymbol{TOE}$ referring to the image reference system I as explained in Chapter 1 paragraph “3D point cloud reconstruction”. The foot pose, θ_{ankle} , was determined by leveraging the previously identified positions of $^{I}\boldsymbol{HEEL}$ and $^{I}\boldsymbol{TOE}$. Specifically, the plantar-dorsiflexion, α_{ankle}, and intra-extra rotation angle of the foot, β_{ankle} were calculated using simple trigonometric formulas.

Two different strategies were implemented based on the phase of the gait cycle:

- For the first frame in the gait cycle, the $\boldsymbol{SMPL_{subject-specific}}$ was positioned in a neutral stance position, meaning both $\boldsymbol{\theta_{hip}}$ and $\boldsymbol{\theta_{knee}}$ were set to zero;
- 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. In order to prevent the possibility of local minima due to incorrect alignments in previous frames, the upper and lower limits were kept consistent with those specified in **Table 5.**

Also, for the dynamic processing, a top-bottom approach was implemented starting from the hip to the knee as shown in **Figure 33**.

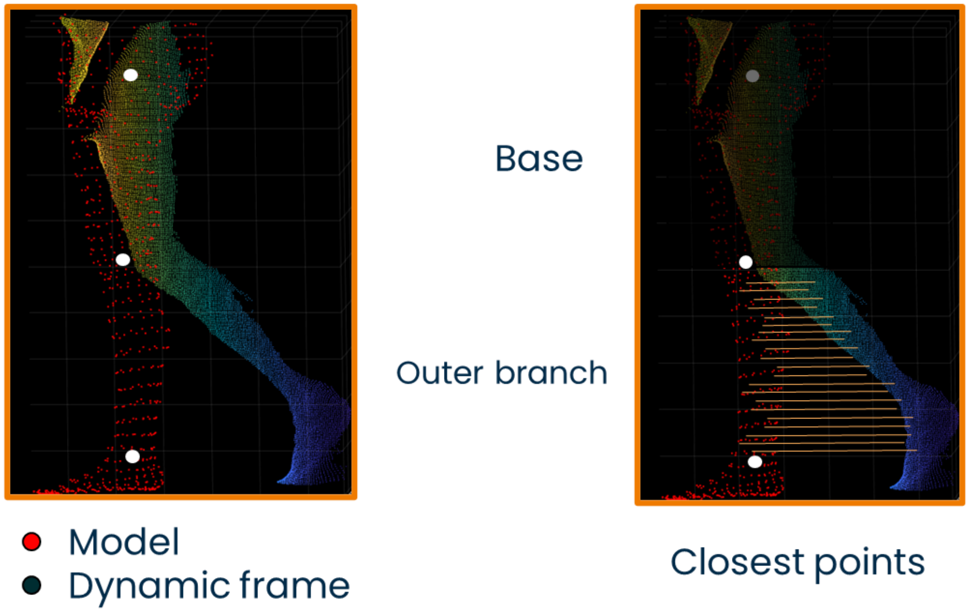


Figure 33. An example of AICP implementation for aligning the subject-specific model around the knee to a dynamic frame.

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 $\boldsymbol{SMPL_{subject-specific}}$ was split around it and the rigid transformation was applied.

The search limits for the AICP implementation, shown in **Table 5.**, were established based on physiological limits and to ensure an admissible configuration.

Table 5. Physiological joint limits for the implementation of AICP algorithm

Joint	*Angle*	*Upper limit*	*Lower limit*
Hip	Flexion-extension	$-\frac{\pi}{4}$	$\frac{\pi}{4}$
	Abduction-adduction	0	0
	Intra-extra rotation	0	0
Knee	Flexion-extension	$-\frac{\pi}{3}$	$\frac{\pi}{3}$
	Abduction-adduction	0	0
	Intra-extra rotation	0	0

Having acquired a sagittal view, the optimization process through the AICP algorithm was limited only to the hip and knee flexion-extension angles. The decision to limit the analysis to flexion and extension movements is justified by the fact that, although the RGB-D camera provides 3D coordinates, the view is restricted solely to the sagittal plane, leaving no reliable and complete information about external rotation movements and abduction-adduction movements. Furthermore, it is important to consider that the fitting process occurs between a 3D subject-specific model and 2D+ dynamic point cloud, thus limiting the selection of the closest points to a single sagittal view.

The AICP process was repeated for a certain number of iterations until the convergence was reached.

Convergence was reached when:

- the discrepancy between the error in consecutive iterations falls below 1%.
- the algorithm is constrained by a maximum of 9 iterations.
- the alignment accuracy is enhanced by incorporating a refinement mechanism based on knee joint position, ${}^{I}\boldsymbol{KJC}$, in the $\boldsymbol{SMPL_{subject-specific}}$. 3D points included between ${}^{I}\boldsymbol{KJC}$ and (${}^{I}\boldsymbol{KJC} \pm \epsilon$) were isolated in both $\boldsymbol{SMPL_{subject-specific}}$ and $\boldsymbol{Dyn}$. AICP alignment process stops when 30% of the points of $\boldsymbol{SMPL_{subject-specific}}$ exhibit an x-coordinate greater than those of $\boldsymbol{Dyn}$ indicating excessive bending beyond the limits allowed by $\boldsymbol{Dyn.}$

After having aligned the $\boldsymbol{SMPL_{subject-specific}}$ to $\boldsymbol{Dyn}$ for each frame, the final joint centers, ${}^{I}\boldsymbol{AJC}$, ${}^{I}\boldsymbol{KJC}$, ${}^{I}\boldsymbol{HJC}$ were estimated by applying the joint regression matrix to the aligned $\boldsymbol{SMPL_{subject-specific}}$ as depicted in **Figure 34**.

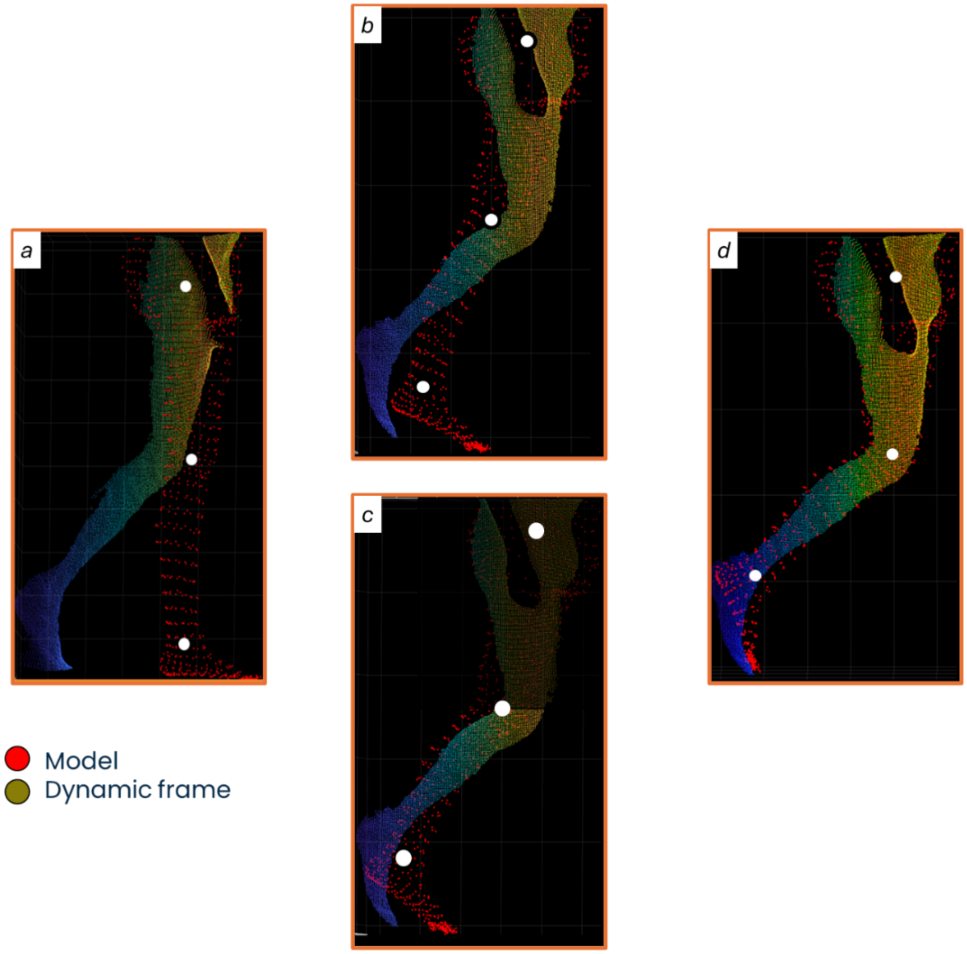


Figure 34. AICP implementation following a top-down approach from hip to ankle. a) subject specific model and a dynamic point cloud (initial condition), b) Rigid alignment around the hip joint, c) Rigid alignment around the knee joint, d) The final alignment between subject specific model and a dynamic point cloud. This process was repeated until the convergence is reached.

4) Joint kinematics estimation

Joint kinematics were determined by the inclination of segments, defined by the lines connecting the joint centers. The plantar-dorsi flexion angle of the ankle was calculated as the angle between the segment which best fits the foot centered in ${}^{I}\boldsymbol{AJC}$ 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.

2.4.2 Performance assessment and statistical analysis

Prior to comparison, the kinematic curves from both the MS and MB systems were processed using a fourth-order Butterworth filter with a cutoff frequency of 7 Hz and then time-normalized to the gait cycle (1-100%) (Bergamini et al., 2014).

For each key gait feature, k, of each gait trial, t, and subject, s, the mean absolute difference (MAD) and mean difference (MD) with respect to the MB estimates along with 95% confidence intervals (95% CI) were computed as follows:

$$MD_{k,s,t} = MB_{k,s,t} - MS_{k,s,t}$$

$$MAD_{k,s,t} = \left|MB_{k,s,t} - MS_{k,s,t}\right|$$

The latter values were then averaged across trials and subjects:

$$MD_k = \frac{1}{N_S}\sum_{s=1}^{N_S}\frac{1}{N_T}\sum_{t=1}^{N_T} MD_{k,s,t}$$

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

Where N_T = 6 is the number of trials and N_S = 10 is the number of subjects.

For each gait feature, each method (MS and MB) and each side, the reliability was evaluated with intraclass correlation based on absolute agreement and 2 way random effects (ICC*(2,k)*) computed based on the data collected over subjects (n = 10) for the different gait cycles (k = 3) (Koo & Li, 2016).

ICC values below 0.5 signify poor reliability, values from 0.5 to 0.75 indicate moderate reliability, values ranging from 0.75 to 0.9 denote good reliability, and values above 0.90 represent excellent reliability (Koo & Li, 2016).

To evaluate the inter-trial variability of each method, each joint angles, k, and each subject, s, the gait variable standard deviation (GVSD) was computed as follows (Sangeux et al., 2016) :

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

Where $T = 100$ is the number of time samples and $N = 6$ is the number of trials.

The latter values were then averaged across subjects.

2.4.3 Results

Results related to the extracted key gait features in terms of MD, MAD and their 95% CI are summarized in **Table 6**. Results for ICC(*2, k*) for both MS and MB protocols, are reported in **Table 7**. Results for GVSD for both MS and MB protocols, are reported in **Table 8**.

Table 6. Mean difference (MD) and mean absolute difference (MAD) of the key gait features along with 95 % CI: 95% of confidence interval averaged over six trials per ten subjects with respect to MB protocol.

Gait Variables (deg)		***MAD (95% CI)***	***MD (95% CI)***
Knee	***Initial Contact***	2.9 [2.3, 3.6]	-1.8 [-2.7, -0.9]
	Load	4.3 [3.4, 5.3]	0.6 [-0.8, 2.1]
	Stance	4.5 [3.6, 5.3]	3.9 [2.9, 4.9]
	Swing	4.5 [3.6, 5.2]	-1.5 [-2.9, -0.1]
Ankle	***Stance***	3.9 [3.1, 4.7]	-3.6 [-2.7, -4.5]
	Swing	3.8 [3.1, 4.5]	2.8 [1.8, 3.7]
Hip	***Stance***	4.3 [3.4, 5.1]	4.1 [3.1, 5.1]

Table 7. Reliability of the markerless (MS) and marker-based system (MB) methods computed for each gait variable and each side.

Gait Variables (deg)		***Side***	***ICC MS***	***ICC MB***
Knee	***Initial Contact***	*R*	0.93	0.94
		L	0.80	0.93
	Load	*R*	0.82	0.86
		L	0.94	0.87
	Stance	*R*	0.85	0.91
		L	0.80	0.80
	Swing	*R*	0.92	0.94
		L	0.85	0.88
Ankle	***Stance***	*R*	0.80	0.75
		L	0.90	0.92
	Swing	*R*	0.80	0.95
		L	0.70	0.94
Hip	***Stance***	*R*	0.90	0.60
		L	0.80	0.90

Table 8: Gait variability standard deviation (GVSD) of the markerless (MS) and marker-based system (MB) methods averaged over six trials per ten subjects.

Joint	***GVSD MS (deg)***	***GVSD MB (deg)***
Knee	4.5 [3.8, 5.2]	3.6 [2.8, 4.3]
Ankle	3.0 [2.6, 3.5]	1.9 [1.4, 2.3]
Hip	2.7 [2.3, 3.1]	2.0 [1.5, 2.6]

An example of the normalized joint kinematics curves of one subject, averaged over trials is reported in **Figure 35.**

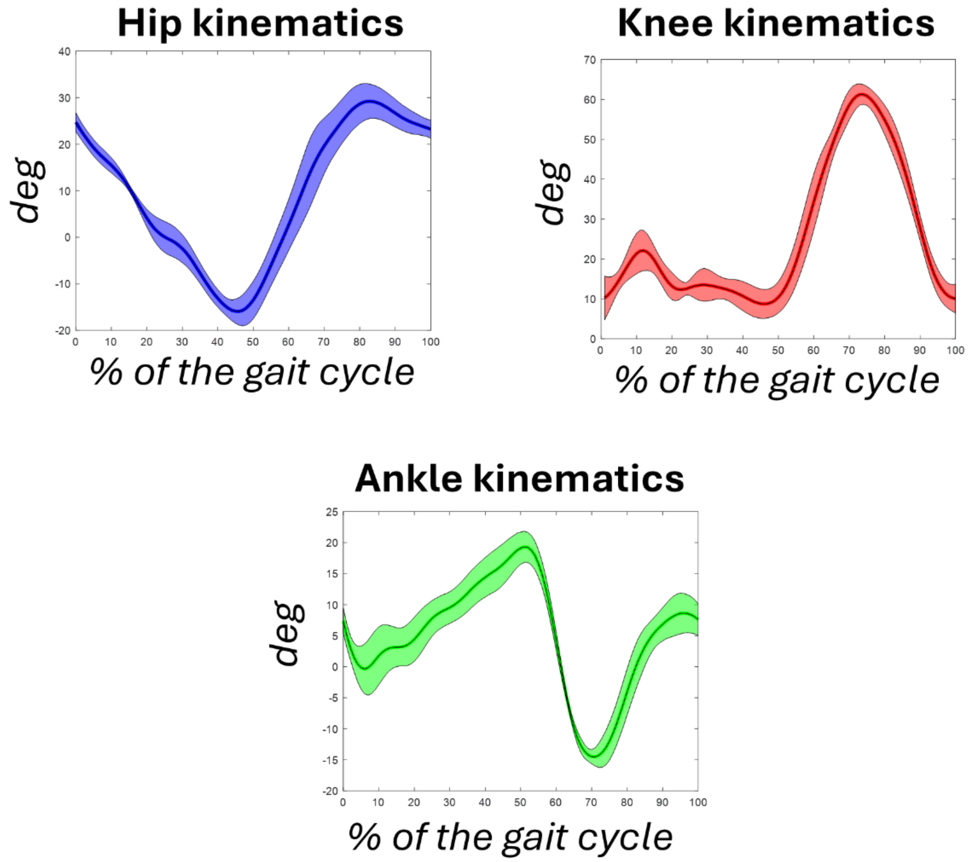


Figure 35. Sagittal lower limb joint kinematics of one subject extracted from the 3D MS protocol (hip, knee and ankle) averaged trials (average: dashed lines; SD: shaded area).

2.4.4 Discussions

For the knee, the MAD ranges from 2.9 deg to 4.5 deg. The lowest MAD of 2.9 deg occurs at initial contact, while the highest MAD of 4.5 deg occurs during the swing phase. For the ankle, the MAD ranges from 3.8 deg to 3.9 deg. For the hip, MAD reported is 4.3 deg during the stance phase. The AICP algorithm takes an average of 50 seconds to fit the 3D model to each dynamic frame (on average 30 minutes for each gait trial).

Clinically, errors between 2 deg and 5 deg are typically considered acceptable but they need careful interpretation (McGinley et al., 2009).

The differences between the proposed 3D MS protocol and MB system are partially due to the different protocols used. The MS protocol calculates joint angles using simple trigonometric formulas by connecting the joint centers, while MB system calculates the joint kinematics through the decomposition of Euler angles.

Residual errors in swing phase represented by K5 and A5 were due to technological limitations as the depth sensor failed in reconstructing depth values at highest velocity. This problem will be detailed in the paragraph “Factors influencing the accuracy of joint kinematics estimation”. Residual errors during the stance phase were 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. This issue will be detailed in the paragraph "Factors influencing the accuracy of 3D lower-limb model creation".

It is important to note that differences between MS and MB protocols are also due to the fact that the 3D method did not involve simultaneous acquisitions because of infrared interference between IR sensors.

Despite those issues, the ICC values for the 3D MS protocol indicated high reliability (ICC>0.75) for all gait features. For the knee at initial contact, the average ICC for the MS system was 0.88, while for the MB system, it was 0.94. This reflects good reliability for both methods. During the load phase, the average ICC for the knee was 0.88 (MS) and 0.86 (MB). Both methods exhibited good reliability. In the stance phase, the average knee ICC was 0.84 (MS) and 0.83 (MB), demonstrating good reliability for both methods. For the swing phase, the average knee ICC was 0.90 (MS) and 0.90 (MB). Both methods showed excellent reliability for this phase. For the ankle during the stance phase, the average ICC was 0.86 (MS) and 0.94 (MB). The MS method demonstrated good reliability while the MB method showed excellent reliability. In the swing phase, the average ankle ICC was 0.75 (MS) and 0.95 (MB). For the hip during the stance phase, the average ICC was 0.85 (MS) and 0.75 (MB). Both methods showed good reliability for the hip in this phase. The lowest ICC value is observed in the ankle angle during the swing phase (A5) confirming that the quality of the depth images influenced the performance of the model fitting.

Despite inaccuracies due to the above-mentioned issues, the 3D MS protocol ensures a very high reliability (ICC ≥ 0.75) which are comparable to the gold standard

(MB system) demonstrating that the estimates of the proposed MS protocol are consistent. It is important to note that the GVSD for each joint angles of the MS system is on average higher than 1 deg with respect to the MB system.

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

It must be highlighted the importance of positioning the subject in the same position during the static acquisitions in order to correctly identify the common points among the three views (frontal, lateral and posterior). When reconstructing a 3D subject-specific lower limb model by combining those views, the accuracy can be affected by slight changes in the subject's positioning between captures. The subject might unintentionally alter the orientation of their feet or legs. Even if the common points are correctly identified as described in paragraph “Subject-specific model calibration”, the reconstruction appears distorted because the subject’s pose has not been reproduced consistently across the views (**Figure 36**).

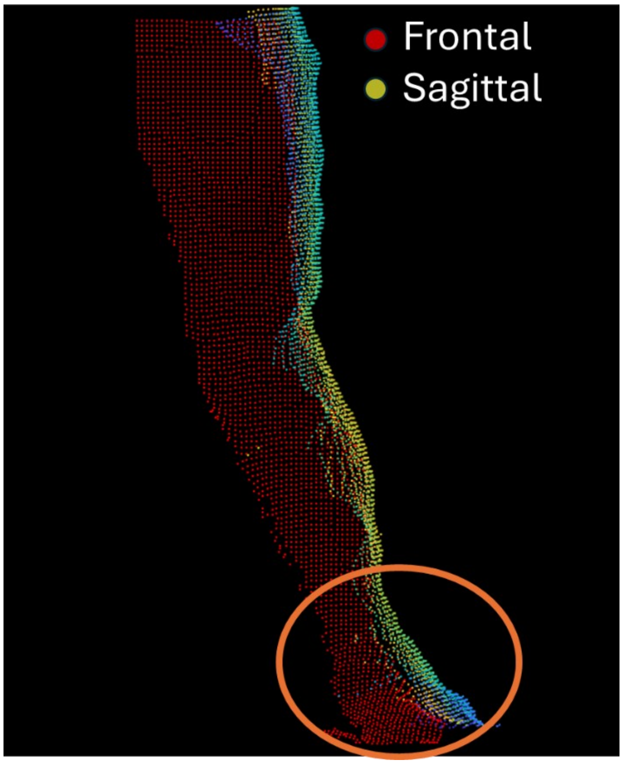


Figure 36. An example of incorrect 3D subject-specific reconstruction caused by a non-reproducible pose across the frontal view (red dot) and sagittal view (yellow dot) at the foot. The thigh and shank in both the frontal and sagittal views, however, are correctly merged.

To mitigate these issues, a proper standardization of the static acquisitions using specific guidelines through the use of a mat with footprints (**Figure 37**) will be included in future studies. This will help ensure that subject's acquisitions are consistently aligned across all views.

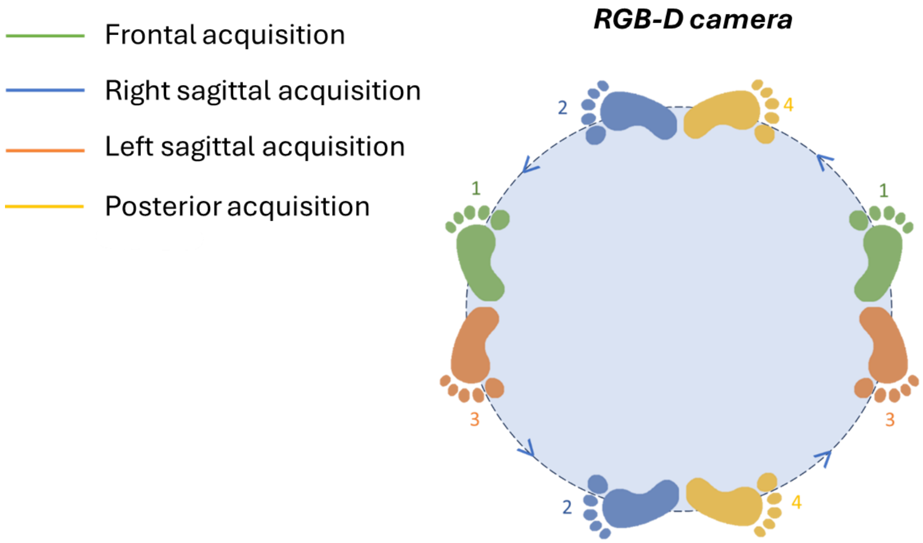


Figure 37. Footprint mat to aid patient in correct positioning during static acquisitions.

2.4.4.2 Factors influencing the accuracy of joint kinematics estimation

This paragraph will examine the sources of residual errors in sagittal lower-limb joint kinematics estimation. A primary factor influencing the accuracy of joint center estimations is associated to technological limitations of the depth sensor. During high-speed movements, it has been noticed that the depth sensor fails to accurately reconstruct depth values since the limited exposure time of RGB-D cameras can lead to motion blurs in captured images, potentially causing artifacts such as holes or fake boundaries (Gao et al., 2015), resulting in improper alignment between the depth image and the RGB image specifically at the shank and foot (**Figure 38**).

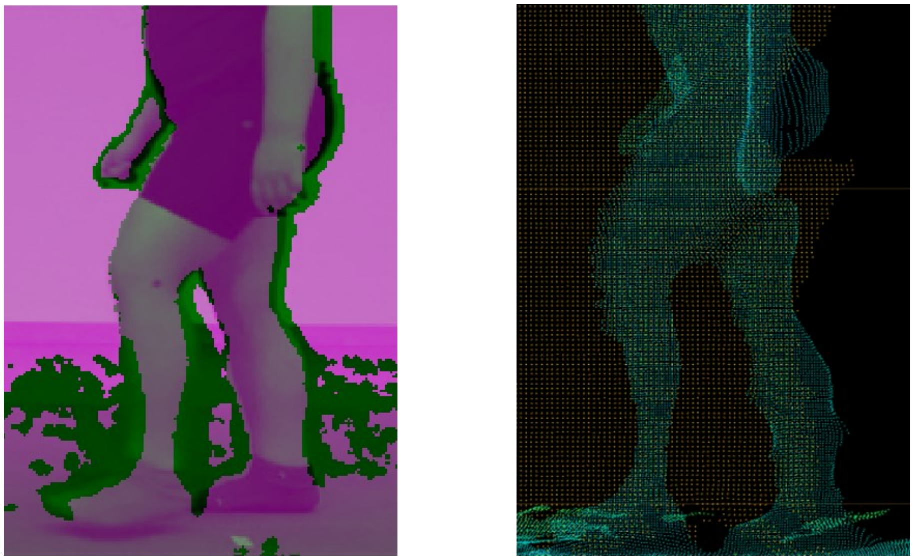


Figure 38. a) Depth map (pink pixels) overlapped on RGB images, b) Dynamic point cloud.

It is important to highlight that the differences between the MS and MB protocols are partly due to the fact that the 3D method did not involve simultaneous acquisitions, as infrared interference between IR sensors prevented this.

To the best of the authors' knowledge, the only MS study involving children with CP was conducted by (Ma et al., 2019). This study assessed the concurrent validity of the Kinect v2 body tracking SDK compared to a stereophotogrammetric system using a frontal view on 10 children with CP (GMFS I-II). The results indicated lower repeatability with respect to the proposed method, with ICC values of 0.8 versus 0.9 for hip kinematics, 0.5 versus 0.9 for knee kinematics, and 0.3 versus 0.8 for ankle kinematics.

(Yeung et al., 2021) explored the impact of five different camera viewing angles on kinematic curve estimates in healthy subjects using the Azure Kinect body tracking SDK. They found that a frontal viewing angle provided superior performance. This study demonstrated comparable performance to our MS protocol for hip and knee kinematics but reported a higher error (RMSE of 10 degrees) for ankle dorsiflexion, compared to a MAE of 4.5 degrees in our protocol.

(Yamamoto et al., 2021) evaluated OpenPose's performance on healthy individuals, finding similar reliability for knee and hip kinematics, with ICC values ranging from 0.60 to 0.98. However, significantly lower reliability was noted for ankle angles, with an ICC of 0.1 for maximum dorsiflexion during the stance and swing phases (A3 and

A5, respectively). In contrast, our MS protocol achieved an ICC from 0.7 to 0.9 for the same measurements.

(Lin et al., 2023) proposed a gait assessment system using the open-source human posture detection algorithm Keypoints And Poses As Objects. This software was applied to both healthy individuals and children with hip dysplasia. The study reported a comparable repeatability with respect to our protocol for the hip kinematics in terms of average ICC values (0.8 vs 0.85) but a lower ICC for the knee kinematics (0.5 vs 0.86 in our method). Notably, this method does not include ankle kinematics computation.

(Hatamzadeh et al., 2024) proposed an AI-based pose estimation algorithms improved by a subject-specific geometric model using two RGB cameras positioned on the front and back of the subject. This method was validated against a stereophotogrammetric system obtaining similar performance with respect to our protocol but on healthy subjects and with a dual cameras set-up (RMSE = 3.42 deg vs MAE = 4.3 deg for the hip kinematics, RMSE = 6.14 deg vs MAE = 4.0 deg for the knee kinematics, RMSE = 7.25 deg vs MAE = 3.85 for the ankle kinematics).

2.5 Preliminary validation of a 3D model-based method for estimating the kinematics of a two-segment foot model using a single RGB-D camera

2.5.1 Clinical relevance and aim of the work

Clubfoot is a congenital condition that affects the development of foot and ankle, with a worldwide incidence of 1 on 1000 live births. It is more frequent on males than on females and it occurs unilateral in 50% of the cases and the right side is more often affected (Gurnett et al., 2008). Also called talipes equinovarus, this pathology is recognizable at birth: its rigidness makes it distinguishable from other positional foot anomalies. This pathology is not passively correctable and, if it is left untreated, it could provoke infections, foot and leg deformities, pain, and limits mobility (Dobbs & Gurnett, 2017).

The common gait features of a patient affected by clubfoot are:

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

In recent years, several companies have developed affordable RGB camera integrated with infrared depth sensor (RGB-Depth) and MS alternatives based on it have been recently proposed to overcome MB limitations. These alternatives (e.g. Azure Kinect body tracking SDK, OpenPose), model the foot as a single segment without articulating the metatarso-phalangeal joint kinematics, which is crucial to guarantee an affective load of the foot and correct progression (Allan et al., 2020). Van den Herrewegen et al., 2014 has proposed a 3D multi-segmental foot model reconstructed through a D3DScan4D (5 scanner units) composed by 4 segments (Shank, Calcaneus, Matatarsus and Hallux) which were manually selected on a static scan. Then, to track the segments in the dynamic scan, the segments from the static scan were aligned with each frame of the dynamic scan using the ICP fitting algorithm.

However, the high cost and the size associated with a scanner system limits its clinical applicability to laboratory settings.

Following the idea of creating a 3D multi-segmental foot of Van den Herrewegen et al., 2014, the aim of this study is to design a MS method based on a single RGB-Depth camera to estimate sagittal ankle and metatarso-phalangeal kinematics using a two-segment 3D foot model composed by two segments: Mid-Rear-foot with ankle joint and Forefoot with metatarso-phalangeal joint (MTP) and explore its clinical applicability on children with foot deformities.

This work has been developed into two parts. The first aims at developing an algorithm to create a 2-segment 3D foot model using a single RGB-D camera by merging four static views (Frontal, Medial, Lateral and Posterior) by aligning three common points on the foot sole identified on each view .Then, the resulting 3D model was calibrated in order to obtain a subject specific model and divided into two segments: Mid-Rear Foot and Forefoot. The second part aims at developing an original method to estimate the sagittal foot joint kinematics during a gait cycle by matching the foot model to each dynamic point cloud using the 3D ICP.

2.5.2 Materials and methods

2.5.2.1 Method description

The subject preparation is the same as described in Chapter 2 “Concurrent validation of a 2D model-based approach on individuals with Cerebral Palsy based on a single RGB-D camera.”, as is the camera positioning. The only difference is that the Azure Kinect camera was used (RGB images: 720 × 1280 pixels at 30 fps, Depth images: 640 × 576 pixels at 30 fps).

Data collection – to create the 3D foot model, four static views (frontal, posterior, medial and lateral) of both feet with the camera placed directly on the floor at a 0.6-meter from the walkway were recorded to reconstruct the 3D foot model.

In addition, to create the lower-limb model as proposed in (Balta et al., 2023), the same experimental protocol described in Chapter 2 “Concurrent validation of a 2D model-based approach on individuals with Cerebral Palsy based on a single RGB-D camera.” was used. Participants were then instructed to walk at a comfortable, self-selected speed along a straight 10-meter walkway. Six gait trials were recorded for each subject, capturing three complete gait cycles for both the right and left sides.

Validation – The validation of the markerless protocol against the stereophotogrammetric system could not be conducted due to interference problems observed in the depth images during recordings from the Qualysis system. The problem originates from both Azure Kinect IR sensor and the Qualysis system operating at the same 850 nm wavelength. This similarity led to a significant degradation in the quality of depth images, marked by a high number of invalid pixels. Moreover, black, non-informative pixels were commonly observed in areas containing reflective markers during joint recordings. This issue is especially critical for MB protocols such as Oxford (Stebbins et al., 2006), which are employed in studying foot kinematics and necessitate a dense placement of markers in a limited space. As illustrated in **Figure 39**, the depth information for the foreground leg and foot was heavily affected, thus adversely influencing the precision of joint center position estimates.

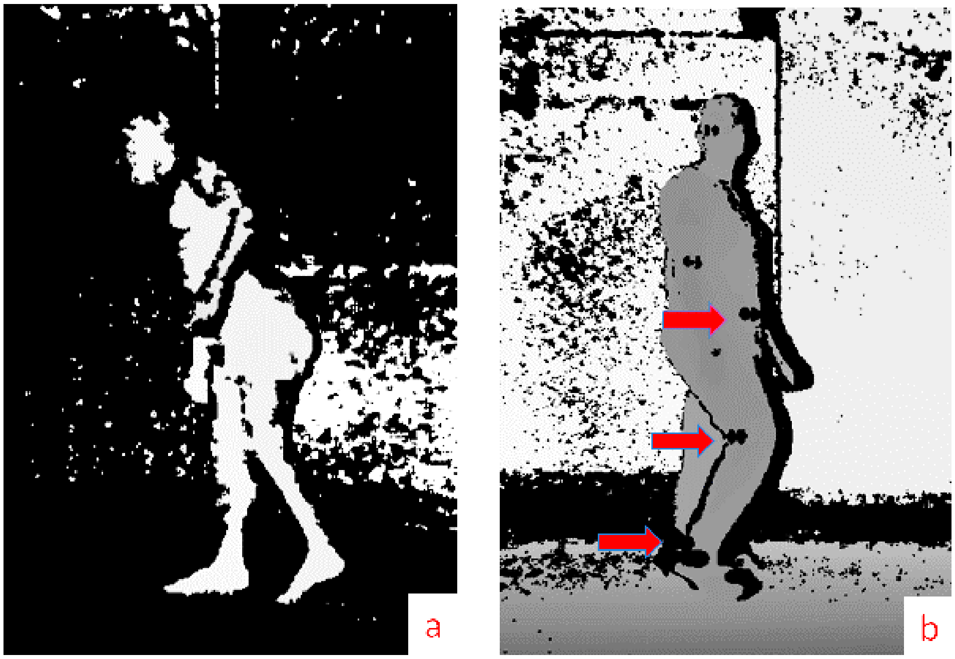


Figure 39. a) Illustration of the IR interference between Azure Kinect and Qualisys system. b) Black invalidated pixels in correspondence of the reflective markers (red arrows)

To avoid concerns about IR interference, the output of the MS protocol was validated against the joint centers positions manually labeled from an expert operator on the 2D RGB images. Sagittal shank and foot kinematics were computed as described in "Joint kinematics estimation".

Figure 40 shows the block diagram of the 3D MS protocol.

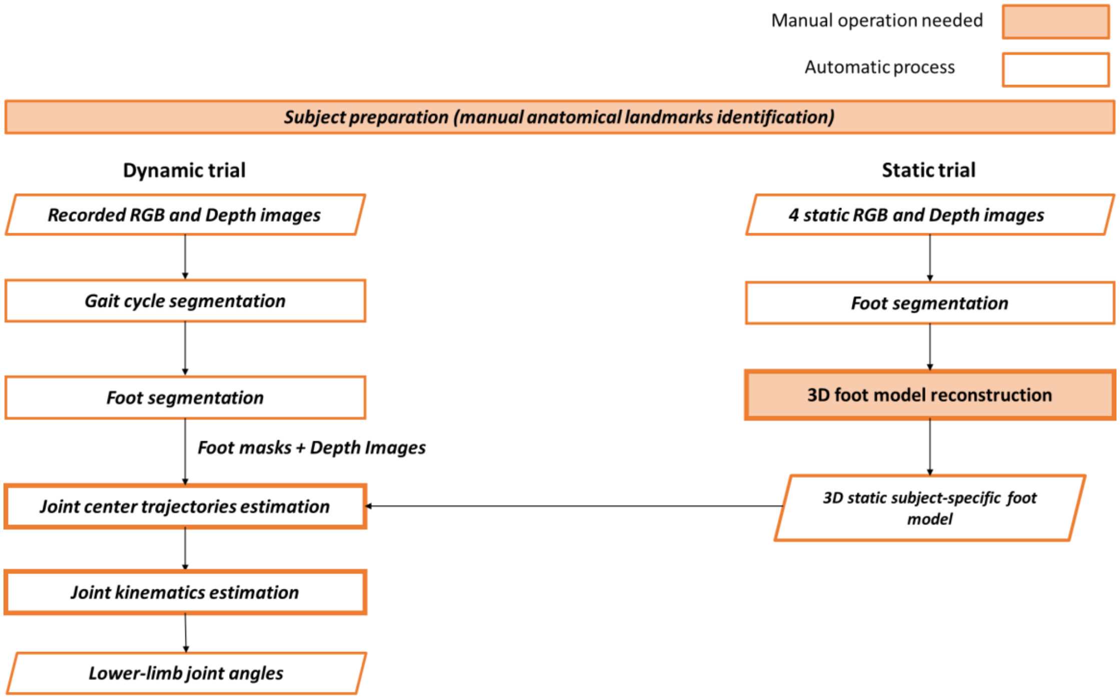

Manual operation needed
Automatic process
Subject preparation (manual anatomical landmarks identification)
Dynamic trial
Static trial
Recorded RGB and Depth images
4 static RGB and Depth images
Gait cycle segmentation
Foot segmentation
Foot segmentation
3D foot model reconstruction
Foot masks + Depth Images
Joint center trajectories estimation
3D static subject-specific foot model
Joint kinematics estimation
Lower-limb joint angles

Figure 40. Block diagram of the proposed 3D markerless protocol.

1) 3D foot model reconstruction

- *Multi-segmental foot model definition*

A 3D subject-specific kinematic foot model was introduced to estimate sagittal foot kinematics. The model consisted in two-segments (mid rear foot and forefoot) connected by a revolute joint (metatarsophalangeal joint). The ankle joint (AJC) coincided to the lateral malleolus (LM) and MTP joint coincided to the 5th metatarsophalangeal point (MTP5).

The primary objective stated in this work is the creation of a two-segments 3D foot model using a single camera which can lead to obtaining more complete information about foot kinematics. In order to achieve this aim, a 3D foot model was created by capturing four views of the foot in a static position: frontal (FRO), posterior (POS), lateral (LAT), and medial (MED) (**Figure 41**).

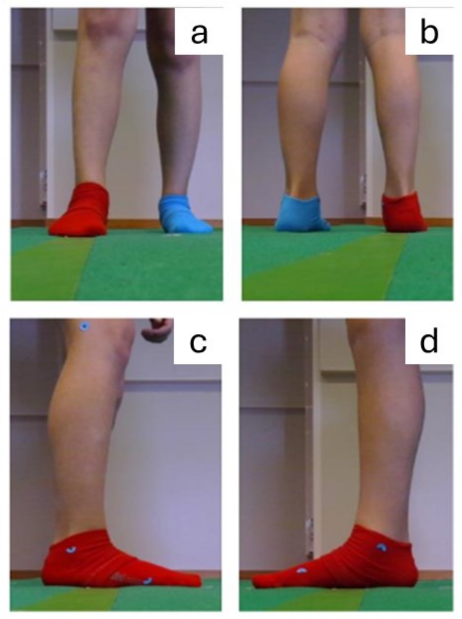


Figure 41. Four static views acquired to create a 3D foot model a) frontal view, b) posterior view, c) lateral view and c) medial view of the right foot.

The process involves a step by-step algorithm divided into four stages. The details of each step are provided below:

1. Point cloud creation: for each view, a 3D point cloud of the foot was generated by exploiting the RGB-Depth information as described in the Chapter 1 – paragraph "3D point cloud reconstruction" (**Figure 42**a).
2. Ground removal and identification of the foot sole: As illustrated in **Figure 42**a, the foot point cloud could be subjected to errors. These inaccuracies could arise from the ground or from shadows due to varying body weight distribution on the feet when stationary. To address this issue, a pre-set threshold based on a small percentage of foot height (i.e. 8%) was applied to remove the 3D points below that threshold ensuring that the anatomical details of the foot were preserved (**Figure 42**b).

Then, the foot sole was identified as the plane fitting the lowest part of the foot point cloud.

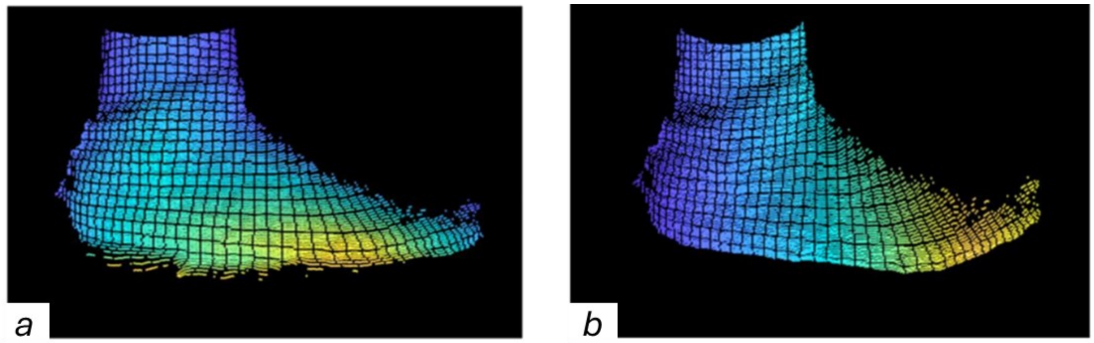


Figure 42. a) Point cloud of the lateral view of the right foot b) Point cloud after removing the ground.

3. Identification of common points on the foot sole: For each view, specific points were identified:
    - the most anterior point (Z1) and the most lateral point (Z2) in FRO;
    - the most anterior point (Z1), the most lateral point (Z2), and the most posterior point (Z3) in LAT;
    - the most anterior (Z2) and the most posterior point (Z3) in MED;
    - the most posterior point (Z3) in POS.

    Prior to the alignment, LAT, MED and POS was rotated of 90°, 180° and 270°, respectively.
4. Alignment of views: FRO was used as the reference, the other three views were aligned following this order:
    1) LAT was aligned to FRO using Z2 as common point;
    2) MED was aligned to FRO using Z1 as common point;
    3) POS was aligned to LAT using Z3 as common point.

Thanks to these alignments, a 3D foot model, ${}^{I}\boldsymbol{M}_{foot}$,was created preserving the foot anatomical features (**Figure 43**).

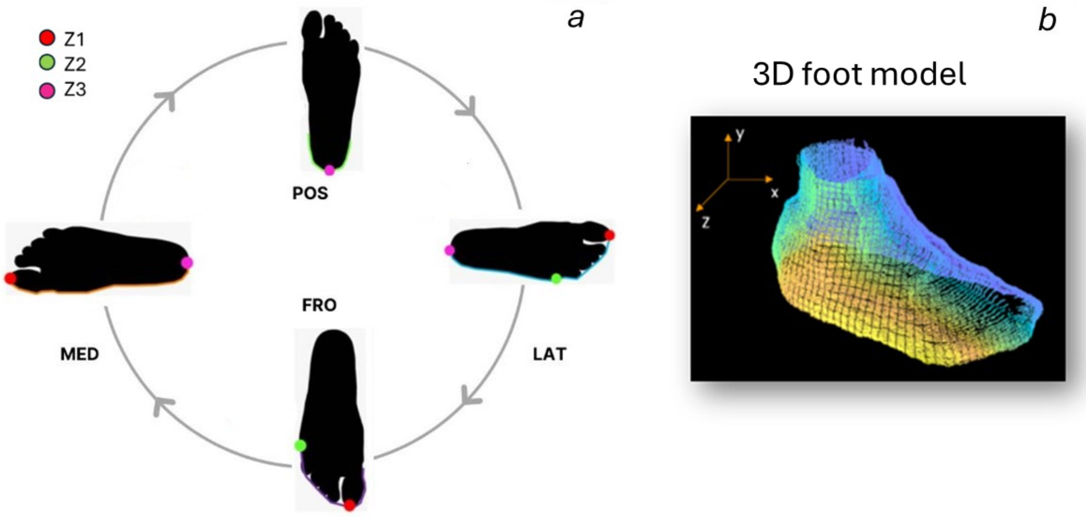


Figure 43. a) Common points for each view b) 3D foot model, ${}^{I}\boldsymbol{M}_{foot}$, obtained from the alignment of four views.

- *Anatomical calibration and body segment templates definition*

The body segments' templates and the relevant coordinate systems were calibrated on the static upright standing acquisition (image "0") by manually selecting two anatomical landmarks (LM, MTP5) to obtain their position vectors in I (${}^{I}\boldsymbol{LM}_0$, ${}^{I}\boldsymbol{MTP5}_0$). Finally, the ${}^{I}\boldsymbol{TOE}_0$ was identified as the most distal point of the foot at the same y-coordinate of ${}^{I}\boldsymbol{MTP5}_0$. To account for potential right/left asymmetries, the subject-specific model was defined for both sides.

From ${}^{I}\boldsymbol{M}_{foot}$, the mid-rear foot portion was extracted to define a template ${}^{I}\boldsymbol{TMP}_{mid-rear-foot}$ where the value of its generic pixel ${}^{I}TMP_{mid-rear-foot}(x,y,z)$ in the I was obtained **Figure 44**b as:

$$ {}^{I}TMP_{mid-rear-foot}(x,y,z) = \left\{ \begin{array}{l|l} 1, & {}^{I}M_{foot}(x,y,z) = 1 \cap CA_{xi} < x < MTP5_{xi} \\ 0, & otherwise \end{array} \right\} $$

Where CA_{xi} is the x-coordinate of heel of the foot identified as in “Gait cycle ”, $MTP5_{xi}$ is the x-coordinate of MTP5 and ${}^{I}M_{foot}(x, y, z)$ is a generic pixel of ${}^{I}\boldsymbol{M}_{foot}$ expressed in the *I*.

The forefoot portion was extracted to define a template ${}^{I}\boldsymbol{TMP}_{forefoot}$ where the value of its generic pixel ${}^{I}TMP_{forefoot}(x, y, z)$ in the *I* was obtained (**Figure 44**c) as:

$$
{}^{I}TMP_{forefoot}(x, y, z) = \left\{ \begin{array}{l|l} 1, & {}^{I}M_{foot}(x, y, z) = 1 \cap MTP5_{xi} < x < TOE_{xi} \\ 0, & otherwise \end{array} \right\}
$$

Where TOE_{xi} is the x-coordinate of the toe.

An ellipsoid was fitted on ${}^{I}\boldsymbol{TMP}_{mid-rear-foot}$. Then, a mid-rear foot coordinate system (mrf_0) was defined with the axes coincident to the inertial ellipsoid principal axes and the origin coinciding with its centroid. The transformation matrix ${}^{I}\boldsymbol{T}_{mrf_0}$ from mrf_0 to *I* was computed by simple geometrical rules.

The same procedure was applied to ${}^{I}\boldsymbol{TMP}_{forefoot}$ to obtain the transformation matrix ${}^{I}\boldsymbol{T}_{ff_0}$ from a forefoot coordinate system (ff_0) to *I*. The transformation matrix ${}^{I}\boldsymbol{T}_{mrf0}$ from mrf_0 to *I* and ${}^{I}\boldsymbol{T}_{ff0}$ from ff_0 to *I* were applied to transform ${}^{I}\boldsymbol{TMP}_{mid-rear-foot}$ in the mrf_0 (${}^{mrf0}\boldsymbol{TMP}_{mid-rear-foot}$) and ${}^{I}\boldsymbol{TMP}_{forefoot}$ in the ff_0 (${}^{ff0}\boldsymbol{TMP}_{forefoot}$).

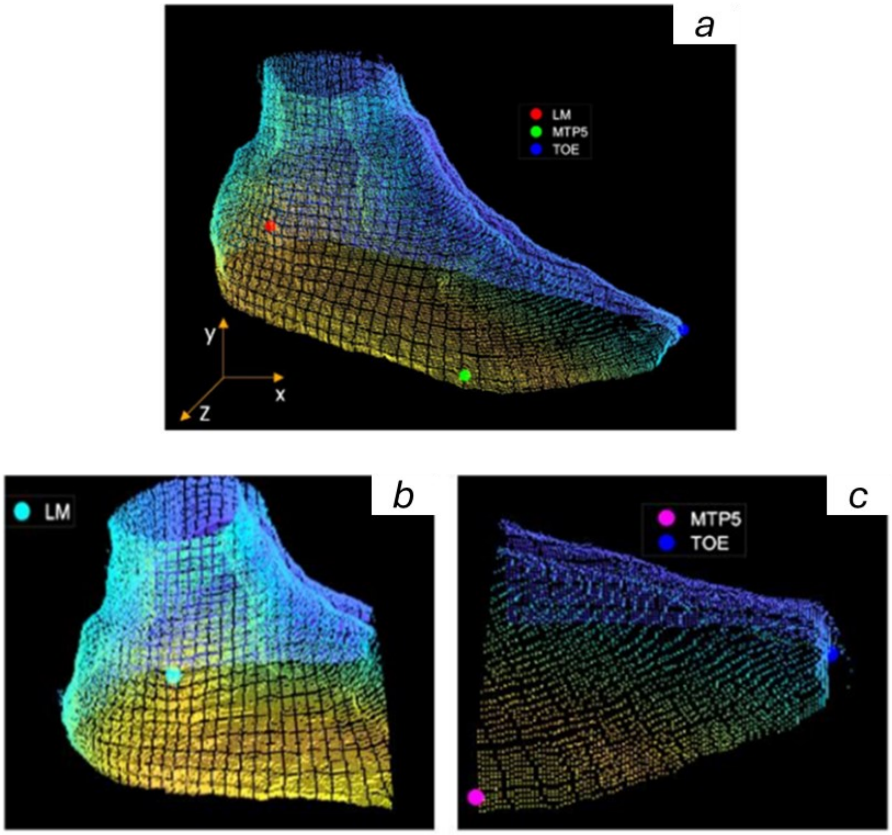


Figure 44. a) 3D foot model after the anatomical calibration b) $^{I}\boldsymbol{TMP}_{mid-rear-foot}$ with its relevant $^{I}\boldsymbol{LM}_{0}$ coordinates c) $^{I}\boldsymbol{TMP}_{forefoot}$ with its relevant $^{I}\boldsymbol{MTP5}_{0}$ and $^{I}\boldsymbol{TOE}_{0}$.

2) 2D shank template reconstruction

A 2D shank template was obtained following the procedure described in Chapter 2 “Concurrent validation of a 2D model-based approach on individuals with Cerebral Palsy based on a single RGB-D camera.”

3) Joint centers trajectories estimation

From each gait trial, the most central gait cycle was selected and analyzed based on the identification of initial foot contacts as described in Chapter 2 “Concurrent validation of a 2D model-based approach on individuals with Cerebral Palsy based on a single RGB-D camera.”

For each frame of the gait cycle, the joint center positions were identified following these steps:

1) Segmentation of the foreground foot: as reported in the previous paragraph, two color filters were implemented to identify the foot in the foreground and create a 2D foot mask. A color filter is based on the RGB values of the pixels, so its effectiveness depends on the lighting conditions of the environment (Cheng et al., 2001).
2) Identification of the depth values of the foreground foot: 2D foot mask was overlapped to the depth image to extract the corresponding depth values.
3) Depth completion technique and creation of 3D foot point cloud: The depth sensor of the Azure Kinect may produce an inaccurate depth map when the captured object is in motion, especially when the foot reaches its highest speed within the gait cycle. The faster the movement, the poorer the quality of the depth reconstruction, as demonstrated by the foot in **Figure 45**a. For this reason, to prevent the loss of anatomical information, the missing depth points were reconstructed to enhance the estimation of the joint centers' positions through a proper depth completion technique.

Depth completion for RGB-Depth images is a technique that aims to recover dense depth maps from sparse depth measurements (Xu et al., 2019). The proposed MS method, in order to recover the missing depth points, includes a depth completion based on a low-pass filter using a 5x5 kernel. The process is described in the following three steps:

A) Selection of target regions: The missing areas in the Depth image were identified by overlaying the 2D foot mask to the Depth image.
B) Kernel design: A 5x5 low-pass kernel was selected, and for each pixel in the missing region, the mean value of its 5x5 neighborhood was computed and used to replace the missing value.
C) Starting point: Missing points are predominantly located in the front part of the foot, where the speed is higher. When the walking direction is towards the right, the kernel starts sliding from the top-left corner of the Depth map, covering existing depth points before reaching the missing ones; this sequence is reversed when the walking direction changes.

After having reconstructed the missing depth points, depth values belonging to the subject's foot were then converted in the 3D coordinate system in order to generate a point cloud, $\boldsymbol{Dyn_{foot}}$, as described in the Chapter 1, paragraph "3D point cloud reconstruction".

The depth image and the relevant point cloud before and after the implementation of depth completion technique are shown in **Figure 45**.

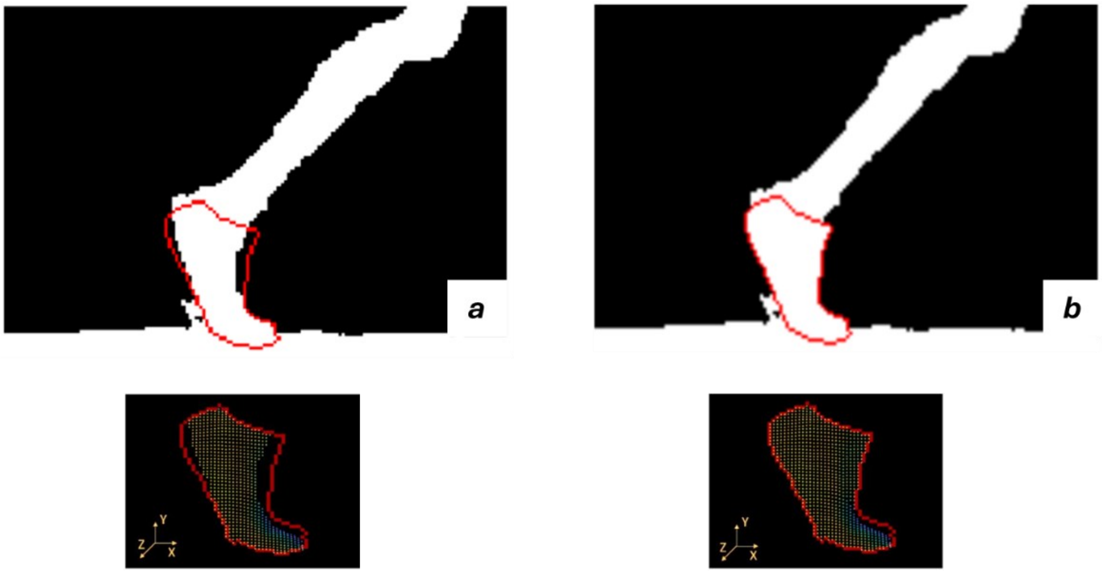


Figure 45. a) Depth image (white) overlapped to the 2D foot mask (red) before Depth completion b) Depth image (white) overlapped to the 2D foot mask (red) after Depth completion.

For each frame of the gait cycle, the joint center positions (LM and MTP5) were identified by aligning each dynamic point cloud and the 3D foot template previously identified using the ICP algorithm.

As explained in paragraph "3D foot model reconstruction", each dynamic point cloud, $\boldsymbol{Dyn_{foot}}$, (**Figure 46**a) was split into two segments representing the mid-rear foot $^{I}\boldsymbol{M}_{mid-rear-foot}$ (**Figure 46**b) and the forefoot $^{I}\boldsymbol{M}_{forefoot}$ (**Figure 46**c) considering the same splitting percentage obtained in the static posture.

The mid-rear-foot coordinate system mrf and the relevant transformation matrix ${}^{I}\boldsymbol{T}_{mrf}$ was defined as described in "3D foot model reconstruction - *Anatomical calibration and body segment templates definition*"

In order to compensate for any difference in size between ${}^{I}\boldsymbol{M}_{mid-rear-foot}$ and ${}^{I}\boldsymbol{TMP}_{mid-rear-foot}$ due to the presence of soft tissue artifacts and the degradation of depth images caused by a fast movement, a proper scaling factor was applied to ${}^{I}\boldsymbol{M}_{mid-rear-foot}$.

The same procedure was applied to ${}^{I}\boldsymbol{M}_{forefoot}$ to obtain the forefoot coordinate system ff and the relevant transformation matrix ${}^{I}\boldsymbol{T}_{ff}$.

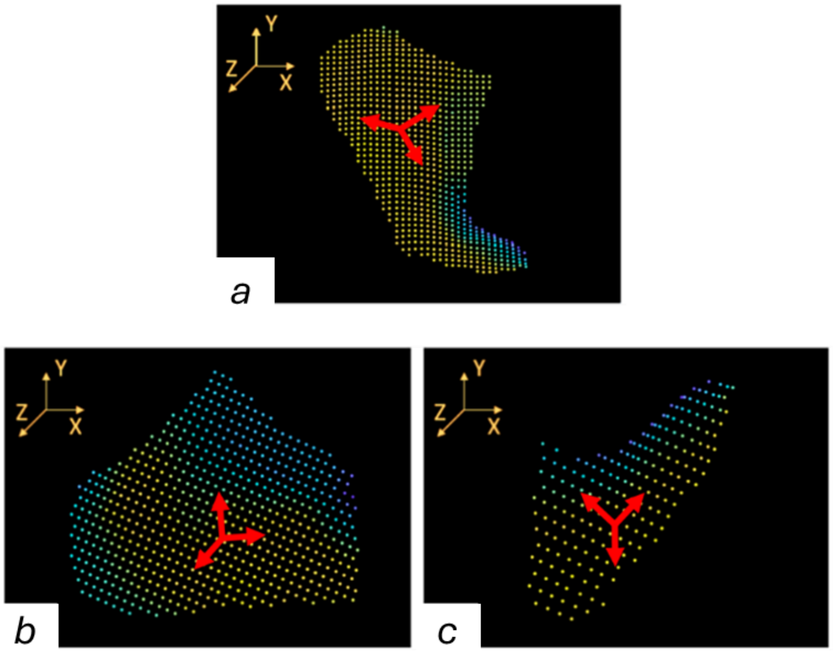


Figure 46. a) A point cloud of the foot for each frame, $\boldsymbol{Dyn_{foot}}$ b) ${}^{I}\boldsymbol{M}_{mid-rear-foot}$ and its relevant ${}^{I}\boldsymbol{T}_{mrf}$ c) ${}^{I}\boldsymbol{M}_{forefoot}$ and its relevant ${}^{I}\boldsymbol{T}_{ff}$.

- *LM trajectories estimation*

The centroids of ${}^{I}\boldsymbol{TMP}_{mid-rear-foot}$ and ${}^{I}\boldsymbol{M}_{mid-rear-foot}$ were made to coincide on the x- and y- axes while an offset equal to the half of the foot width was added on the z-axis to place ${}^{I}\boldsymbol{M}_{mid-rear-foot}$ laterally with respect to ${}^{I}\boldsymbol{TMP}_{mid-rear-foot}$ (**Figure 47**).

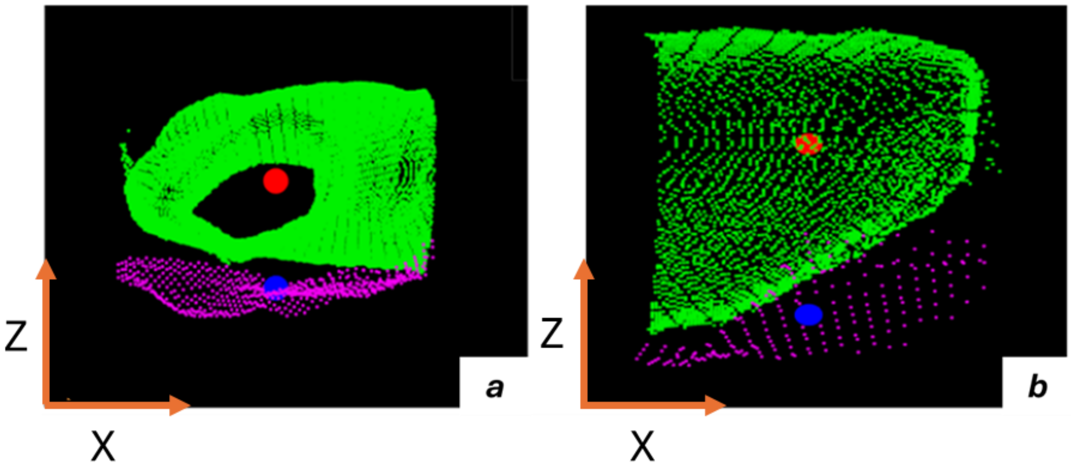


Figure 47. a) The centroids of ${}^{I}\boldsymbol{TMP}_{mid-rear-foot}$ (red) and ${}^{I}\boldsymbol{M}_{mid-rear-foot}$ (blue) were made to coincide on the x- and y- axes while on the z-axis an offset equal to the half of the foot width was added to position ${}^{I}\boldsymbol{M}_{mid-rear-foot}$ laterally with respect to ${}^{I}\boldsymbol{TMP}_{mid-rear-foot}$ b) The same procedure was applied to ${}^{I}\boldsymbol{TMP}_{forefoot}$ and ${}^{I}\boldsymbol{M}_{forefoot}$.

Using a 3D ICP technique (Besl & McKay, 1992), the ${}^{mrf0}\boldsymbol{TMP}_{mid-rear-foot}$ was matched with the ${}^{mrf}\boldsymbol{M}_{mid-rear-foot}$ and the relevant transformation matrix ${}^{mrf0}\boldsymbol{T}_{mrf}$ (4x4), determined (**Figure 48**a). Finally, ${}^{I}\boldsymbol{LM}$ was determined for each frame by referencing the position of LM in the template, ${}^{I}\boldsymbol{LM}_{0}$, through the application of the following three successive transformations:

$${}^{I}\boldsymbol{LM} = {}^{I}\boldsymbol{T}_{mrf}\,{}^{mrf}\boldsymbol{T}_{mrf0}\,{}^{mrf0}\boldsymbol{T}_{I}\,{}^{I}\boldsymbol{LM}_{0}$$

- *MTP5 trajectories estimation*

Similarly to the LM trajectories estimation, the centroids of ${}^{I}\boldsymbol{TMP}_{forefoot}$ and ${}^{I}\boldsymbol{M}_{forefoot}$ were made to coincide on both *x*- and *y*- axes while on the *z*-axis an offset equal to the half of the foot width was introduced to place ${}^{I}\boldsymbol{M}_{forefoot}$ laterally with respect to

${}^{I}\boldsymbol{TMP}_{forefoot}$. Using a 3D ICP technique (Besl & McKay, 1992), the ${}^{ff0}\boldsymbol{TMP}_{forefoot}$ was matched with the ${}^{ff}\boldsymbol{M}_{forefoot}$ and the relevant transformation matrix ${}^{ff0}\boldsymbol{T}_{ff}$ (4x4), determined (**Figure 48**b).

Finally, ${}^{I}\boldsymbol{MTP5}$ was determined for each frame by referencing the position of MTP5 in the template ${}^{I}\boldsymbol{MTP5}_{0}$ through the application of the following three successive transformations:

$${}^{I}\boldsymbol{MTP5} = {}^{I}\boldsymbol{T}_{ff}\,{}^{ff}\boldsymbol{T}_{ff_0}\,{}^{ff_0}\boldsymbol{T}_{I}\,{}^{I}\boldsymbol{MTP5}_{0}$$

- *LE trajectories estimation*

For each frame, ${}^{I}\boldsymbol{LE}$ was extracted as described in (Balta et al., 2023) and explained in "Chapter 2 - Concurrent validation of a 2D model-based approach on individuals with Cerebral Palsy based on a single RGB-D camera."

- *TOE trajectories estimation*

For each frame, ${}^{I}\mathbf{TOE}$ was extracted as the most distal part of $\boldsymbol{Dyn_{foot}}$.

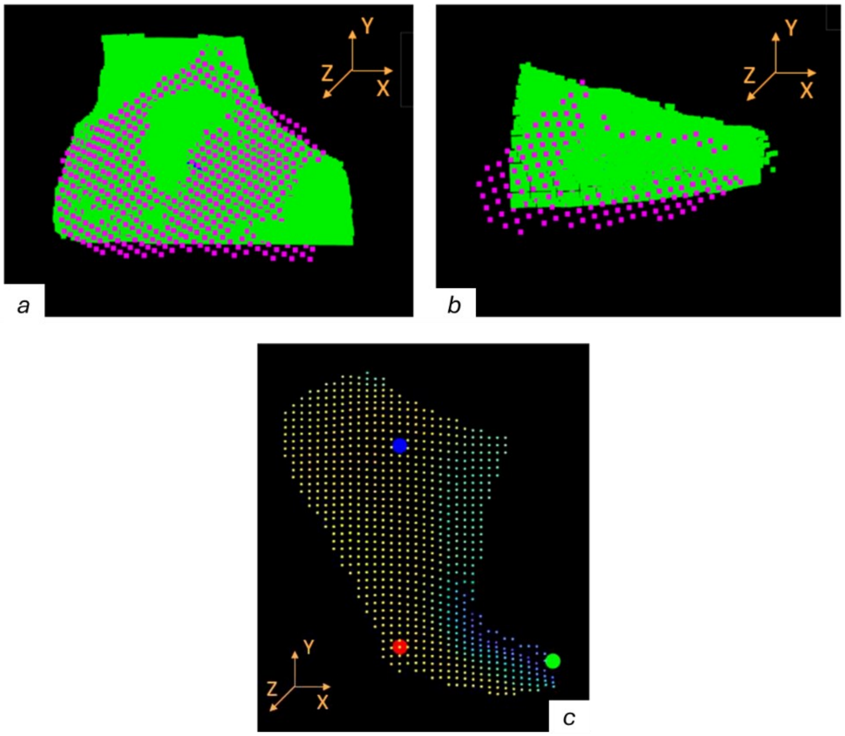


Figure 48. Joint trajectories estimation. a) $^{I}\boldsymbol{TMP}_{mid-rear-foot}$ was matched with the $^{I}\boldsymbol{M}_{mid-rear-foot}$, b) $^{I}\boldsymbol{TMP}_{forefoot}$ was matched with the $^{I}\boldsymbol{M}_{forefoot}$, c) $^{I}\boldsymbol{LM}$ (blue), $^{I}\boldsymbol{MTP5}$ (red) and $^{I}\boldsymbol{TOE}$ (green).

4) Joint kinematics estimation

Joint kinematics was determined based on the segment inclination as defined by the lines connecting the joint centers. First of all, $^{I}\boldsymbol{LM}$, $^{I}\boldsymbol{MTP5}$ and $^{I}\boldsymbol{TOE}$ were projected in the 2D image plane. For the ankle, the plantar-dorsi flexion angle was determined as the angle between $^{I}\boldsymbol{LE}$ - $^{I}\boldsymbol{LM}$ and the $^{I}\boldsymbol{LM}$ - $^{I}\boldsymbol{MTP5}$ vectors. for the metatarsophalangeal joint, the flexion-extension angle was determined as the angle between $^{I}\boldsymbol{LM}$ - $^{I}\boldsymbol{MTP5}$ and the $^{I}\boldsymbol{MTP5}$ - $^{I}\boldsymbol{TOE}$ vectors.

2.5.3 Performance assessment and statistical analysis

Prior to comparison, the kinematic curves from the MS system and those extracted from manual labeling (LAB) were processed using a fourth-order Butterworth filter with a cutoff frequency of 7 Hz and then time-normalized to the gait cycle (1-100%) (Bergamini et al., 2014). The mean absolute error (MAE) and the mean absolute percentage error (MAPE) between the foot length obtained from the proposed MS protocol and the measured one were computed.

For each subject, *s*, gait trial, *t*, and joint, *j*, the performance of the proposed MS method in estimating the foot kinematics were assessed in terms of offset and waveform similarity (Picerno et al., 2008). The offset was computed as the absolute difference between the mean value of the kinematic curves obtained from the proposed MS protocol ($\overline{MS}$) and kinematic curves extracted from the manual labeling, $\overline{LAB}$, within a gait cycle:

$$Offset_{s,t,j} = \left|\overline{LAB_{s,t,J}} - \overline{MS_{s,t,J}}\right|$$

The latter values were then averaged across trials and subjects:

$$Offset_j = \frac{1}{N_S}\sum_{s=1}^{N_S}\frac{1}{N_T}\sum_{t=1}^{N_T} Offset_{s,t,j}$$

Where N_T = 6 is the number of trials and N_S = 10 is the number of subjects.

For each subject, *s*, gait trial, *t*, and joint, *j*, the waveform similarity was evaluated as the root mean square error (RMSE) of the MS joint kinematic curves with respect to the joint kinematic curves from the manual labeling (LAB), after removing their mean values (Picerno et al., 2008):

$$RMSE_{s,t,j} = RMS((LAB_{s,t,j} - \overline{LAB_{s,t,J}}) - (MS_{s,t,j} - \overline{MS_{s,t,J}}))$$

For each joint, the latter values were then averaged across trials and subjects:

$$RMSE_j = \frac{1}{N_S}\sum_{s=1}^{N_S}\frac{1}{N_T}\sum_{t=1}^{N_T} RMSE_{s,t,j}$$

Where N_T = 6 is the number of trials and N_S = 10 is the number of subjects.

2.5.4 Results

Results related to the accuracy of 3D model foot reconstruction, in terms of MAE and MAPE, are reported in **Table 9**.

Table 9. Mean Absolute Error and Mean Absolute Percentage error between the actual foot length and the automatic estimated value averaged across subjects.

3D foot model	*MAE (mm)*	*MAPE (%)*
Right	12.3 ± 7.6	5.2 ± 3.0
Left	16 ± 4.4	5.7 ± 2.4

Results for sagittal foot kinematics, in terms of offset and RMSE are reported in **Table 10**.

Table 10. The average root-mean-square errors (RMSE) value and the average offset between the joint kinematics curves estimated by the MS method and those extracted from the manual labeling are computed over the gait cycle and averaged across trials and subjects

Ankle	***Offset (deg)***	***RMSE (deg)***
Right	2.3 ± 1.7	4.8 ± 0.7
Left	3.5 ± 2.0	4.9 ± 1.6
Metatarsophalangeal	***Offset (deg)***	***RMSE (deg)***
Right	3.5 ± 3.0	4.8 ± 0.7
Left	6.5 ± 4.0	5.3 ± 0.5

An ensemble view of the normalized joint kinematics curves, averaged over trials and subjects, is reported in **Figure 49.** and **Figure 50**.

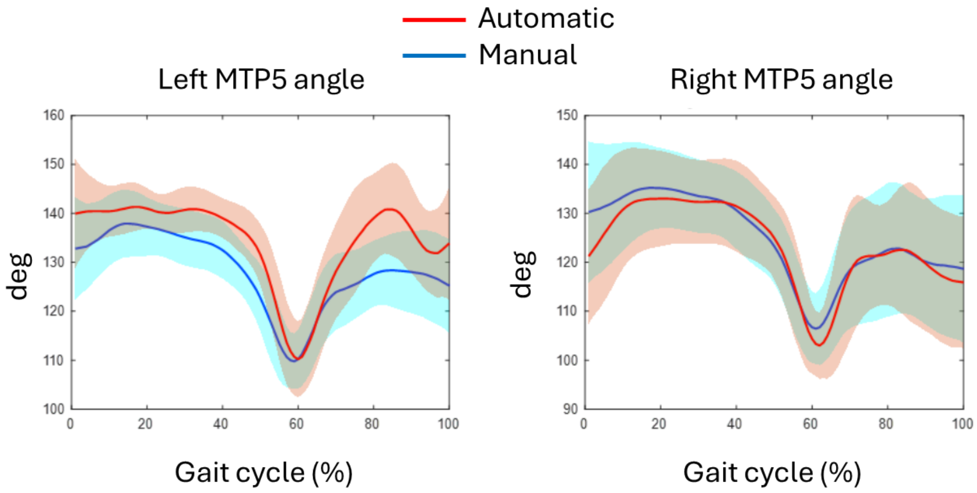


Figure 49. Metatarsophalangeal joint kinematics averaged over subjects and trials (average: solid lines; SD: shaded area; red = MS system; blue = manual labeling).

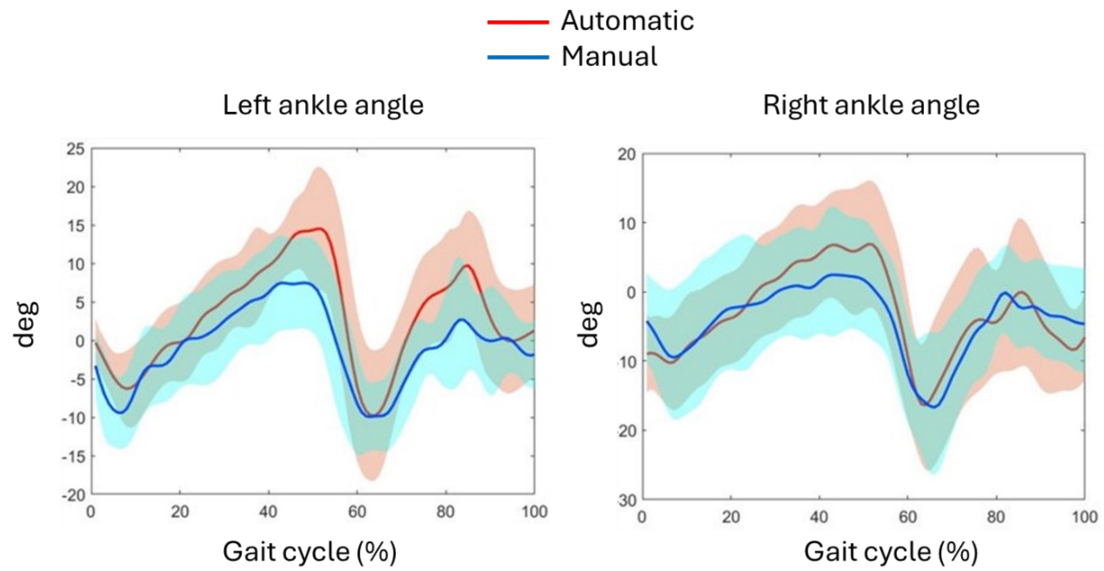


Figure 50. Ankle joint kinematics averaged over subjects and trials (average: solid lines; SD: shaded area; red = MS system; blue = manual labeling).

2.5.5 Discussions

2.5.5.1 Factors influencing the accuracy of foot model creation

The process of creating the foot model is semi-automated. It must be highlighted that it is crucial to properly position the subject during the static acquisitions to accurately identify the common points among the four views. One potential source of errors could be attributed to the incorrect positioning of the subject during the static acquisition phase. The most significant errors occur when the subject's shank is not perpendicular to the ground causing different foot shapes across the views due to a different body weight distribution as shown in **Figure 51**.a. Another problem which cloud affect the accuracy of the model reconstruction is a misalignment between the foot axis and sagittal plane during the acquisition of lateral and medial side. If this alignment is not respected (as shown in **Figure 51**.b), the model reconstruction could require manual adjustments since common points (identified as explained in "3D foot model reconstruction") could represent different foot portion across the four views.

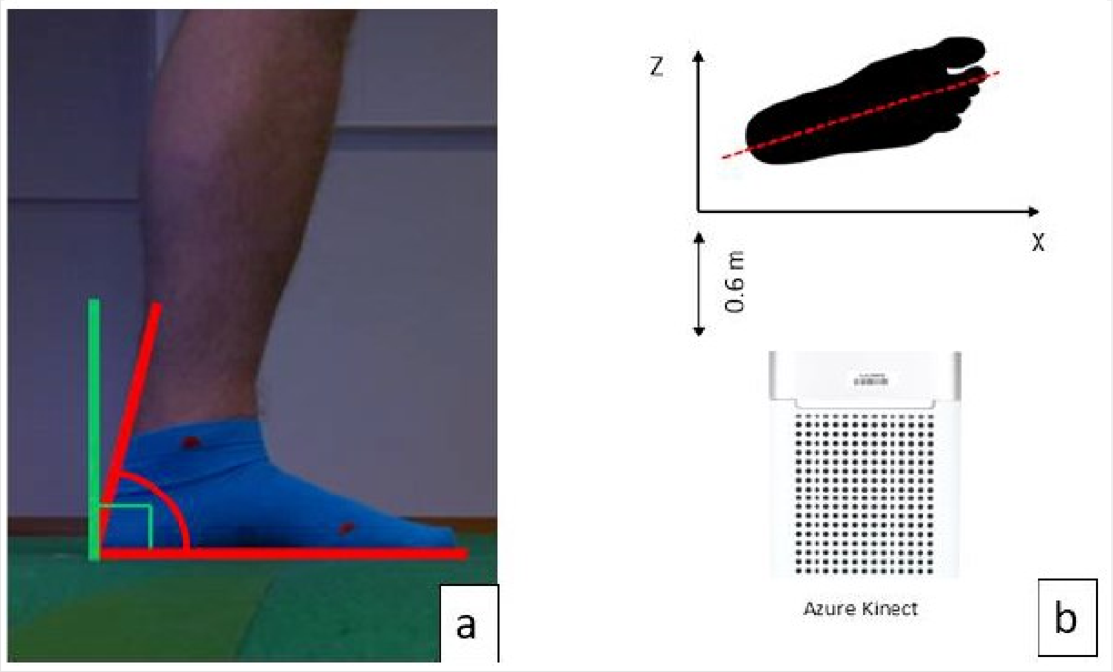


Figure 51. a) The incorrect position on the static acquisition for the medial view of left foot (red lines). The patient should stand as indicated by the green line. b) Representation of misalignment of principal foot axis (red dotted line) and the X-axis during the static acquisition for the lateral view.

2.5.5.2 Factors influencing the accuracy of joint kinematics estimation

This paragraph will examine the sources of residual errors in foot kinematics estimation related to the proposed depth completion technique and the ICP algorithm. A primary factor influencing the accuracy of joint center estimations is associated to technological limitations of the depth sensor. In particular, the errors are caused by inaccuracies in the measurement of the depth values from the depth sensor during high-speed movements resulting also in a small number of points belonging to the distal part of the shank and foot. During high-speed movements, it has been noticed that the depth sensor fails in accurately reconstructing depth values since the limited exposure time of RGB-D cameras can lead to motion blurs in captured images, potentially causing artifacts such as holes or fake boundaries (Gao et al., 2015), resulting in improper alignment between the depth image and the RGB image specifically at the foot (**Figure 52**).

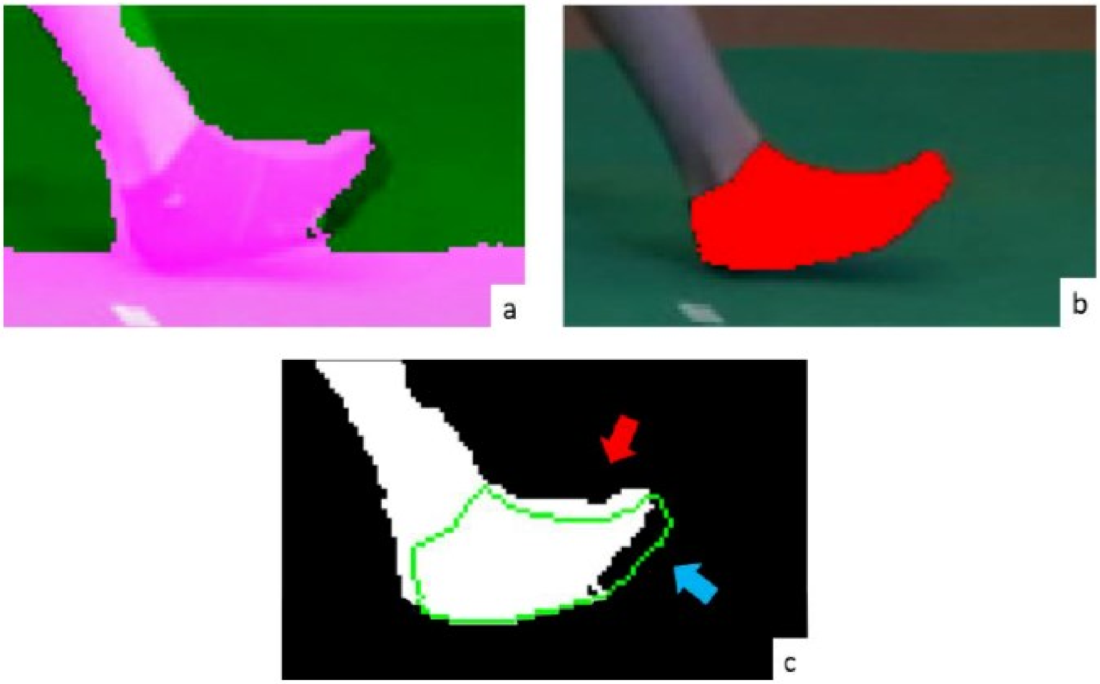


Figure 52. a) Depth map (pink pixels) overlapped to RGB image. b) RGB mask (red pixels) overlapped to RGB image. c) RGB mask edge (green line) overlapped to depth map (white pixels). The red arrow indicates the misalignment while blue arrow indicates the missed depth values.

Even when implementing a specific depth completion technique, a significant number of missing points could lead to an inaccurate reconstruction of the missing depth values, which would not represent the actual 3D shape of the foot.

The identification of the left foot, wearing a blue sock, was challenging during the stance phase due to the color similarity between the sock and the green carpet, as well as due to the shadows caused by the foot itself. These issues can result in the misidentification of the foot on the depth image, causing the generation of inaccurate point clouds and consequently an inaccurate alignment through the ICP algorithm.

Furthermore, during the swing phase, the rapid movement of the foot causes the blue color of the sock to blur. This blurring effect complicates the manual labeling of anatomical landmarks for validating joint kinematics, making it difficult to correctly label the appropriate pixels.

The ICP algorithm is a rigid alignment technique that presupposes the objects to be aligned are rigid segments. However, this assumption is problematic when applied to the human foot, which is not a rigid body and can undergo shape variations, particularly during the swing phase, due to soft tissue artifacts (Van den Herrewegen et al., 2014) and important foot deformities.

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

The proposed 3D method for studying foot kinematics cannot be directly compared with other 3D methods which estimate foot joint angles using 3D scanners, as some methods compare joint kinematics against the stereophotogrammetric system only qualitatively and only during the stance phase (Van den Herrewegen et al., 2014). Moreover, other studies (Jiang et al., 2023) proposed low-cost systems based on depth sensors for the reconstruction of a multi-segment foot model and validated the 3D foot reconstruction in terms of root mean square error with respect to reference commercial system (e.g. laser scanners) avoiding a direct comparison with the proposed MS protocol.

To the best of our knowledge, the only study implementing a MS method based on a single RGB camera for foot kinematics, including the rearfoot-forefoot flexion/extension angle, was proposed by (Surer et al., 2011). This study employed a multi-rigid body model with three segments (shank, rearfoot, and forefoot) connected by cylindrical hinges, analyzing two degrees of freedom: ankle plantar/dorsi-flexion and rearfoot-forefoot flexion/extension. This study reported an RMSE of 2 deg for the ankle kinematics and 3.1 deg for the MTP joint kinematics. However, it is worth noting that this validation was conducted only on three healthy subjects and was limited to the stance phase, excluding the swing phase, where we observed that the alignment is particularly deteriorated due to depth limitations.

References

Albert, J. A., Owolabi, V., Gebel, A., Brahms, C. M., Granacher, U., & Arnrich, B. (2020). 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 (Switzerland)*, *20*(18), 1–22. https://doi.org/10.3390/s20185104

Allan, J. J., McClelland, J. A., Munteanu, S. E., Buldt, A. K., Landorf, K. B., Roddy, E., Auhl, M., & Menz, H. B. (2020). First metatarsophalangeal joint range of motion is associated with lower limb kinematics in individuals with first metatarsophalangeal joint osteoarthritis. *Journal of Foot and Ankle Research*, *13*(1). https://doi.org/10.1186/s13047-020-00404-0

Amprimo, G., Pettiti, G., Priano, L., Mauro, A., & Ferraris, C. (2021). *Kinect-based solution for the home monitoring of gait and balance in elderly people with and without neurological diseases*.

Balta, D., Figari, G., Paolini, G., Pantzar-Castilla, E., Riad, J., Croce, U. D., & Cereatti, A. (2023). 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*, 1. https://doi.org/10.1109/ACCESS.2023.3340622

Balta, D., Salvi, M., Molinari, F., Figari, G., Paolini, G., Croce, U. Della, & Cereatti, A. (2020). A two-dimensional clinical gait analysis protocol based on markerless recordings from a single RGB-Depth camera. *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 1–6. https://doi.org/10.1109/MeMeA49120.2020.9137183

Benedetti, M. G., Catani, F., Leardini, A., Pignotti, E., & Giannini, S. (1998). Data management in gait analysis for clinical applications. *Clinical Biomechanics*, *13*(3), 204–215. https://doi.org/https://doi.org/10.1016/S0268-0033(97)00041-7

Bergamini, E., Ligorio, G., Summa, A., Vannozzi, G., Cappozzo, A., & Sabatini, A. M. (2014). Estimating orientation using magnetic and inertial sensors and different sensor fusion approaches: Accuracy assessment in manual and locomotion tasks. *Sensors (Switzerland)*, *14*(10), 18625–18649. https://doi.org/10.3390/s141018625

Bertram, J., Krüger, T., Röhling, H. M., Jelusic, A., Mansow-Model, S., Schniepp, R., Wuehr, M., & Otte, K. (2023). Accuracy and repeatability of the Microsoft Azure Kinect for clinical measurement of motor function. *PLoS ONE*, *18*(1 January). https://doi.org/10.1371/journal.pone.0279697

Besl, P. J., & McKay, N. D. (1992). A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*(2), 239–256. https://doi.org/10.1109/34.121791

Cappello, A., Cappozzo, A., La Palombara, P. F., Lucchetti, L., & Leardini, A. (1997). Multiple anatomical landmark calibration for optimal bone pose estimation. *Human Movement Science*, *16*(2), 259–274. https://doi.org/https://doi.org/10.1016/S0167-9457(96)00055-3

Castelli, A., Paolini, G., Cereatti, A., & Della Croce, U. (2015). A 2D markerless gait analysis methodology: Validation on healthy subjects. *Computational and Mathematical Methods in Medicine*, *2015*. https://doi.org/10.1155/2015/186780

Cereatti, A., Bonci, T., Akbarshahi, M., Aminian, K., Barré, A., Begon, M., Benoit, D. L., Charbonnier, C., Dal Maso, F., Fantozzi, S., Lin, C.-C., Lu, T.-W., Pandy, M. G., Stagni, R., van den Bogert, A. J., & Camomilla, V. (2017). Standardization proposal of soft tissue artefact description for data sharing in human motion measurements. *Journal of Biomechanics*, *62*, 5–13. https://doi.org/https://doi.org/10.1016/j.jbiomech.2017.02.004

Cereatti, A., Rosso, C., Nazarian, A., DeAngelis, J. P., Ramappa, A. J., & Croce, U. Della. (2015). Scapular motion tracking using acromion skin marker cluster: In vitro accuracy assessment. *Journal of Medical and Biological Engineering*, *35*(1), 94–103. https://doi.org/10.1007/s40846-015-0010-2

Ceseracciu, E., Sawacha, Z., & Cobelli, C. (2014). Comparison of Markerless and Marker-Based Motion Capture Technologies through Simultaneous Data Collection during Gait: Proof of Concept. *PLOS ONE*, *9*(3), e87640-. https://doi.org/10.1371/journal.pone.0087640

Chen, S. W., Lin, S. H., Liao, L. De, Lai, H. Y., Pei, Y. C., Kuo, T. S., Lin, C. T., Chang, J. Y., Chen, Y. Y., Lo, Y. C., Chen, S. Y., Wu, R., & Tsang, S. (2011). Quantification and recognition of parkinsonian gait from monocular video imaging using kernel-based principal component analysis. *BioMedical Engineering Online*, *10*. https://doi.org/10.1186/1475-925X-10-99

Cheng, H. D., Jiang, X. H., Sun, Y., & Wang, J. (2001). Color image segmentation: advances and prospects. *Pattern Recognition*, *34*(12), 2259–2281. https://doi.org/https://doi.org/10.1016/S0031-3203(00)00149-7

Cimolin, V., Vismara, L., Ferraris, C., Amprimo, G., Pettiti, G., Lopez, R., Galli, M., Cremascoli, R., Sinagra, S., Mauro, A., & Priano, L. (2022). Computation of Gait Parameters in Post Stroke and Parkinson's Disease: A Comparative Study Using RGB-D Sensors and Optoelectronic Systems. *Sensors*, *22*(3). https://doi.org/10.3390/s22030824

Clark, R. A., Bower, K. J., Mentiplay, B. F., Paterson, K., & Pua, Y. H. (2013). Concurrent validity of the Microsoft Kinect for assessment of spatiotemporal gait variables. *Journal of Biomechanics*, *46*(15), 2722–2725. https://doi.org/10.1016/j.jbiomech.2013.08.011

Clark, R. A., Vernon, S., Mentiplay, B. F., Miller, K. J., McGinley, J. L., Pua, Y. H., Paterson, K., & Bower, K. J. (2015). Instrumenting gait assessment using the Kinect in people living with stroke: reliability and association with balance tests. *Journal of Neuroengineering and Rehabilitation*, *12*, 15. https://doi.org/10.1186/s12984-015-0006-8

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology*. Pearson education.

Davis, R. B., Õunpuu, S., Tyburski, D., & Gage, J. R. (1991). A gait analysis data collection and reduction technique. *Human Movement Science*, *10*(5), 575–587. https://doi.org/https://doi.org/10.1016/0167-9457(91)90046-Z

Dobbs, M. B., & Gurnett, C. A. (2017). The 2017 ABJS Nicolas Andry Award: Advancing Personalized Medicine for Clubfoot Through Translational Research. *Clinical Orthopaedics and Related Research*, *475*(6), 1716–1725. https://doi.org/10.1007/s11999-017-5290-0

Ferraris, C., Amprimo, G., Masi, G., Vismara, L., Cremascoli, R., Sinagra, S., Pettiti, G., Mauro, A., & Priano, L. (2022). Evaluation of Arm Swing Features and Asymmetry during Gait in Parkinson's Disease Using the Azure Kinect Sensor. *Sensors*, *22*(16). https://doi.org/10.3390/s22166282

Ferraris, C., Amprimo, G., Pettiti, G., Masi, G., & Priano, L. (2022). Automatic Detector of Gait Alterations using RGB-D sensor and supervised classifiers: a preliminary study. *Proceedings - IEEE Symposium on Computers and Communications*, *2022-June*. https://doi.org/10.1109/ISCC55528.2022.9912923

Ferraris, C., Cimolin, V., Vismara, L., Votta, V., Amprimo, G., Cremascoli, R., Galli, M., Nerino, R., Mauro, A., & Priano, L. (2021). Monitoring of gait parameters in post-stroke individuals: A feasibility study using rgb-d sensors. *Sensors*, *21*(17). https://doi.org/10.3390/s21175945

Gao, Y., Yang, Y., Zhen, Y., & Dai, Q. (2015). Depth error elimination for RGB-D cameras. *ACM Transactions on Intelligent Systems and Technology*, *6*(2). https://doi.org/10.1145/2735959

Goffredo Michela and Carter, J. N. and N. M. S. (2009). 2D Markerless Gait Analysis. In P. and N. M. and H. J. Vander Sloten Jos and Verdonck (Ed.), *4th European Conference of the International Federation for Medical and Biological Engineering* (pp. 67–71). Springer Berlin Heidelberg.

Gurnett, C. A., Alaee, F., Kruse, L. M., Desruisseau, D. M., Hecht, J. T., Wise, C. A., Bowcock, A. M., & Dobbs, M. B. (2008). Asymmetric Lower-Limb Malformations in Individuals with Homeobox PITX1 Gene Mutation. *The American Journal of Human Genetics*, *83*(5), 616–622. https://doi.org/10.1016/j.ajhg.2008.10.004

Harvey, A., & Gorter, J. W. (2011). Video gait analysis for ambulatory children with cerebral palsy: Why, when, where and how! *Gait and Posture*, *33*(3), 501–503. https://doi.org/10.1016/j.gaitpost.2010.11.025

Hatamzadeh, M., Busé, L., Chorin, F., Alliez, P., Favreau, J. D., & Zory, R. (2022). A kinematic-geometric model based on ankles' depth trajectory in frontal plane for gait analysis using a single RGB-D camera. *Journal of Biomechanics*, *145*. https://doi.org/10.1016/j.jbiomech.2022.111358

Hatamzadeh, M., Busé, L., Turcot, K., & Zory, R. (2024). Improved markerless gait kinematics measurement using a biomechanically-aware algorithm with subject-specific geometric modeling. *Measurement: Journal of the International Measurement Confederation*, *234*. https://doi.org/10.1016/j.measurement.2024.114857

Herrera, D. C., Kannala, J., & Heikkilä, J. (2012). Joint Depth and Color Camera Calibration with Distortion Correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(10), 2058–2064. https://doi.org/10.1109/TPAMI.2012.125

Hesse, N., Baumgartner, S., Gut, A., & Van Hedel, H. J. A. (2023). Concurrent Validity of a Custom Method for Markerless 3D Full-Body Motion Tracking of Children and Young

Adults based on a Single RGB-D Camera. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. https://doi.org/10.1109/TNSRE.2023.3251440

Jiang, D., Li, J. W., Geng, X., Ma, X., & Chen, W. M. (2023). Fast tool to evaluate 3D movements of the foot-ankle complex using multi-view depth sensors. *Medicine in Novel Technology and Devices*, *17*. https://doi.org/10.1016/j.medntd.2023.100212

Bouguet, J.-Y. (2022). Camera Calibration Toolbox for Matlab. *CaltechDATA*.

Kadaba, M. P., Ramakrishnan, H. K., & Wootten, M. E. (1990). Measurement of Lower Extremity Kinematics During Level Walking. In *Journal of Orthopaedic Research* (Vol. 8383). Orthopaedic Research Society.

Kojovic, N., Natraj, S., Mohanty, S. P., Maillart, T., & Schaer, M. (2021). Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-94378-z

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/https://doi.org/10.1016/j.jcm.2016.02.012

Larsen, K. L., Maanum, G., Frøslie, K. F., & Jahnsen, R. (2012). Ambulant adults with spastic cerebral palsy: The validity of lower limb joint angle measurements from sagittal video recordings. *Gait and Posture*, *35*(2), 186–191. https://doi.org/10.1016/j.gaitpost.2011.09.004

Latorre, J., Colomer, C., Alcañiz, M., & Llorens, R. (2019). Gait analysis with the Kinect v2: Normative study with healthy individuals and comprehensive study of its sensitivity, validity, and reliability in individuals with stroke. *Journal of NeuroEngineering and Rehabilitation*, *16*(1). https://doi.org/10.1186/s12984-019-0568-y

Latorre, J., Llorens, R., Colomer, C., & Alcañiz, M. (2018). Reliability and comparison of Kinect-based methods for estimating spatiotemporal gait parameters of healthy and post-stroke individuals. *Journal of Biomechanics*, *72*, 268–273. https://doi.org/10.1016/j.jbiomech.2018.03.008

Leu, A., Ristic-Durrant, D., & Graser, A. (2011). A robust markerless vision-based human gait analysis system. *SACI 2011 - 6th IEEE International Symposium on Applied Computational Intelligence and Informatics, Proceedings*, 415–420. https://doi.org/10.1109/SACI.2011.5873039

Li, T., Chen, J., Hu, C., Ma, Y., Wu, Z., Wan, W., Huang, Y., Jia, F., Gong, C., Wan, S., & Li, L. (2018). Automatic timed up-and-go sub-task segmentation for Parkinson's disease patients using video-based activity classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(11), 2189–2199. https://doi.org/10.1109/TNSRE.2018.2875738

Lin, J., Wang, Y., Sha, J., Li, Y., Fan, Z., Lei, W., & Yan, Y. (2023). Clinical reliability and validity of a video-based markerless gait evaluation method. *Frontiers in Pediatrics*, *11*. https://doi.org/10.3389/fped.2023.1331176

Lonini, L., Moon, Y., Embry, K., Cotton, R. J., McKenzie, K., Jenz, S., & Jayaraman, A. (2022). Video-Based Pose Estimation for Gait Analysis in Stroke Survivors during Clinical Assessments: A Proof-of-Concept Study. *Digital Biomarkers*, *6*(1), 9–18. https://doi.org/10.1159/000520732

Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J. (n.d.). *SMPL: A Skinned Multi-Person Linear Model*.

Ma, Y., Mithraratne, K., Wilson, N. C., Wang, X., Ma, Y., & Zhang, Y. (2019). The validity and reliability of a kinect v2-based gait analysis system for children with cerebral palsy. *Sensors (Switzerland)*, *19*(7). https://doi.org/10.3390/s19071660

McGinley, J. L., Baker, R., Wolfe, R., & Morris, M. E. (2009). The reliability of three-dimensional kinematic gait measurements: A systematic review. In *Gait and Posture* (Vol. 29, Issue 3, pp. 360–369). https://doi.org/10.1016/j.gaitpost.2008.09.003

Mündermann, L., Corazza, S., & Andriacchi, T. P. (2006). The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. In *Journal of NeuroEngineering and Rehabilitation* (Vol. 3). https://doi.org/10.1186/1743-0003-3-6

Nguyen, M.-H., Hsiao, C.-C., Cheng, W.-H., & Huang, C.-C. (2022). Practical 3D Human Skeleton Tracking Based on Multi-View and Multi-Kinect Fusion. *Multimedia Syst.*, *28*(2), 529–552. https://doi.org/10.1007/s00530-021-00846-x

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

Pantzar-Castilla, E., Cereatti, A., Figari, G., Valeri, N., Paolini, G., Della Croce, U., Magnuson, A., & Riad, J. (2018). Knee joint sagittal plane movement in cerebral palsy: a comparative study of 2-dimensional markerless video and 3-dimensional gait analysis. *Acta Orthopaedica*, *89*(6), 656–661. https://doi.org/10.1080/17453674.2018.1525195

Parrilla Eduardo, Ballestrer, A., Parra, F., Ruescas, A. V., Uriel, J., Garrido, D., & Alemany, S. (2019). *MOVE 4D: Accurate High-Speed 3D Body Models in Motion*. 30–32. https://doi.org/10.15221/19.030

Pellegrini, S., Schindler, K., & Nardi, D. (2008). A Generalisation of the ICP Algorithm for Articulated Bodies. *British Machine Vision Conference*. https://api.semanticscholar.org/CorpusID:12104382

Picerno, P., Cereatti, A., & Cappozzo, A. (2008). Joint kinematics estimate using wearable inertial and magnetic sensing modules. *Gait & Posture*, *28*(4), 588–595. https://doi.org/https://doi.org/10.1016/j.gaitpost.2008.04.003

Rodda, J., & Graham, H. K. (2001). Classification of gait patterns in spastic hemiplegia and spastic diplegia: a basis for a management algorithm. *European Journal of Neurology*, *8*(s5), 98–108. https://doi.org/https://doi.org/10.1046/j.1468-1331.2001.00042.x

Saboune, J., & Charpillet, F. (2005). Markerless Human Motion Capture for Gait Analysis. *CoRR*, *abs/cs/0510063*. http://arxiv.org/abs/cs/0510063

Salvi, M., & Molinari, F. (2018). Multi-tissue and multi-scale approach for nuclei segmentation in H&E stained images. *BioMedical Engineering OnLine*, *17*(1), 89. https://doi.org/10.1186/s12938-018-0518-0

Sangeux, M., Passmore, E., Graham, H. K., & Tirosh, O. (2016). The gait standard deviation, a single measure of kinematic variability. *Gait and Posture*, *46*, 194–200. https://doi.org/10.1016/j.gaitpost.2016.03.015

Stebbins, J., Harrington, M., Thompson, N., Zavatsky, A., & Theologis, T. (2006). Repeatability of a model for measuring multi-segment foot kinematics in children. *Gait and Posture*, *23*(4), 401–410. https://doi.org/10.1016/j.gaitpost.2005.03.002

Stricker, M., Hinde, D., Rolland, A., Salzman, N., Watson, A., & Almonroeder, T. G. (2021). Quantifying step length using two-dimensional video in individuals with Parkinson's disease. *Physiotherapy Theory and Practice*, *37*(1), 252–255. https://doi.org/10.1080/09593985.2019.1594472

Surer, E., Cereatti, A., Grosso, E., & Croce, U. Della. (2011). A markerless estimation of the ankle-foot complex 2D kinematics during stance. *Gait and Posture*, *33*(4), 532–537. https://doi.org/10.1016/j.gaitpost.2011.01.003

Van den Herrewegen, I., Cuppens, K., Broeckx, M., Barisch-Fritz, B., Vander Sloten, J., Leardini, A., & Peeraer, L. (2014). Dynamic 3D scanning as a markerless method to calculate multi-segment foot kinematics during stance phase: Methodology and first application. *Journal of Biomechanics*, *47*(11), 2531–2539. https://doi.org/10.1016/j.jbiomech.2014.06.010

Xu, Y., Zhu, X., Shi, J., Zhang, G., Bao, H., & Li, H. (2019). *Depth Completion from Sparse LiDAR Data with Depth-Normal Constraints*. http://arxiv.org/abs/1910.06727

Yamamoto, M., Shimatani, K., Hasegawa, M., Kurita, Y., Ishige, Y., & Takemura, H. (2021). Accuracy of Temporo-Spatial and Lower Limb Joint Kinematics Parameters Using OpenPose for Various Gait Patterns with Orthosis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 2666–2675. https://doi.org/10.1109/TNSRE.2021.3135879

Yeung, L. F., Yang, Z., Cheng, K. C. C., Du, D., & Tong, R. K. Y. (2021). Effects of camera viewing angles on tracking kinematic gait patterns using Azure Kinect, Kinect v2 and Orbbec Astra Pro v2. *Gait and Posture*, *87*, 19–26. https://doi.org/10.1016/j.gaitpost.2021.04.005

Chapter 3

3. AI-based approaches for clinical movement analysis

3.1 Introduction

The methods and the results presented in this chapter have been published in (Balta, Kuo, Wang, Porco, Morozova, Schladen, Cereatti, et al., 2022a; Balta, Kuo, Wang, Porco, Schladen, Cereatti, et al., 2022; Balta, Kuo, Wang, Porco, Schladen, Cereatti, Lum, et al., 2022)

In this chapter, the second category of methods, those associated with the AI-based approach, will be presented. These markerless (MS) techniques for motion analysis use artificial intelligence (AI) to analyze human movement without the need for physical markers making them ideal for situations requiring rapid deployment and adaptability. Convolutional Neural Networks (CNNs) are employed across the majority of these methods to infer poses from images or video frames directly. These networks analyze pixel data to predict joint locations, thus creating a real-time skeletal map of the human body. AI-based approaches, enhanced by deep learning, are capable of automatically learning from large datasets and interpreting complex motion patterns. These methods are generally model-free, meaning they can estimate joint centers positions directly from the data without relying on a predefined model. However, to improve accuracy, some AI approaches could incorporate model-based techniques. In this latter category, there are body tracking software development kits (SDKs) integrated with proprietary

RGB-D camera (i.e. Kinect v2 or Azure Kinect) (Clark et al., 2013; Ma et al., 2019; Romeo et al., 2021) as well as commercial software such as Theia Markerless.

Body tracking software development kits (SDKs) integrated with RGB-D cameras have been typically designed for animation or gaming for ease of implementation. They offer plug-and-play functionality that allows for quick setup without the need for extensive customization or deep understanding of the underlying algorithms. This makes them highly user-friendly and efficient for immediate use. Furthermore, these systems are specifically optimized for certain applications such as animation and gaming, where they excel by providing real-time feedback and interaction.

The main disadvantage of these methods is their non-compliance with clinical standards and terminology (Clark et al., 2019). A significant limitation is their "black box" nature, which prevents the fine-tuning of model parameters for specific pathological conditions, adversely affecting both external validity and performance (Büker et al., 2023). Additionally, body tracking SDKs are designed for specific hardware solutions, which hampers their ability to be generalized.

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

Within this general context, the first part of this chapter aims: (*i*) at investigating whether motion tracking through the body tracking SDK integrated into the Azure Kinect DK could be employed to perform gait analysis for clinical purposes and (*ii*) at comparing the performance of the above-mentioned SDK to a 2D deterministic model-based approach proposed by (Balta et al., 2020, 2023) and previous explained in the Chapter 2 – “Concurrent validation of a 2D model-based approach on individuals with Cerebral Palsy based on a single RGB-D camera.”

Open-source model-free methods that use deep learning, such as AlphaPose, OpenPose, DeepLabCut (D’Antonio et al., 2021; Lv et al., 2021; Mathis et al., 2018; Moro et al., 2020; Moro, Marchesi, et al., 2022; Ruescas-Nicolau et al., 2024; Stenum et al., 2021; Yamamoto et al., 2021), often rely on real or synthetic movement data during the training phase. The training datasets employed typically do not meet clinical analysis standards (Wu et al., 2002), lacking clear anatomical or functional guidelines for defining joint centers (Cereatti et al., 2010) and these datasets usually do not include individuals with disabilities. Despite those issues, open-source methods that utilize deep learning are distinguished from black-box alternatives by their adaptable hyperparameters, which can be finely tuned to meet the specific needs of various applications. In particular, an essential aspect of deploying these algorithms in new environments is the transfer learning phase. This process is crucial when the original training datasets do not include diverse representations, particularly of individuals with impairments or unique movement patterns. Transfer learning involves taking a deep-learning algorithm that has been initially trained on a broad dataset and refining it on a more targeted dataset that has been manually annotated. This manual annotation, which involves labeling key points like joint centers on video, enables the algorithm to learn specific features relevant for specific clinical applications, thereby enhancing its performance.

For all the reasons mentioned above, AI model-free approaches are particularly suitable for the analysis of infants' general movements within a home environment to identify movement disorders as early as possible. Deterministic model-based approaches, as detailed in Chapter 2, often rely on fitting algorithms that map predefined models to the recorded movements. However, due to the small physical size of infants, the performance of these computer vision algorithms can deteriorate. Additionally, to obtain a subject-specific model, at least two static acquisitions are required. This requirement is especially challenging because, unlike adults, infants are uncooperative during data acquisition and cannot follow instructions or strike poses on demand. This uncooperativeness complicates the process of obtaining static poses necessary to calibrate a subject-specific model. In contrast, AI model-free systems do not rely on predefined models and static poses, making them more adaptable and practical for use with infants.

Currently, the gold-standard for the early identification of movement disorders is the General Movement Assessment (GMA) (Heinz et al., 1997) which necessitates extensive training and validation for assessors making it unsuitable to be implemented in a patient's home. Since an early intervention depends on early identification, an automatic home monitoring is particularly crucial for identifying childhood neuromotor disorder. Recent advancements in computer vision techniques have significantly enhanced the automated analysis of infant movements, building on over a decade of research focused on 2D video analysis (Adde et al., 2010; Ihlen et al., 2020;

Moro, Marchesi, et al., 2022; Stagni et al., 2023). However, 3D recording based on a multiple-camera setup offers additional benefits such as higher spatial resolution and accuracy, though its widespread use has been constrained by high costs and computational demands (Marcroft et al., 2015). A commercial RGB-D sensor that combines an RGB camera with a depth sensor could provide a low-cost, compact solution suitable for both clinical and home settings, facilitating ongoing, naturalistic assessments. The adoption of AI model-free MS approaches, combined with an RGB-D camera that allow for 3D estimation, holds significant potential for shifting the analysis from the clinic to the subject's home. This approach makes continuous monitoring and assessment of infant movements more practical and convenient.

For this reason, the purpose of the second part of this chapter is to propose a novel MS protocol for infants' upper body movements analysis based on a single RGB-D camera that features a simplified instrumental setup, suitable for home use, to a purposely developed algorithm for 3D pose estimation and general movements (GM) metrics extraction.

To focus the reader's attention to the importance of these algorithms and the necessity of investigating their clinical applicability, the first paragraph of this chapter provides an introduction about the state of the art.

3.2 State of the art

In this paragraph, some of the main AI-based algorithms primarily used in the literature for movement analysis, or that have been utilized within this thesis, are listed, and described. Some of these employ a multi-camera approach, while others use only a single camera.

Theia markerless

Theia3D (Theia Markerless Inc., Kingston, ON, Canada) is a commercial AI model-based software which employs a deep learning-based MS motion capture method that uses synchronized video data for 3D human pose estimation (Kanko et al., 2021). This system leverages deep convolutional neural networks, trained on over 500,000 manually annotated digital images of humans in various settings, to track fifty-one key features in new images fed into the system. This technique allows the system to identify the 2D positions of these points of interest in the provided video data. Then, a tailored articulated multi-body model is adjusted to fit these features positions in three dimensions, and an inverse kinematic multi-body optimization approach is utilized to estimate the 3D pose of the subject during the recorded physical activities.

Azure Kinect body tracking software development kit (SDK)

The Azure Kinect Body Tracking SDK (Romeo et al., 2021) employs advanced techniques to estimate 3D coordinates of human body joints, integrating deep learning algorithms to achieve accurate and robust pose estimation. The body tracking SDK allows for the creation of a skeleton composed of thirty-two points of interest, including the feet and hands (**Figure 53**). This estimation can also be performed in real-time. The SDK provides estimates of 3D joint centers along with their associated confidence level which represent the probability that a given estimate is accurate or not.

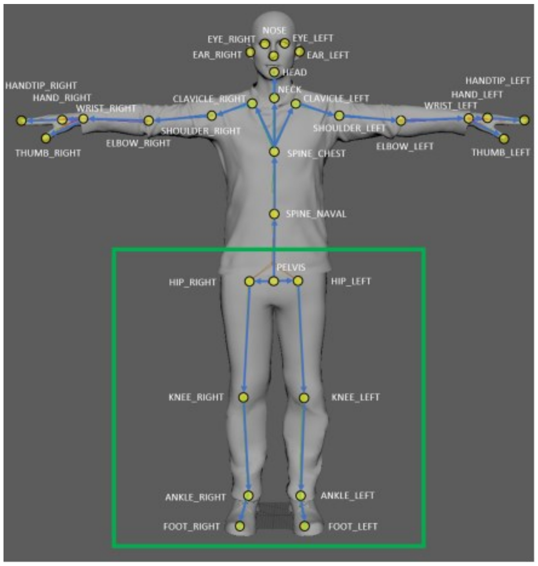


Figure 53. Joint positions and skeleton provided by the Azure Kinect SDK.

While specific implementation details are proprietary and not fully disclosed by Microsoft, the general approach incorporates a deep learning approach to estimate 3D joint centers coordinates. The principal steps are presented in detail and shown in **Figure 54**:

1) Initial Pose Estimation Using Deep Learning: The process starts with the CNN analyzing the RGB and Depth data to identify and estimate the initial positions of body joints. This step provides a rough estimation of the pose by predicting the 3D locations of various joints, such as the elbows, knees, and spine, based on the input data from the Kinect's sensors;

2) Estimates refinement: After the initial pose estimation, predefined information about the joints and the physical constraints between them, such as bone lengths and joint limits was introduced ensuring that the estimated poses are physically plausible;

3) Optimization Algorithm (model fitting): The system then applies an optimization algorithm to fit a predefined skeletal model to the observed data. This step adjusts the positions and orientations of the joints of the model to minimize the discrepancy between the observed joint locations (as estimated by the CNN) and the model predictions;

4) Constraints and Regularization: The optimization algorithm considers also biomechanical constraints (like joint angles and limb lengths) and may use regularization techniques to prevent overfitting and ensure smooth and realistic movements. This helps in handling occlusions or ambiguities in the depth data where CNN’s predictions might be less reliable.

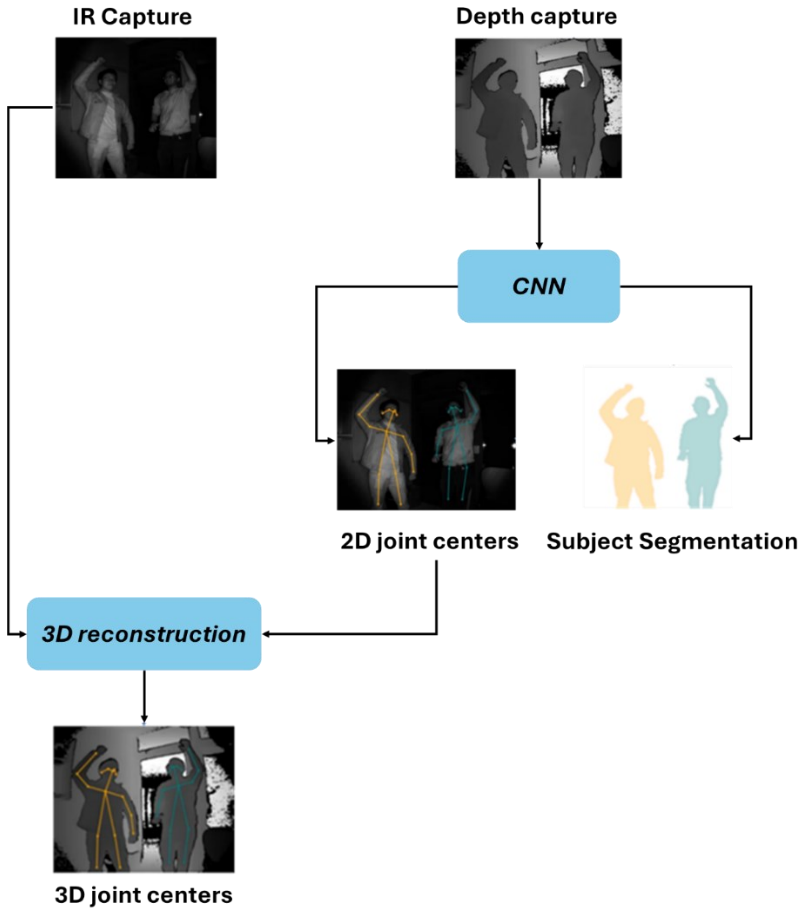


Figure 54. Block diagram of the Body Tracking SDK of the Azure Kinect

For the study of general movements in infants, DeepLabCut (DLC) was used which is an advanced version of DeepCut. This progression includes several key developments: starting from the original DeepCut, advancing through DeeperCut, which improved the network's accuracy and robustness, and culminating in DLC, which enhances usability through a dedicated interface. In the next section, the different steps from the original DeepCut, through DeeperCut, to DLC are outlined.

DeepCut

DeepCut (Rajchl et al., 2016) utilizes CNNs to detect keypoints in images, producing confidence maps for each body part. These maps are 2D images where each pixel's value indicates its likelihood of representing a specific keypoint. This method marks a significant advancement in computer vision by integrating detection and pose estimation, which improves accuracy and reliability in complex scenes with multiple individuals.

Unlike traditional methods that require detecting individuals before estimating their poses, DeepCut combines these tasks. It simultaneously infers the number of people in a scene, identifies occluded body parts, and resolves ambiguities between body parts of individuals close to each other.

DeepCut proposes a partitioning and labeling formulation using an Integer Linear Program that optimally arranges detected body parts while respecting geometric and visual constraints, ensuring anatomical accuracy. This formulation processes hypotheses generated by CNN-based part detectors, performing non-maximum suppression to select and group the most probable configurations of body parts.

The methodology behind DeepCut involves an innovative application of Integer Linear Program to solve the joint partitioning and labeling problem. This problem selects a subset of body parts from a pool of candidates, labels each part according to its class, and groups them based on the individuals they belong to. The Integer Linear Program framework excels at identifying optimal solutions that maximize the overall confidence of the detected parts, assembling them into coherent human poses.

DeeperCut

DeeperCut (Insafutdinov et al., 2016) improves upon the DeepCut algorithm by leveraging recent advancements in machine learning to enhance human pose estimation, particularly in scenarios involving multiple people. This algorithm integrates several key enhancements:

- "Deeper" Detection: DeeperCut uses robust body part detectors based on the ResNet (He et al., 2015) architecture, a deep convolutional neural network known for its effectiveness in complex visual recognition tasks. These ResNet-based detectors help DeeperCut in achieving competitive performance on pose estimation benchmarks;
- "Stronger" Connectivity: This algorithm incorporates image-conditioned pairwise terms directly within the ResNet layers. This integration allows for the accurate and efficient association of body parts within CNN's

workflow, improving the algorithm's ability to manage scenes with multiple individuals.

- "Faster" Performance: By embedding pairwise part-to-part predictions within the fully-connected ResNet network, DeeperCut significantly reduces processing time—cutting it down by two to three orders of magnitude.

DeeperCut's use of ResNet provides a large receptive field, enabling accurate predictions of body part locations and facilitating the modeling of spatial relationships between adjacent parts. This capability is crucial for generating pairwise probabilities that enhance multi-person pose estimation. By eliminating separate post-processing stages like Integer Linear Program and integrating critical functionalities directly into the ResNet architecture, DeeperCut offers an efficient and robust approach to pose estimation. This makes it highly effective for accurately capturing human poses in environments with multiple interacting individuals.

DeepLabCut

DeepLabCut (DLC) (Mathis et al., 2018) is an advanced machine learning tool designed for animal pose estimation. Developed by researchers from the Max Planck Institute of Neurobiology and Harvard University, DLC leverages deep learning to understand and predict the posture of animals by tracking key body parts.

A distinctive feature of DLC is its implementation of transfer learning. This technique involves starting with a CNN that has been pre-trained on a broad dataset, which may not be directly related to the current task, and then fine-tuning this algorithm for specific applications such as human pose estimation using a much smaller set of targeted data. This method drastically cuts down the necessary data volume and computational effort required to reach high levels of accuracy.

To customize the network for particular research needs, scientists manually annotate a collection of images by identifying on each image the relevant body parts. These annotated images are then used to refine the pre-trained network, enabling it to recognize these parts more effectively in similar situations. For training DLC, it is necessary to select from a large initial dataset only the most relevant images for the manual annotation. Using techniques like k-means clustering, researchers can group similar images, facilitating the selection of a diverse and representative subset of images for training. This step ensures that the dataset encompasses a broad array of poses and contexts, enhancing the network's ability to generalize across different scenarios.

DLC also includes a user-friendly graphical user interface, and it is available as a software package compatible with various operating systems. This makes it especially accessible to researchers who may not have extensive expertise in machine learning, allowing them to apply sophisticated pose estimation techniques in their studies.

3.3 2D gait analysis based on Azure Kinect body tracking SDK

Most recently, several companies are producing inexpensive RGB-D cameras (e.g. Microsoft Kinect, IntelRealSense D435) that come with SDK for the real-time tracking of body position and orientation.

In 2013, Microsoft released the second version of Kinect (Kinect v2) which included a machine learning approach for tracking joint centers that have been trained on massive amounts of labeled depth data. In particular, a randomized Decision Forest (an ensemble of decision trees) was implemented to process the depth data (Shotton et al., 2013.). Each tree in the forest classifies each pixel in the image as part of a particular body part based on its depth value. By processing the depth images through these decision trees, the system can label each pixel as being part of a specific joint or body segment. Once initial estimates of joint locations are made through the decision trees, the Kinect v2 SDK refines these estimates using the mean shift algorithm—a robust technique used for finding the peaks of a density function. This helps in accurately localizing joint centers by clustering nearby data points that are labeled as belonging to the same joint. Moreover, the mean shift algorithm helps in smoothing the data around the estimated joint locations, reducing the noise from the decision tree predictions, and enhancing the precision of the final joint location estimates. However, this technology was primarily focused on gaming purposes.

Recently (in 2020), Microsoft released a new RGB-D camera (Azure Kinect) which, compared to the previous versions of Kinect, is targeted toward other markets such as logistics, robotics, health care, and retail. The Azure Kinect includes an IR sensor for distance estimation that has greater accuracy than its predecessors and a novel motion tracking algorithm (body tracking SDK) for the estimation of the body joints’ 3D positions and orientations which is based on deep neural networks (i.e. CNN).

In this context, the purpose of this study is twofold:

1. To investigate the use of the body tracking SDK integrated into the Azure Kinect DK for conducting clinical gait analysis;

2. To compare the performance of the body tracking SDK with a 2D deterministic model based approach developed by (Balta et al., 2020, 2023) and explained in Chapter 2 – “Concurrent validation of a 2D model-based approach on individuals with Cerebral Palsy based on a single RGB-D camera.”.

3.3.1 Materials and methods

A. Subjects: Gait data were collected from five healthy subjects.

B. Data collection and subject preparation were the same described in Chapter 2 “Concurrent validation of a 2D model-based approach on individuals with Cerebral Palsy based on a single RGB-D camera.” Azure Kinect camera was used (RGB images: 720 × 1280 pixels at 30 fps, Depth images: 640 × 576 pixels at 30 fps).

C. Validation – A 12-camera stereo-photogrammetric system (Vicon-Vero) was utilized to collect 3D reference data at a rate of 100 fps. Sixteen retro-reflective spherical markers, each 14 mm in diameter, were attached to the subjects following the Davis protocol (Davis et al., 1991). The calculations of 3D reference joint angles were conducted using Nexus software using the “*Plug-in-Gait*” lower limb model.

The acquisitions were not performed synchronously by the MB based and MS systems as interferences in the depth map reconstruction were observed in the Azure Kinect recordings. The wavelength of the Azure Kinect IR sensor is the same as the Vicon Vero system (850 nm) and this resulted in extremely poor quality depth images with many invalidated pixels. Black uninformative pixels were present in the synchronous acquisitions, particularly in correspondence with the positions of the reflective markers. As can be seen in **Figure 55** the depth information about the foreground leg, which is the leg under investigation, is very poor thus affecting the joint center’s position estimations.

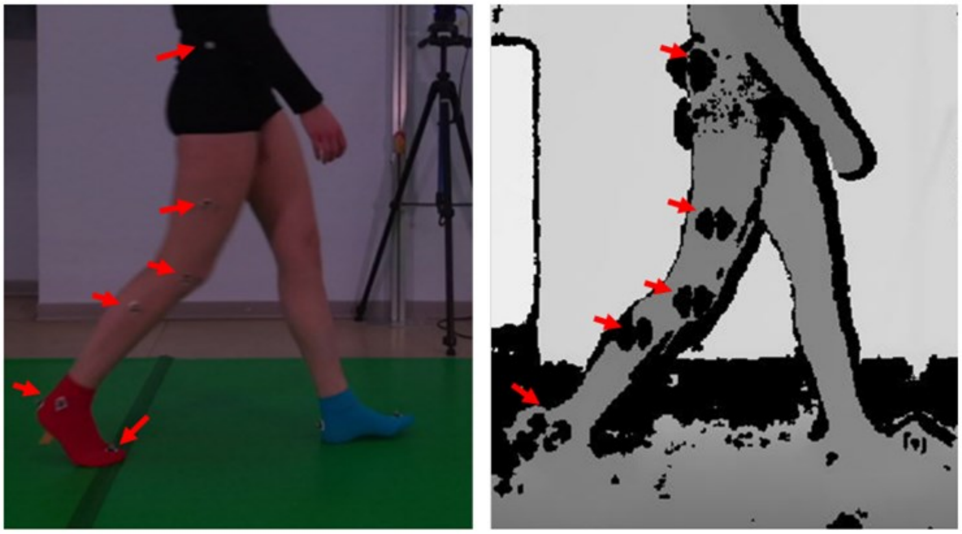


Figure 55. Illustration of the RGB (left) and the depth (right) images captured with the Azure Kinect during synchronous recording with the Vicon system. Black invalidated pixels can be seen particularly in correspondence of the reflective markers (red arrows).

For this reason, the same trial was repeated twice to be acquired separately with the two systems (Azure Kinect and Vicon.Vero) under the hypothesis of repeatability of the gesture.

3.3.2 Data Processing

Using the body tracking SDK (SDK), the coordinates for the hip joint (${}^{I}\boldsymbol{HJC}$), knee joint (${}^{I}\boldsymbol{KJC}$), ankle joint (${}^{I}\boldsymbol{AJC}$), and the toe of the foot (${}^{I}\boldsymbol{TOE}$) were identified. Joint kinematics were evaluated based on the inclination of segments created by connecting these joint centers. Specifically, the plantar-dorsi flexion angle of the ankle is determined by the angle between the ${}^{I}\boldsymbol{KJC}$ - ${}^{I}\boldsymbol{AJC}$ and ${}^{I}\boldsymbol{AJC}$ - ${}^{I}\boldsymbol{TOE}$ vectors. The knee joint's flexion-extension angle is calculated from the angle between the ${}^{I}\boldsymbol{HJC}$ - ${}^{I}\boldsymbol{KJC}$ and ${}^{I}\boldsymbol{KJC}$ - ${}^{I}\boldsymbol{AJC}$ vectors. Furthermore, the flexion-extension angle of the hip joint is assessed by measuring the angle between the ${}^{I}\boldsymbol{HJC}$ - ${}^{I}\boldsymbol{KJC}$ vector and the horizontal-axis.

Joint angles of 3D MB recordings were extracted through the Vicon-Nexus software.

Joint angles using the 2D deterministic model-based protocol proposed by Balta and colleagues (MLM) were computed as described in Chapter 2 – paragraph “Concurrent validation of a 2D model-based approach on individuals with Cerebral Palsy based on a single RGB-D camera.”

An overview of the comparison between MLM, SDK and MB protocol is shown in **Figure 56**.

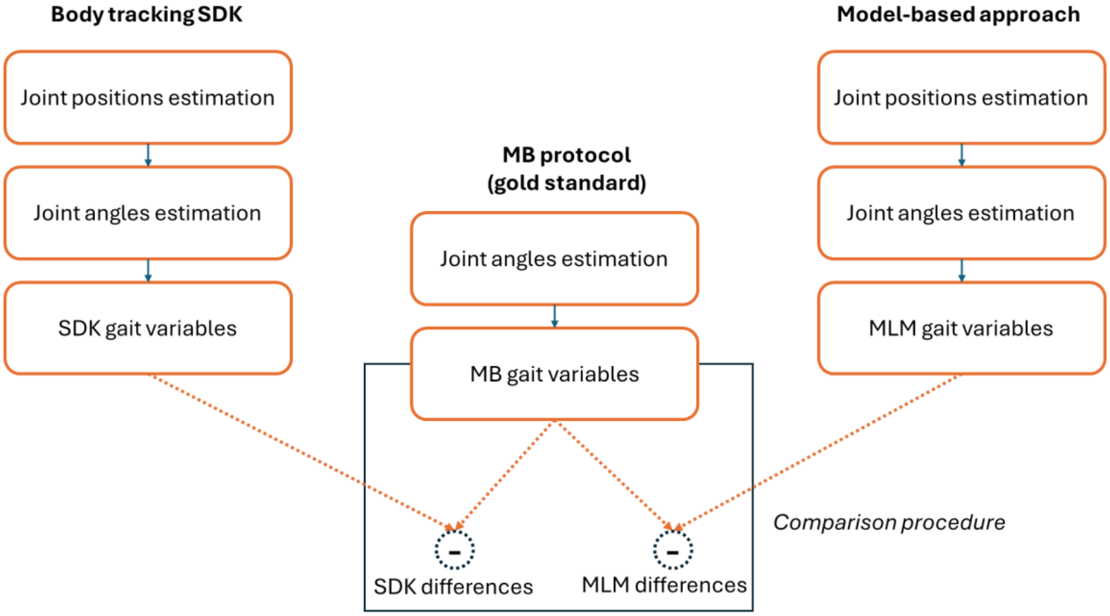


Figure 56. Overview of the comparison procedure between MLM, SDK and MB protocol.

3.3.3 Performance assessment and statistical analysis

Prior to comparison, the kinematic curves from both the MLM, SDK and MB protocols were processed using a fourth-order Butterworth filter with a cutoff frequency of 7 Hz and then time-normalized to the gait cycle (1-100%).(Bergamini et al., 2014). A set of clinically relevant key gait features were extracted according to (Benedetti et al., 1998) from the MB, MLM and SDK sagittal lower limb joint kinematics as detailed in Chapter 2-paragraph “Concurrent validation of a 2D model-based approach on individuals with Cerebral Palsy based on a single RGB-D camera.”.

For each gait feature and each method (MLM, SDK and MB), the mean values and the standard deviation across trials and subjects were calculated.

The performance of MLM and SDK methods were assessed in terms of mean difference (MD) compared to the MB system for each gait feature.

3.3.4 Results

Results related to the extracted key gait features in terms of mean and standard deviation for MLM, SDK and MB protocols, are reported in **Table 11**.

Results related to the extracted key gait features in terms of mean difference for MLM and SDK methods with respect to MB protocol are summarized in **Table 12**.

Table 11. Mean and standard deviation of the key gait features over ten trials per five subjects computed for the marker-based (MB), deterministic model-based markerless (MLM) and body tracking SDK methods.

Gait variables (deg)		*MLM mean (SD)*	*SDK mean (SD)*	*MB mean (SD)*
Knee	*Initial Contact*	10.5 (5.2)	7.4 (2.5)	6.1 (4)
	Load	18.6 (6)	13.5 (6)	14.2 (4.8)
	Stance	5.7 (5.9)	7.7 (3)	6.3 (3.1)
	Swing	64.5 (4.6)	59.7 (4.8)	64 (2.6)
Ankle	*Stance*	7.2 (2.6)	46.5 (4)	18.4 (3.2)
	Swing	-23.4 (6.9)	25.7 (3.6)	-13.9 (8.5)
Hip	*Stance*	-12.4 (3.8)	-16.8 (3.1)	-9.1 (4.6)

Table 12. Mean difference (MD) of the key gait features averaged over ten trials per five subjects computed for deterministic model-based markerless (MLM) and body tracking SDK methods with respect to MB protocol

Gait Variables (deg)		***MLM MD***	***SDK MD***
Knee	***Initial Contact***	3.9	1.3
	Load	4	-0.7
	Stance	-0.6	1.4
	Swing	0.5	-4.3
Ankle	***Stance***	-11.2	28.1
	Swing	-9.5	39.6
Hip	***Stance***	-3.3	-7.7

The resulting sagittal hip, knee and ankle angles extracted from MLM, SDK and MB systems estimated for ten gait trials for one subject are reported in **Figure 57** as percentage of the gait cycle.

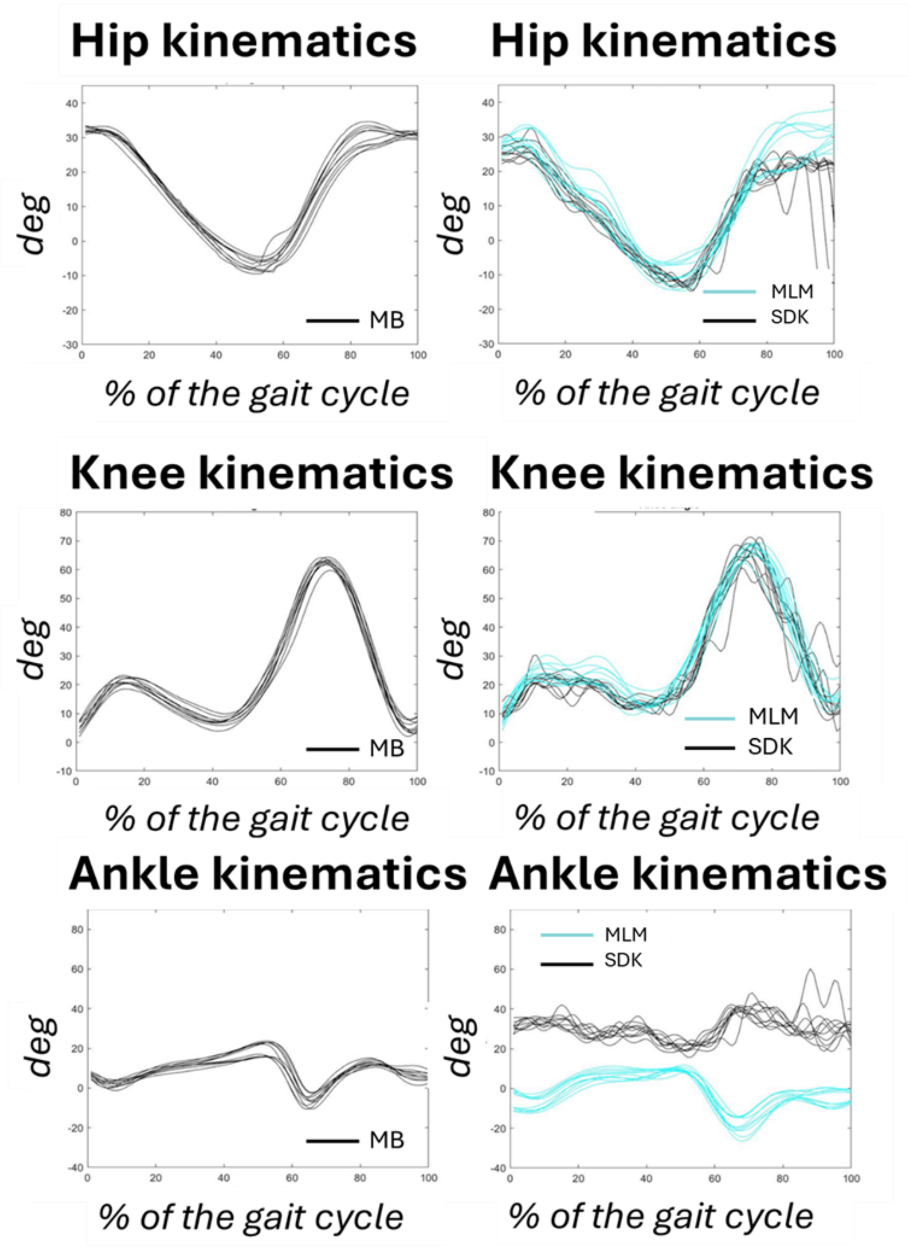


Figure 57. Illustration of sagittal hip, knee and ankle angles of ten gait trials for a single subject estimated with the marker-based (black line - left), deterministic model-based (blue line - right) and body tracking SDK (black line – right) methods.

3.3.5 Discussions

The differences between the two MS protocols and MB system are partially due to the different protocols used. Both MS systems (MLM and SDK) calculate joint angles using simple trigonometric formulas by connecting the joint centers, while MB system calculates the joint kinematics through the decomposition of Euler angles.

The overall differences between the gait features estimated with the MB and those estimated with the body tracking SDK ranged from -0.7 deg in the estimation of the knee max flexion in the load phase to 39.6 deg in the estimation of the ankle max dorsiflexion in swing phase. The differences between the gait features estimated with the MB method and those estimated with the MLM ranged from -0.6 deg in the estimation of the knee max extension in the stance phase to -11.2 deg in the estimation of the ankle max dorsiflexion in stance phase. The standard deviation of the gait features grand mean ranges from 2.5 deg to 4.8 deg for the MLM method and from 2.6 deg to 8.5 deg for the SDK method. The variability was high for the ankle in the swing phase both for the MLM and the MB methods with SD values of 6.9 deg and 8.5 deg, respectively. The lower variability was found for the MB method in the knee feature during the swing phase (SD = 2.6 deg), for the MLM method in the ankle in stance phase (SD = 2.6 deg), and the SDK method in the knee parameter at the initial contact (SD = 2.5 deg).

The SDK differences with respect to the MB method for the ankle in stance and swing phase were respectively 28.1 deg and 39.6 deg while the MLM method underestimates the MB values of 11.2 deg and 9.5 deg. These differences can be attributed to how the two methods define the foot segment for the ankle angle computation. While the MLM method defines the foot segment starting from the LM position as the segment that best fits the segmented foot, in the SDK method the foot inclination was defined starting from the ankle coordinate to the toe coordinate, which is not representative of the actual foot inclination (**Figure 59**). The MLM performances were higher in the stance and swing phase of knee kinematics with respect to the SDK method. For what concerns the hip angles, both methods overestimate the hip flexion during the stance phase, the MLM showed good performances with a mean difference with respect to the MB estimations of only 3.3 deg while the SDK mean difference was 7.7 deg. The overall performances of the SDK are also affected by the fact that 2 out of 5 subjects showed left-right confusion during the gait cycle (the foreground limb was incorrectly identified as the background limb), as shown in **Figure 58**. This inevitably affected the computation of the joint angles.

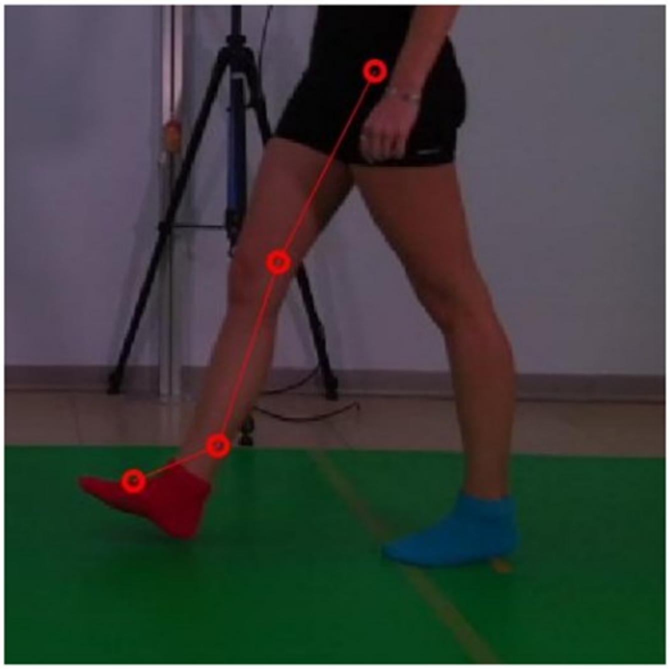


Figure 58. Example of wrong leg identification with the body tracking SDK method. Left hip, knee, ankle, and toe coordinates were misidentified as the right limb ones.

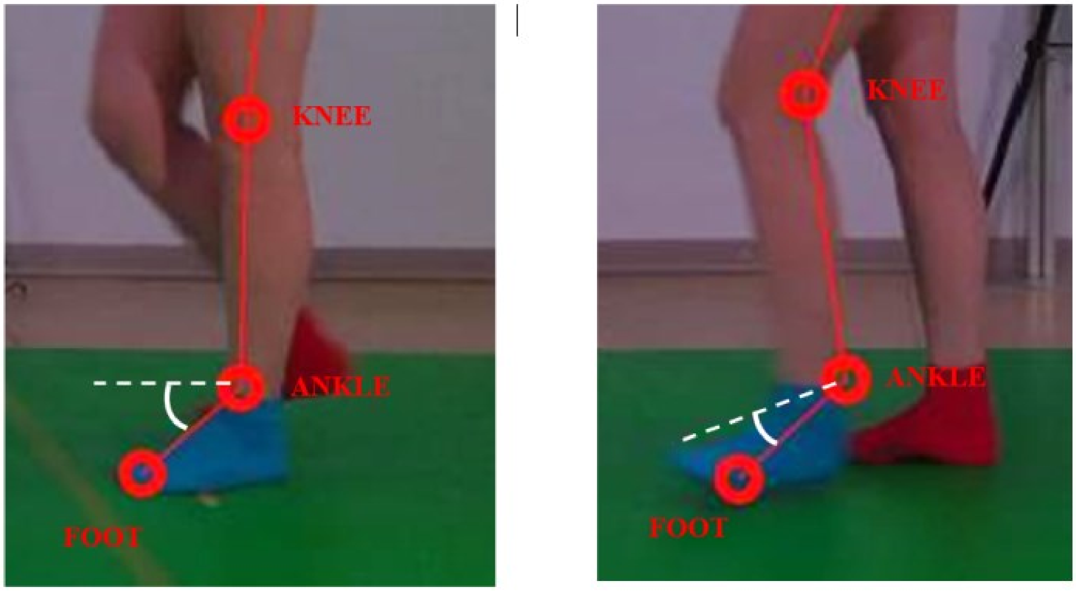


Figure 59. Illustration of ankle and foot 2D coordinates. The angle formed between ANKLE and FOOT is not representative of the actual foot inclination.

3.4 3D upper limb analysis on preterm infants in a home environment

3.4.1 Introduction

Cerebral palsy (CP) is defined as a group of neuromotor impairments resulting from brain injuries occurring around the time of birth, such as periventricular leukomalacia, intracerebral hemorrhage, infection, and infant stroke (Metz et al., 2022). Research including a systematic review and meta-analysis (Oskoui et al., 2013) places the global prevalence of CP at approximately 2.11 per 1000 births. However, studies from different regions like Africa (El-Tallawy et al., 2014), Asia (Wang et al., 2021), and North America (Christensen et al., 2014) suggest an increasing trend, exceeding three cases per 1000 births, potentially linked to higher survival rates of early, preterm infants (Graham et al., 2016). Typically, CP is diagnosed between 12–24 months in high-income countries, but this can extend up to five years in lower-resource settings (Novak et al., 2017), often delayed by factors like the absence of definitive biomarkers, the potential distress of false positives for families, and the lack of cure (Te Velde et al., 2019).

One significant advancement in early detection has been the general movement assessment (GMA), an visual evaluation tool that gained prominence as the importance of infants' spontaneous movements was recognized towards the end of the 20th century (Heinz et al., 1997; Prechtl, 1990). The GMA identifies two key movement patterns - cramped-synchronized movements and the absence of fidgety movements between three to five months - which are strong indicators of later CP diagnosis (Einspieler & Prechtl, 2005). Although effective, the GMA requires extensive training for observers, making it challenging to implement widely in clinical settings (Silva et al., 2021).

This backdrop underscores the necessity for a broad, accessible screening method to enhance early intervention. Engaging families in the monitoring and therapeutic processes could facilitate earlier detection and intervention, particularly for infants showing subtle early symptoms who are at a higher risk of delayed diagnosis and treatment (Hekne et al., 2021).

The use of MB and MS systems for neuromotor assessment has been explored. Multi-camera, 3D MB analysis is well-established (Disselhorst-Klug et al., 2012; Mazzarella et al., 2020). In particular, in the literature, there are different studies which have paved the way to quantitatively evaluate spontaneous motor activity developments in infants at risk for developing CP (Disselhorst-Klug et al., 2012; Karch et al., 2012a; Meinecke et al., 2006).

However, its requirement for a lab setting limits its general application and physical attachments of markers could interfere with the natural movements as shown in **Figure 60**.

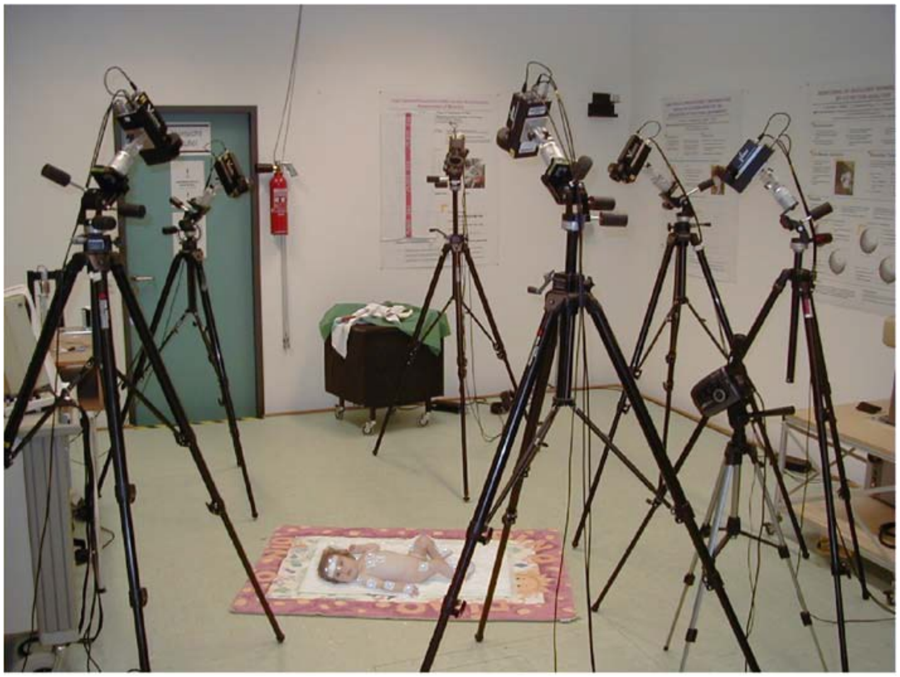


Figure 60. Marker-based system for the general movement analysis of newborns at risk for developing spasticity.

In contrast, MS approaches using single-camera setups promise greater accessibility and non-intrusiveness, maintaining the integrity of natural infant movements without the use of physical markers (Silva et al., 2021).

Recent advances have been made in computer vision techniques to automate the MS analysis of infant movements, with over a decade of research focusing on 2D video analysis (Adde et al., 2010; Ihlen et al., 2020; Moro, Pastore, et al., 2022; Stagni et al., 2023). However, 3D recording based on multi-camera setup offers additional benefits such as higher spatial resolution and accuracy, though its widespread use has been constrained by high costs and computational demands (Marcroft et al., 2015). Using a commercial RGB-D sensor that combines an RGB camera with a depth sensor could provide a low-cost, compact solution suitable for both clinical and home settings, facilitating ongoing, naturalistic assessments (Marcroft et al., 2015).

The current study uses such a sensor to track infants' upper body movements, applying DeepLabCut (Mathis et al., 2018), a AI model-free algorithm for 2D pose estimation, and subsequently a purposely developed method to reconstruct 3D trajectories of selected points of interest.

The second aim of this study is exploring the applicability of selected metrics for the early identification of movement disorders on eight infants recorded at home at 3,4 and 5 months of age and on a pair of twins with divergent health profiles.

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of The Catholic University of America (protocol #19-0012, initially approved 7 May 2019) for studies involving humans.

3.4.2 Method Description

The camera employed was an Intel RealSense D435 (fs = 30 Hz), an RGB-D sensor that integrates a high-definition RGB camera and a depth sensor. This camera captures detailed color images and depth-perceived images based on how far objects are from it, using stereoscopic vision technique. Although both image types were synchronized, some minor alignment issues remained between them.

A series of steps was executed to reconstruct 3D coordinates of selected points of interest from the acquired RGB-D videos and to compute the associated GM metrics (**Figure 61**).

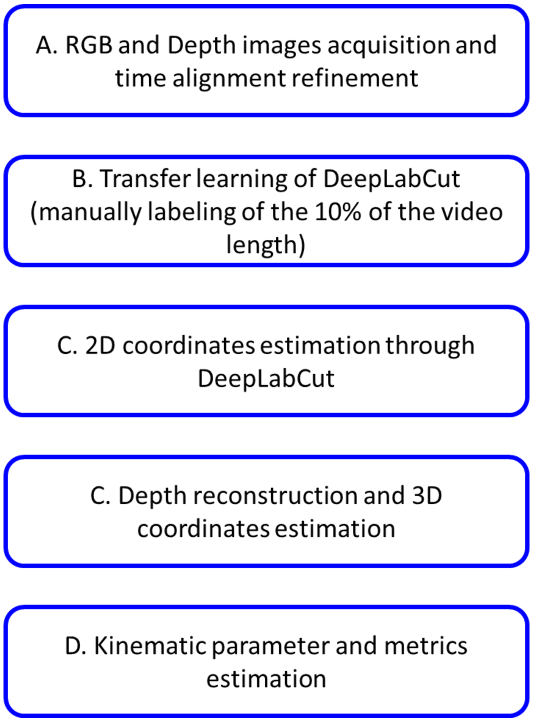

A. RGB and Depth images acquisition and time alignment refinement
B. Transfer learning of DeepLabCut (manually labeling of the 10% of the video length)
C. 2D coordinates estimation through DeepLabCut
C. Depth reconstruction and 3D coordinates estimation
D. Kinematic parameter and metrics estimation

Figure 61. Flow-chart of the proposed MS method.

1) RGB and Depth Images Acquisition and Time Alignment Refinement

The process of capturing images with the RGB camera and depth sensor required precise synchronization, as the frame rates were not consistently steady. Enhancements were made beyond the manufacturer's provided software using the timestamps from the acquisition software to address three specific alignment issues:

- When the timestamp of an RGB image was significantly closer to other RGB timestamps compared to the nearest depth image timestamp, a gap of the proper number of frames was inserted in the sequence of depth frames.
- In cases where a depth image timestamp was significantly closer to other depth timestamps compared to the nearest RGB image timestamp, a gap of the proper number of frames was inserted in the sequence of RGB frames.
- If the time difference between an RGB and a depth image timestamp was less than half of the nominal sampling interval (approximately 17 ms), the frames were considered sufficiently aligned.

All resulting gaps were reconstructed by using cubic spline interpolation.

2) 2D Tracking Algorithm

RGB images were first converted into video files using ImageJ (National Institute of Health, Bethesda, MD, USA) (Schindelin et al., 2015) and subsequently processed with the DeepLabCut (DLC) image processing tool (Swiss Federal Institute of Technology, Lausanne, Switzerland). DLC was specifically trained to detect six Points of Interest (PoIs) on the infant's upper body: left and right shoulders (LS and RS), elbows (LE and RE), and wrists (LW and RW). During the training process, PoIs were manually labeled on 10% of the frames of each video. These frames were selected by DLC through a k-means algorithm that chose frames based on variability in pixel characteristics. DLC then identified the PoIs in the RGB frames and provided confidence scores for each detection. When a PoI was occluded, it was assigned a low confidence score.

3) Depth Reconstruction and 3D Coordinates Estimation

We developed a method to obtain the 3D coordinates of PoIs tracked in the RGB images by utilizing depth sensor data. We addressed three main issues that could result in incorrect or undefined 3D PoI positions:

1) If the RGB location of a tracked PoI was over a "black area" in the corresponding depth image, which lacked depth information, this resulted in undefined depth coordinates (**Figure 62**a).
2) As the tracking algorithm relied solely on RGB data, the estimated location of a PoI might fall over another body segment in the RGB image, such as a shoulder covered by the head. To correct this, we used confidence levels from the DLC; depth values from frames with confidence levels below 0.6 were removed (**Figure 62**b).
3) Residual spatial misalignment between RGB and depth images could cause errors when estimating the depth coordinate of a tracked PoI, particularly near significant depth discontinuities. We calculated the first derivative of the PoI depth coordinate and removed values exceeding a threshold set according to physical motion limits of the subject (**Figure 62**c).

All gaps in depth coordinates were filled using cubic spline interpolation (**Figure 63**).

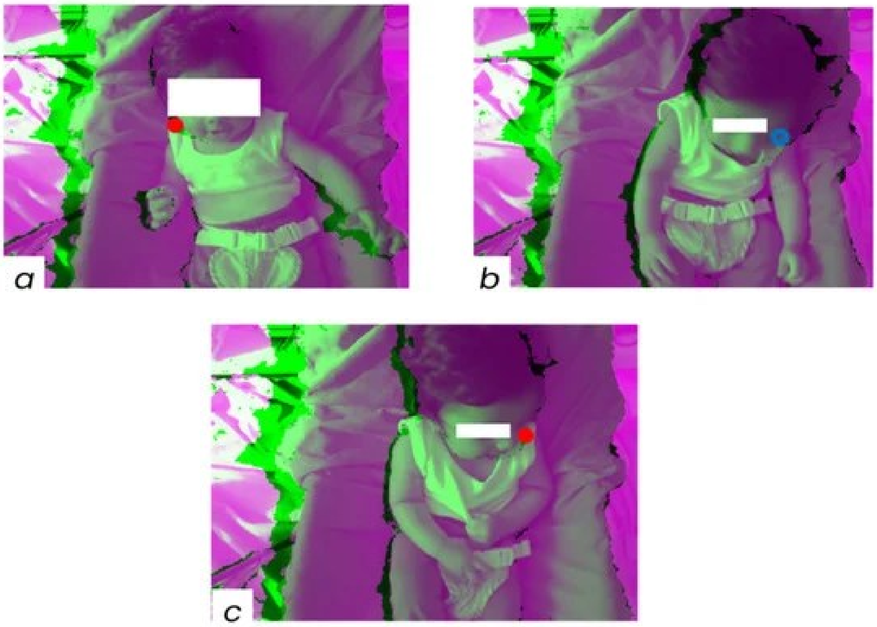


Figure 62. Factors Leading to Undefined 3D Positions of Points of Interest (PoIs): (a) Right Shoulder (RS) positioned on the 'black area', (b) occlusion of the Left Shoulder (LS) by the head, and (c) residual spatial misalignment between RGB and depth data. Colored circles indicate PoIs.

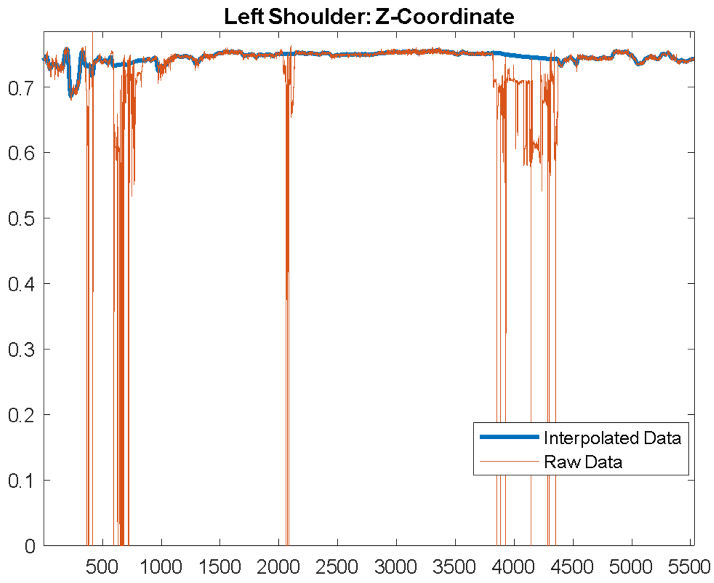


Figure 63. Cubic spline interpolation to fill all gaps in depth coordinate of Left shoulder.

4) Kinematic parameters and metrics estimation

This study also includes the computation of a subset of metrics proposed in the literature starting from these 3D trajectories.

From 3D trajectories of 6 PoIs, the following metrics for quantifying GMs were computed (Disselhorst-Klug et al., 2012; Kanemaru et al., 2013a; Karch et al., 2012b; Meinecke et al., 2006):

- *Metric 1: Area differing from the moving average trajectories*

Area where the wrist trajectories deviate from their moving average (Disselhorst-Klug et al., 2012), adjusted for the length of the moving average window (two seconds).

This parameter identifies deviations in a trajectory's movement from its moving average and describes the variability and diversity of the movements performed. In general, a smaller value indicates a lower variety of movements.

Figure 64 exemplifies the x-coordinate of the left wrist trajectory. The upper figure displays TD's coordinate, while the lower figure depicts that of an affected baby. As illustrated in **Figure 64**, the trajectory of the TD baby demonstrates a higher area deviating from the moving average.

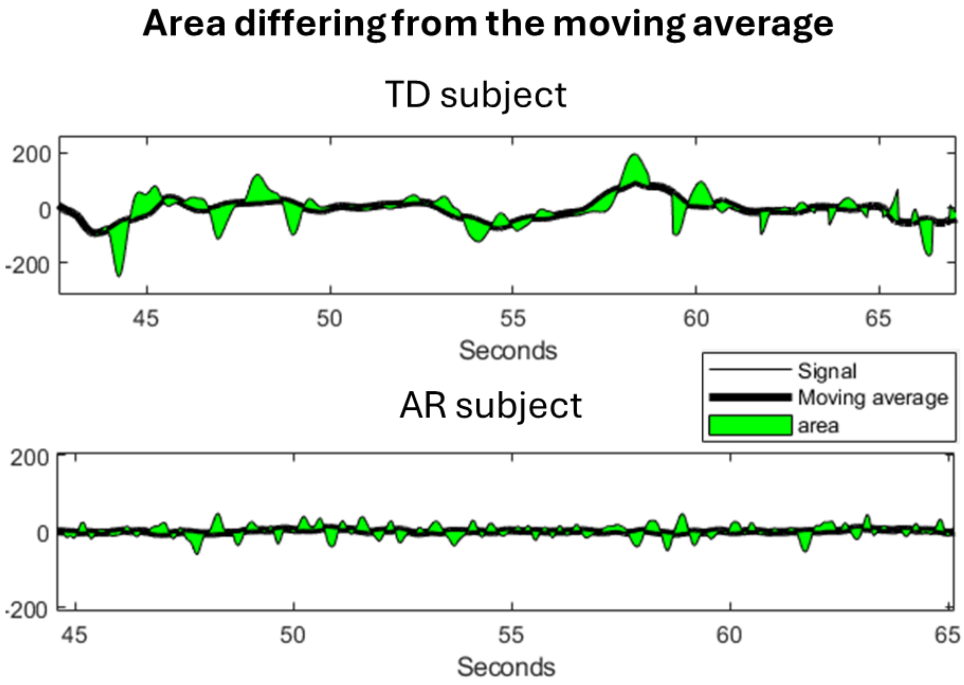


Figure 64: Area differing from the moving average of the x- coordinate of right wrist for TD (a) and an AR child (b)

- *Metric 2: Area out of the standard deviation of the moving average trajectories*

Area where wrist trajectories fell outside the standard deviation of their moving average (Disselhorst-Klug et al., 2012), adjusted for the duration in which the trajectories exceeded the standard deviation (specifics on normalization were not provided in the reference work). As for the previous parameter, this metric represents another way to characterize the variability of the movement.

Figure 65 exemplifies the x-coordinate of the left wrist trajectory. The upper figure displays TD's coordinate, while the lower figure depicts that of an AR child. As illustrated in **Figure 65**, the trajectory of the AR baby demonstrates a smaller area deviating from the standard deviation.

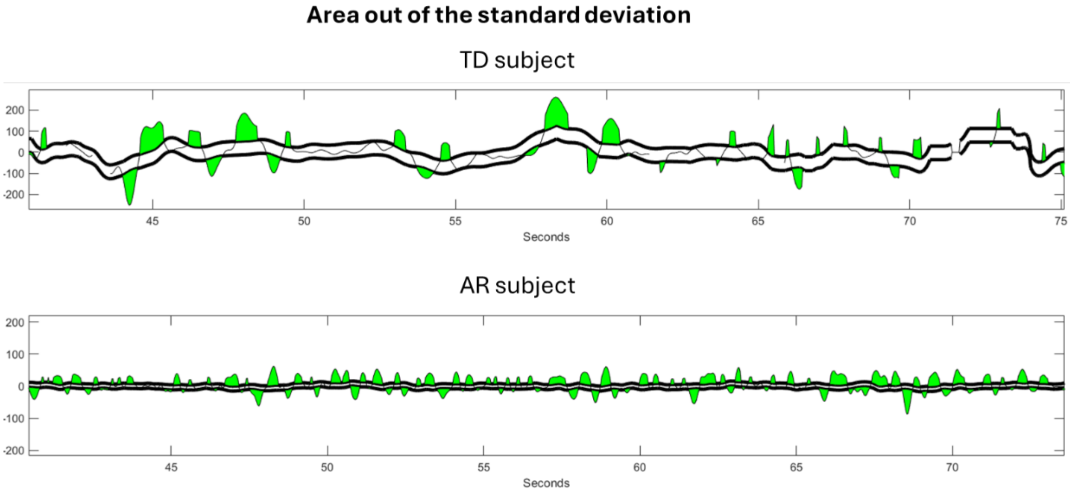


Figure 65. Area out of standard deviation of moving average – trajectory of the right wrist for a healthy (a) and an affected child (b)

- *Metric 3 and Metric 4: Periodicity in the wrist trajectories/velocities*

The movement of TD children typically exhibits a high degree of complexity, whereas the movement of an AR child with impairments tends to be more monotonous and repetitive, often showing a periodic pattern (**Figure 66**).

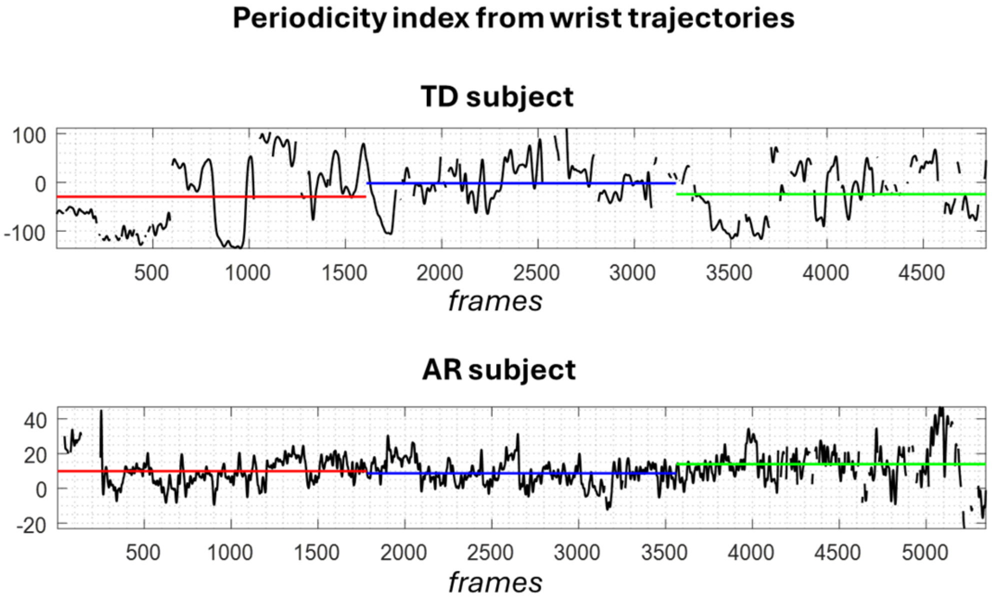


Figure 66. Periodicity index – trajectory of the right wrist for a TD and an AR subject.

To measure periodicity, the diagram includes the neutral axis and three horizontal lines, each representing the arithmetic mean for one-third of the measurement duration. According to the method by (Disselhorst-Klug et al., 2012; Meinecke et al., 2006), the first step is to determine the number of intersections between the trajectory and the arithmetic mean. Next, the distance between each pair of intersections, denoted as $d_{t,i}$ is calculated in terms of frame numbers. The mean and standard deviation of these distances, d_t and $\sigma_{t,d}$ respectively. The periodicity parameter is defined as:

$$P_t = \frac{1}{\sigma_{t,d,x} + d_{t,x}} + \frac{1}{\sigma_{t,d,y} + d_{t,y}} + \frac{1}{\sigma_{t,d,z} + d_{t,z}}$$

Periodicity was analyzed for both the left and right wrists along the three spatial axes, with these individual calculations being integrated into a single parameter. This parameter is designed to evaluate the regularity and frequency of the trajectory intersecting the mean lines, providing a comprehensive measure of movement periodicity.

- *Metric 5: Area differing from the moving average velocities*

Area where the wrist velocities deviate from their moving average (Meinecke et al., 2006), adjusted for the length of the moving average window (two seconds). The velocity of the AR baby is expected to have a smaller area deviating from the moving average with respect to the TD subject.

- *Metric 6: Area outside of the standard deviation of the moving average velocities*

Area where speed profiles of the wrists fell outside the standard deviation of their moving average (Disselhorst-Klug et al., 2012), adjusted for the duration in which the trajectories exceeded the standard deviation. The velocity of the AR baby is expected to have a smaller area deviating from the standard deviation with respect to the TD subject.

- *Metric 7: Skewness of the velocity of the wrists*

In a normal distribution, the skewness is zero, and typically, skewness values range between -3 and +3. Skewness of the velocity reflects the observation of unequal and asymmetrical distribution of movement velocity (Disselhorst-Klug et al., 2012).

- *Metric 8: Cross correlation between wrist accelerations*

To investigate coordination among limb movements, we calculated the cross-correlations at zero time lag between wrist accelerations (Disselhorst-Klug et al., 2012).

- *Metric 9: Range of motion of the elbow angle*

The range of motion (ROM) of the elbow angle (EA), defined as the angle between the forearm segment and the upper arm segment. TD subject is expected to show a higher ROM with respect to the AR child.

- *Metric 10: Bouts of activity from wrists trajectories*

To consider the impact of extended periods without upper limb movements on the estimated parameters, bouts of activity were introduced. Time intervals during which the infants' wrists were in motion were extracted from the rest of the acquisition. These bouts were defined as periods where wrist speed exceeded a fixed threshold, set at 5% of the wrist's maximum velocity.

- *Metric 11: Stereotypy score*

Movement variation is a key indicator of infant motor development (Karch et al., 2012a). A lack of variation in spontaneous infant movements, characterized by repetitive, stereotyped movements, can be an early sign of significant neurological issues. Traditionally, assessing movement variation has relied on subjective observations. The term "variation of movements" refers to a diversity of patterns that differ in aspects like speed, amplitude, and joint angle. Conversely, movements that are stereotypical and monotonous lack this variation. Such movements often repeat throughout a recording and display similar time-series trajectories across several degrees of freedom, such as in elbow flexion-extension. While there may be minor differences in speed, amplitude, or onset, the fundamental shapes of these trajectories remain consistent.

A segment in the interval $[t_1,\ t_2]$ is considered a movement segment, M, if the speed $v(t)$ in this segment exceeds a basic threshold v_1, defined as 5% of the maximum velocity value of the entire measurement, and at least for one frame exceeds a higher threshold v_2, defined as 20% of the maximum velocity value of the entire measurement, on x, y, and z axes.

A movement segment is classified as stereotyped if its time-series trajectory exhibits a high degree of similarity to other segments from the same recording. To assess this similarity, movement segments are compared using the Dynamic Time-Warping (DTW) distance (Sakoe, 1978). For each pair of movement segments (I_1, I_2), the DTW-distance $dtw(I_1, I_2)$ is calculated. The similarity between each pair of dates (t_1, t_2), where t_1 is within the interval of movement segment I_1 and t_2 is within the interval of movement segment I_2, is determined by computing the DTW-distance of these segments. If one of the dates does not fall within a movement segment, the similarity is zero.

The similarity values of the elbow angle, the shoulder angle with respect to the vertical axis, and the shoulder angle with respect to the antero-posterior axis for one limb can be summed to produce a two-dimensional function sim_1 (**Figure 67**a), which evaluates the similarity of movements from one limb between each date t_1 and all other possible dates t_2. A mean similarity $sim_2(t)$ is then calculated by averaging the similarity values of all movement segments for each t (**Figure 67**b).

The final stereotypy score is the maximum value of the moving average of the $sim_2(t)$.

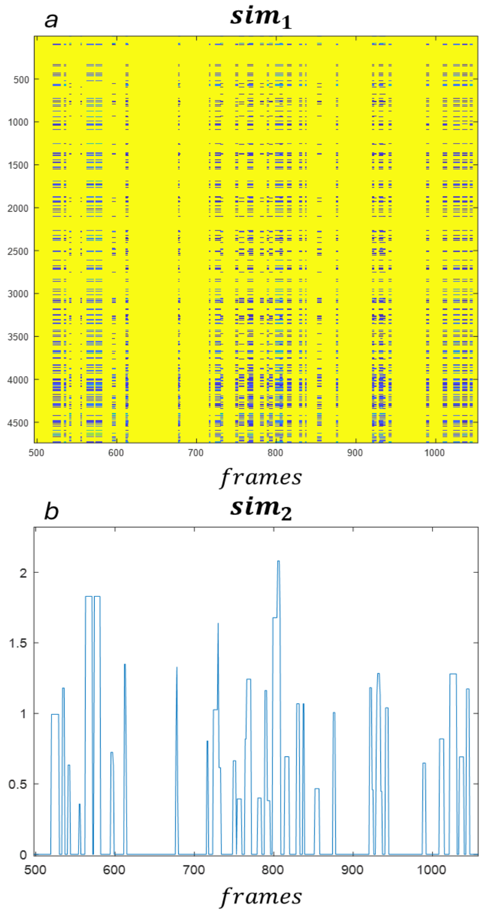


Figure 67. Stereotypy score: a) The function sim_1 is a two-dimensional evaluation of movement similarity from one limb between each date t_1 and every other date t_2. b) The mean similarity $sim_2(t)$ is determined by averaging the similarity values of all movement segments for each time point t.

- *Metric 12: Jerk Index*

An active period was defined as the time interval when the average velocity is higher than the minimum velocity value plus 5% of the maximum velocity value for at least one of the wrists. The Jerk Index (Kanemaru et al., 2013b) was determined by integrating the square of the jerk magnitude (the rate of change of acceleration) over time, normalized by the distance moved. This process was applied to each limb including only the active periods from the entire time-series, using the following equation:

$$C = \frac{1}{2} * \frac{(\dddot{x}^2 + \dddot{y}^2 + \dddot{z}^2)}{\bar{V}}$$

Where x, y, and z are the 3D wrist's location. C is the jerk index; V is the tangential velocity. A high jerk index suggests that the movement is unsteady (not smooth).

- *Metric 13: Kurtosis of wrists acceleration*

To assess the distribution of movement data, the shape of the acceleration's probability distribution was analyzed by calculating the kurtosis (β) for each wrist. A higher β value, indicating a more sharply peaked and heavy-tailed distribution, suggested intermittent and repetitive limb movements (Kanemaru et al., 2013b).

- *Metric 14: Correlation between limb velocities*

To assess coordination between limb movements, we computed cross-correlations at zero time lag between tangential limb velocities on the active periods. This index quantifies the similarity in velocity waveforms between the two limbs. Generally, AR subjects exhibit less variability in their movements compared to TD subjects, which results in a higher correlation between the movements of the right and left wrists (Kanemaru et al., 2013b).

- *Metric 15: Average velocity of limb movements*

The velocities along the x, y, and z axes were derived from the position of each wrist. We then calculated the average velocity by averaging instantaneous velocities over the entire acquisition. Typically, TD subjects exhibit faster movements compared to AR subjects (Kanemaru et al., 2013b).

3.4.3 Validation tests

The effectiveness of the MS method described above in reconstructing the 3D coordinates estimation was assessed using both a physical model (Balta, Kuo, Wang, Porco, Schladen, Cereatti, Lum, et al., 2022) and actual infants (Balta, Kuo, Wang, Porco, Schladen, Cereatti, Cereatti, et al., 2022; Balta, Kuo, Wang, Porco, Schladen, Cereatti, Lum, et al., 2022). The validation was carried out by comparing the automatic values with those manually measured on both a doll and the actual infant.

1) *Validation on a physical model*

To evaluate the proposed method's accuracy in reconstructing 3D PoI coordinates, a doll was placed on a turntable rotating at 33⅓ rpm and recorded for 5 seconds (**Figure 68**). Recordings were made with the camera's image plane both parallel to the rotation plane (0° acquisition) and at a 45° angle to the horizontal plane to better capture potential occlusions of the PoIs. The length of the upper arm (UA) was defined as the three-dimensional distance between the shoulder and elbow PoIs, while the forearm (FA) length was defined as the distance between the elbow and wrist PoIs. These measurements were taken for both arms and compared against manually measured reference values. Additionally, the angles at the elbows and shoulders were calculated: the elbow angle (EA) was the angle between the forearm and upper arm, and the shoulder angle (SA) was the angle between the upper arm and a line connecting both shoulders. The wrist's linear velocity (RW) was also determined from its 3D position over time and compared to a reference value derived from the turntable's nominal angular velocity and the radius of the wrist's path.

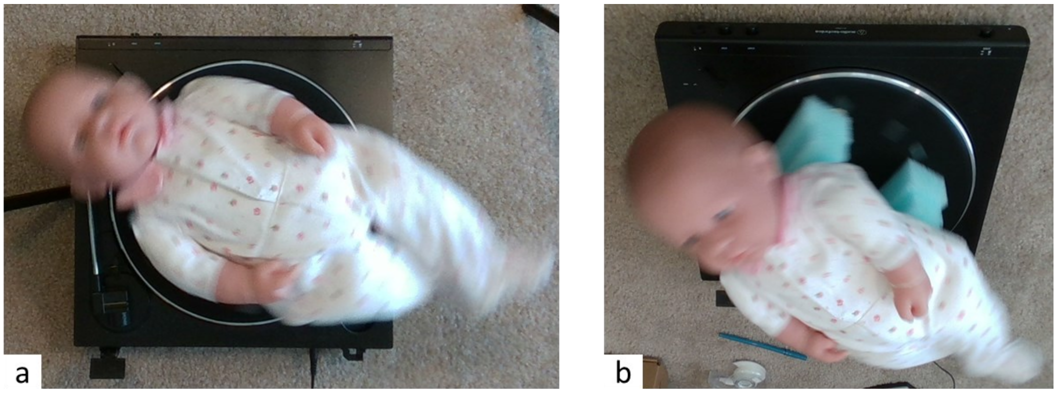


Figure 68. The doll lying on the turntable seen from the 0° view (a) and 45° view (b).

2) *Validation on a real baby*

An infant was recorded while seated in a baby seat at three different ages: 4, 5, and 6 months. Each recording session lasted approximately 30 seconds. From these recordings, bouts of upper body movement were identified by analyzing the motion of the right (R) and left (L) wrists. Only bouts lasting more than 0.5 seconds were further analyzed. The same measurements that were calculated for the turntable experiments—such as segment lengths, angles, and velocities—were also determined for each identified bout of movement. The segment lengths were manually measured on the infant, and PoIs were identified through manual palpation and marked using a black felt pen. These PoIs were then identified on a static image to determine the reference values for the angles and velocities.

3.4.4 Results

1) Validation on a physical model

The analysis of the 3D video data captured at 0° and 45° angles involved addressing numerous gaps that occurred due to the three potential issues outlined earlier and depicted in **Figure 62**. **Table 13** details the frequency and maximum duration of these gaps for each tracked PoI.

Table 13. Number of gaps and maximum duration for each cause of gaps.

Issue causing gaps		***# of gaps [max duration (s)]***	
		45°	0°
Time alignment		52 [0.33]	53 [0.33]
Occlusions + "Black area"	***LS***	49 [0.43]	7 [0.07]
	LE	2 [0.033]	2 [0.033]
	LW	0	0
	RS	58 [0.6]	1 [0.033s]
	RE	0	0
	RW	0	0

The impact of the training set size on the accuracy of the segment length estimates during the 0° acquisition is illustrated in **Figure 69**, where the corresponding mean absolute errors (MAE) are presented.

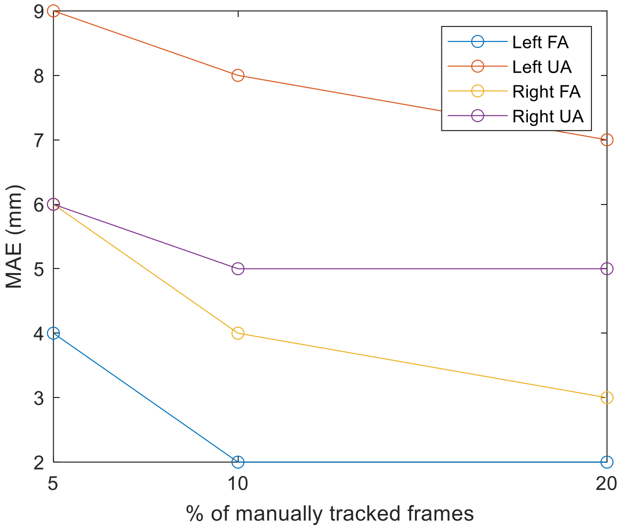


Figure 69. 0° Acquisition segment lengths MAE for training sets of three sizes (5%, 10% and 20%)

Table 14 presents the MAE and the mean absolute percentage error (MAPE) for the estimated values of the right (R) and left (L) UA and FA segment lengths, EA, SA, and the linear velocity of the wrist (RW), compared to their manually measured reference values. These metrics are provided for both the 0° and 45° video acquisitions.

Table 14. MAE and MAPE of the estimates of the UA and FA segment lengths, EA and SA and linear velocity of the RW

	MAE (MAPE %)		***Estimated value***		*Ref*
	0°	**45°**	***0°***	***45°***	-
***UA segment length** [mm]*	8 (13%)	15 (25%)	*58 ± 6*	*58 ± 7*	*60*
***FA segment length** [mm]*	2 (4%)	7 (14%)	*52 ± 8*	*52 ± 9*	*50*
***EA** [°]*	4.5 (18%)	8.8 (36%)	*21.2 ± 5.0*	*29.1 ± 8.5*	*24.5*
***SA** [°]*	4.2 (14%)	10.0 (32%)	*27.2 ± 4.8*	*23.9 ± 9.2*	*30.8*
***Linear velocity** [m/s]*	0.03 (10%)	0.05 (16%)	*0.3 ± 0.1*	*0.3 ± 0.1*	*0.3*

2) Validation on a real baby

The same measurements obtained from the turntable recordings were calculated for each activity bout. The total number of bouts observed during the infant's recordings, along with their average duration at each timepoint, are presented in **Table 15**.

Table 15. The number of bouts and their mean duration during the infant's acquisitions

	Side	***4 months***	***5 months***	***6 months***
# Bouts	**L**	6	4	3
# Bouts	**R**	8	9	4
***Bouts duration** (s)*	**L**	1.11 ± 0.53	1.03 ± 0.69	1.53 ± 0.72
***Bouts duration** (s)*	**R**	1 ± 0.3	1.89 ± 0.4	0.98 ± 0.2

Table 16 presents the MAE of the estimated anthropometric parameters, specifically the segment lengths of the right and left upper arms (RUA, LUA) and forearms (RFA, LFA), compared to the reference values for the infant's recordings at each timepoint. The values for the right and left limbs have been averaged.

Table 16. The MAE of the estimates of the anthropometric parameters at each timepoint

	4 months		***5 months***		***6 months***	
	estimate	***MAE***	***estimate***	***MAE***	***estimate***	***MAE***
***UA segment length** (mm)*	102 ± 11	8	110 ± 11	11	111 ± 32	8
***FA segment length** (mm)*	80 ± 17	11	91 ± 18	13	94 ± 9	10

Table 17 shows the average Range of Motion (RoM) for the elbow angle (EA) and shoulder angle (SA) during the activity bouts recorded at each timepoint for the infant. The values for the right and left sides are listed separately.

Table 17. The mean ROM during the bouts of EA and SA for infant's acquisitions at each time point.

	Side	*4 months*	*5 months*	*6 months*
RoM_EA (°)	**L**	130	106	117
RoM_EA (°)	**R**	80	110	72
RoM_SA (°)	**L**	120	116	171
RoM_SA (°)	**R**	64	130	93

Table 18 details the hand path and average velocity observed during the activity bouts at each timepoint. The data for the right and left sides are presented separately.

Table 18. The mean velocity and hand path during the bouts at each timepoint.

	Side	*4 months*	*5 months*	*6 months*
Mean velocity *(m/s)*	**L**	0.02 ± 0.011	0.03 ± 0.008	0.007 ± 0.003
Mean velocity *(m/s)*	**R**	0.03 ± 0.005	0.010 ± 0.002	0.014 ± 0.009
Hand path *(mm)*	**L**	900	708	449
Hand path *(mm)*	**R**	300	210	652

3.4.5 Discussions

The goal of this study is to preliminarily validate an innovative MS protocol for infants using a simple, cost-effective setup. This setup includes a single commercial RGB-Depth camera and open-source motion tracking algorithm powered by a deep neural network. To validate this method, it was applied to controlled movements, such as a doll on a turntable, to capture clinically relevant metrics like segment lengths, joint angles, and velocities of PoIs, which serve as reference values. The tracking algorithm, initially designed for animal pose estimation, was able to limit errors in estimating upper body segment lengths of the doll to less than 15 mm. Accuracy significantly improved when the training set size was increased from 5% to 10% of the acquired frames. However, further increasing the training set to 20% provided limited additional benefits. This indicates that using approximately 10% of the frames for training strikes an optimal balance between accuracy and the time required for manual annotation (about 30 minutes for a 6.6-second acquisition with 20% of frames used). The results also demonstrate better performance when the camera's image plane is aligned parallel (0°) to the motion plane compared to a 45° angle, providing a key guideline for setting up infant motion recordings. While data gaps and the algorithm's limited ability to track

PoIs during occlusions or unclear views affect the accuracy of segment length measurements, these issues have less impact on wrist velocity estimates, thereby enhancing the reliability of clinical parameters derived from wrist trajectory data. Various factors contribute to the method's limitations:

a. Hardware Limitations: The irregular frame rates of the RGB and depth sensors lead to gaps in the captured images. "Black areas" found in depth images also cause additional gaps in the data sequence. Despite attempts to fill these gaps with cubic interpolation splines, the extended duration of gaps leads to significant errors in 3D position estimates;

b. Algorithm Limitations: The motion tracking algorithm used was not specifically designed for tracking infants but rather animals, and it has limited customization options for improving performance. Additionally, because it relies solely on RGB images, it does not take into consideration the depth data available.

c. Camera Orientation Limitations: camera orientation relative to the motion plane plays an essential role, especially as this method aims to track the upper body motions of seated infants. It is recommended to place the camera directly in front of the infant to minimize occlusions. Additionally, this preliminary validation focuses solely on planar movement and therefore does not account for errors arising from off-plane movements.

d. Tracking Limitations: MS-based motion tracking methods typically monitor movements of body surface areas rather than specific points like internal joint centers, which can introduce errors due to the three-dimensional nature of human joints.

As future studies, addressing these issues could involve using an RGB-Depth camera with an enhanced depth sensor, positioning it appropriately based on expected motion, and developing an optimized tracking algorithm directly for 3D images that includes tracking joint centers directly as internal points. Another challenge is the time required to manually label even 10% of the RGB images; a specialized tracking algorithm could alleviate the burden of generating the training set.

It is important to note that this validation conducted on upper and forearm lengths, wrist linear velocity, elbow and shoulder angles was conducted because these parameters are closely related to the metrics described in the “Kinematic parameters and metrics estimation”. Elbow angle and wrist velocity were directly validated. Wrist trajectory was not directly validated; however, the length of the forearm can provide an indication of the accuracy of the reconstructed wrist trajectory. No specific validation was performed for wrist acceleration and jerk, but it can be asserted that these metrics might be likely affected by greater errors compared to wrist velocity, as they are obtained through derivatives of the wrist trajectories, which inherently introduces additional noise.

Despite existing limitations, the results from this study demonstrate valuable insights for potential applications in recording infant motion and extracting clinical biomarkers for early detection of movement disorders.

3.5 Case study 1: Exploratory Analysis of General Movements in 3-5 Month Old Infants Using a single RGB-Depth sensor

3.5.1 Experimental setup and protocol

Eight parents from the community volunteered to record videos of their infants at home. The infants were seated in baby chairs covered with green fabric to facilitate subject segmentation and to identify PoIs. Placed in front of an RGB camera with a depth sensor, the infants' natural movements were recorded for approximately three minutes at three distinct ages: 3 months, 4 months, and 5 months. To maintain consistency, the same washable baby seat was used throughout the testing periods. Efforts were made to control lighting conditions and minimize human interaction as much as possible within the home setting, ensuring the protocol remained simple while replicating natural conditions. The results and their implications for clinical practice are analyzed, with clinical insights from two pediatric experts in neurology and physiatry. The first nine metrics, reported in the paragraph “Kinematic parameters and metrics estimation” were evaluated.

3.5.2 Results

In this study, two expert physicians reviewed recorded videos of infants at each time point to detect any indications that an infant may not be developing typically (AR). While not every video received comments, a comprehensive assessment was made for each infant. The physicians concurred that four infants (S1, S5, S7, and S8) seemed to be TD, whereas one infant (S2) exhibited signs of potential atypical development. Opinions differed between the two clinicians regarding the developmental status of the remaining three infants (S3, S4, and S6). The first seven GM metrics presented in “Kinematic parameters and metrics estimation” derived from the 3D PoI kinematics of the upper body using the introduced MS technique are displayed in **Figure 70** for each infant at each assessed age, for both the left and right sides, and aggregated. To correlate the clinical assessments with the extracted metrics, the value ranges for infants without signs of atypical development were highlighted in gray on each graph.

Figure 71 and **Figure 72** illustrate the cross-correlation between accelerations of the left and right wrists and the range of motion for the elbow angle at each time point (3, 4, and 5 months) for every subject, respectively. **Figure 73** presents the average and standard deviation of the durations of movement bouts for each infant at each time point, alongside the total number of bouts and the duration of movement expressed as a percentage of the total recording time. Infants not exhibiting signs of atypical development (marked in gray) are shown to move their arms more, particularly at the 4 and 5-month evaluations.

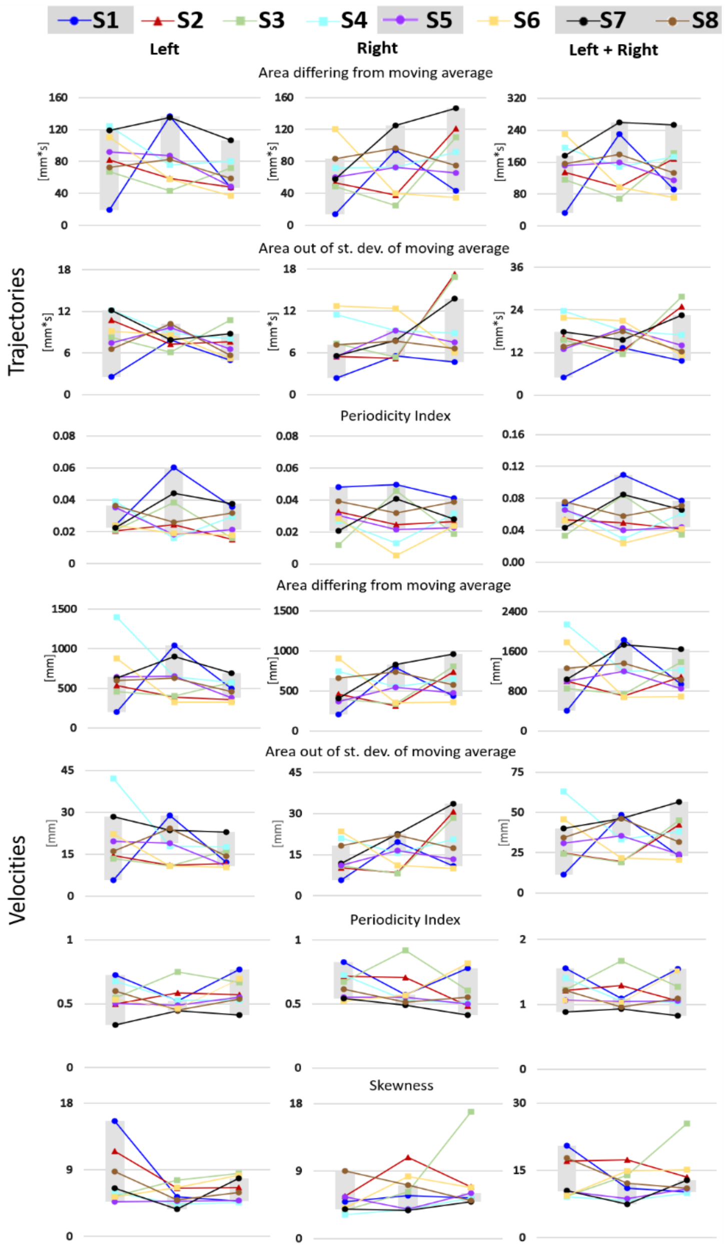


Figure 70. Metrics derived from wrist trajectories and velocities for each infant at three time points (3, 4, and 5 months). Infants showing typical development as agreed upon by both physicians are marked with circles. Those raising concerns of atypical development from both physicians are indicated with triangles, while those evaluated differently by each physician are marked with squares. The gray interval at each time point represents infants suggesting typical development.

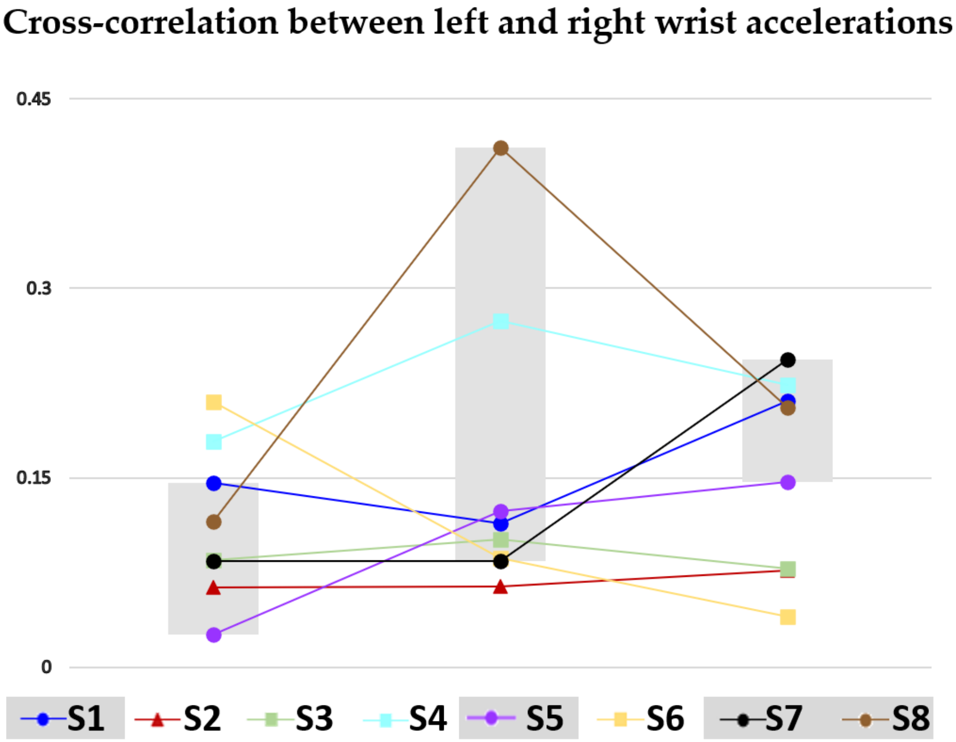


Figure 71. Cross-correlation of left and right wrist acceleration for each infant at 3, 4, and 5 months. Infants not indicating atypical development as determined by both physicians are marked with circles. Those causing concern about atypical development from both physicians are indicated with triangles, while those evaluated differently by each physician are denoted by squares. The gray interval at each time point represents infants not suggesting atypical development.

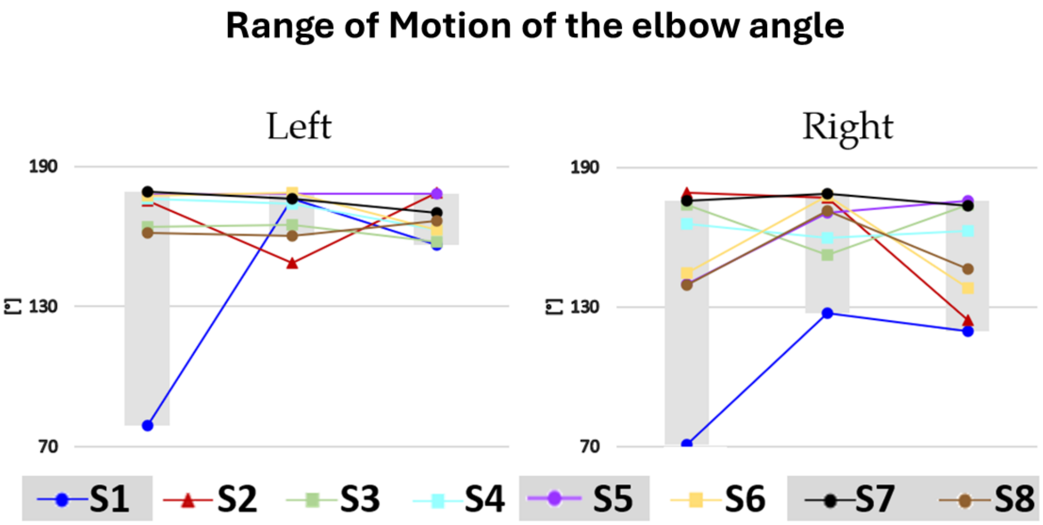
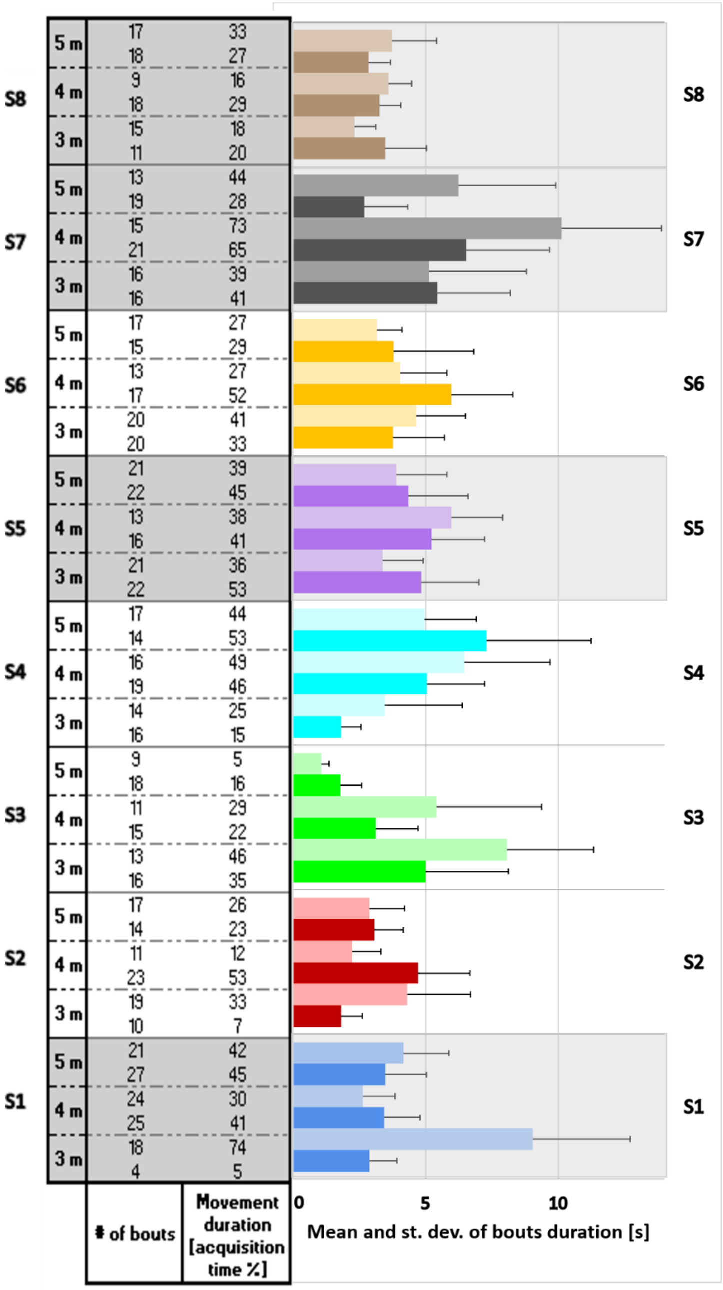


Figure 72. Elbow Angle Range of Motion for the Left and Right Sides for Each Infant at Each Time Point (3, 4, and 5 Months). Infants displaying typical development as agreed upon by both physicians are marked with circles. Those raising concerns about atypical development from both physicians are indicated with triangles, and infants assessed differently by the physicians are represented by squares. The gray interval at each time point denotes infants not suggesting atypical development.



		# of bouts	Movement duration [acquisition time %]
S8	5 m	17	33
		18	27
	4 m	9	16
		18	29
	3 m	15	18
		11	20
S7	5 m	13	44
		19	28
	4 m	15	73
		21	65
	3 m	16	39
		16	41
S6	5 m	17	27
		15	29
	4 m	13	27
		17	52
	3 m	20	41
		20	33
S5	5 m	21	39
		22	45
	4 m	13	38
		16	41
	3 m	21	36
		22	53
S4	5 m	17	44
		14	53
	4 m	16	49
		19	46
	3 m	14	25
		16	15
S3	5 m	9	5
		18	16
	4 m	11	29
		15	22
	3 m	13	46
		16	35
S2	5 m	17	26
		14	23
	4 m	11	12
		23	53
	3 m	19	33
		10	7
S1	5 m	21	42
		27	45
	4 m	24	30
		25	41
	3 m	18	74
		4	5

Figure 73. Mean and Standard Deviation of Bout Duration for Each Infant at Time Points 3, 4, and 5 Months. The left side is represented in a lighter color, while the right side is depicted in a darker color. The number of bouts and the duration of movement, expressed as a percentage of the acquisition time, are detailed in the table on the left. Subjects showing no signs of atypical development are highlighted in gray.

3.5.3 Discussions

The recording and analysis of an infant's upper body movements in a familiar setting has proven to be a complex task due to both technological and environmental challenges. Technologies previously utilized often fail, as they are generally tailored for adults or older children's movement analysis (Cappozzo et al., 2005). The small size of infants makes securely and safely attaching markers challenging, hence MS methods are particularly beneficial for analyzing their movements.

MS methods have introduced new possibilities for movement analysis, though initially these were largely limited to two dimensions (Balta et al., 2020; Castelli et al., 2015; Moro, Pastore, et al., 2022; Stagni et al., 2023; Surer et al., 2011).

Currently, available low-cost RGB-D cameras in the consumer electronics market have made it possible to extend MS techniques to 3D movement analysis without complicating the experimental setup. This simplicity is critical as it allows the techniques to be used outside of laboratory settings and supports repeated measurements over time. This is especially important in the study of infants' movements, where sensorimotor integration rapidly unfolds in the first months of life through activity-dependent neuronal modeling (McIntyre et al., 2011). Regular and routine monitoring of infants' movements in the familiar home setting enhances the likelihood of early identification of abnormal movement patterns and the timely introduction of interventions to prevent the loss of neural connections and functions (Novak et al., 2017).

In this study, we applied a MS method to the RGB images captured from a commercial RGB-D camera, using selected upper body PoIs extracted from the RGB video frames along with recorded depth information to reconstruct 3D PoI kinematics (Balta, Kuo, Wang, Porco, Schladen, Cereatti, Cereatti, et al., 2022). A novel metric and previously established metrics that were initially proposed for quantifying GM (Disselhorst-Klug et al., 2012; Meinecke et al., 2006) recognized for their effectiveness in early detection of movement disorders in infants (Einspieler & Prechtl, 2005) were computed. A key aspect of our method is its applicability in non-clinical environments like the home, however, it is important to highlight that the GMA has not been replicated in our protocol. In fact, infants were acquired in casual home environment, seated in standard infant seats rather than lying supine, and recorded from the front with a commercial camera tripod instead of requiring a specialized overhead camera setup. This alteration in posture likely influenced the GM parameter trends compared to those previously reported (Disselhorst-Klug et al., 2012; Meinecke et al., 2006) for three to five-month-old infants. Due to the limited size of our sample, meaningful statistical analysis was not feasible. We focused instead on depicting trends across the three-, four-, and five-month measurement periods. For details, see **Figure 70** for parameters 1-7, **Figure 71** for parameter 8, and **Figure 72** for elbow angle range of motion plots. Significant differences between TD and non-TD infants were not expected in our cohort, as none had documented injuries that would classify them as at-risk. Although variability in the data generally made challenging to compare our results to the expected ones, trends in two metrics aligned with existing literature.

Metric #1, which measures the area where the wrist trajectory deviates from its moving average, has been thought to quantify GM diversity and fluidity. We expected from the literature (Disselhorst-Klug et al., 2012; Meinecke et al., 2006) no significant change in this metric for TD infants from three to five months, whereas non-TD infants were anticipated to show smaller areas. In alignment with this, at four months, infants flagged by evaluators as AR (S2, S3, S4, S6) displayed smaller metric values compared to their typically developing counterparts (S1, S5, S7, S8). This metric was expected to decrease consistently in non-TD infants over the specified period; however, we observed considerable variability across timepoints.

Similarly, Metric #8, the cross-correlation of acceleration between the left and right wrists, indicative of movement similarity and coordination, conformed to previous findings. TD infants are expected to exhibit synchronous, coordinated movements within the three-to-five-month window, unlike non-TD infants who would likely show asynchronous, uncoordinated patterns (Disselhorst-Klug et al., 2012). This metric should rise in TD infants between three and five months, but not in non-TD children. At five months, all infants without concerns (S1, S5, S7, S8) showed higher metric values than at three months, although for S8, who peaked at four months and then decreased, still remaining above the initial three-month value. In contrast, S6 (split concern) showed a consistent decline across the period, and infants S2 (agreed concern) and S3 (split concern) showed no change, maintaining low values well below the TD range at five months. Infant S4 (split concern) did not exhibit a decreasing or stable trend over time. Overall, 7 of the 8 infants followed expected patterns.

We introduced a novel exploratory metric, elbow range of motion (**Figure 72**), observing that most infants' left arm ROM stayed within a narrow range from three to five months. S1 (no concern) showed very limited range initially but increased to match the cohort average by four months. S2 (concern) exhibited one of the highest ROMs at three and five months but the lowest at four months. More variability was observed on the right side, where S1's ROM was notably low at three months and increased at four and five months, although not as dramatically as on the left. S2 (concern) experienced a significant decrease in ROM at five months. A larger dataset will be required to assess the utility of this new metric for screening non-TD infants.

Our analysis suggests that these metrics are interconnected and influenced by the environment. Several adjustments are recommended for future studies based on video inspections. Enhanced control of environmental factors could reduce data variability. Despite protocol instructions to keep the infant's visual field clear during testing, ensuring this was challenging in home settings. In one video, a sibling's approach caused an infant to laterally shift their movement. In another, an infant frequently placed a hand in their mouth, indicating that multiple recording sessions might be necessary for such cases. The use of a standard baby seat, while maintaining consistency and hygiene, restricted movements more than the traditional GM assessment protocol for supine infants. Future research should consider standardized postural supports for younger seated infants.

Clinically, assessments from the two clinicians involved in this study about whether an infant is developing typically are based on a comprehensive range of motor characteristics, including hand movements and bringing hands to the mouth, which fall outside the eight kinematic parameters proposed by Prechtl et al. (Disselhorst-Klug et al., 2012; Meinecke et al., 2006) and employed in this study. Among the characteristics considered by the two clinicians, there are midline gaze, hand-to-midline movements, visual field preferences, visual attentiveness, social smiling, and social engagement, all of which are critically observed by clinicians. Integrating quantifiable clinical criteria with GM kinematics could enhance the effectiveness of 3D MS movement assessments in infants.

While reviewing the videos, the two clinicians also took into account the infants' states, such as sleepiness or distractions from nearby individuals, which could complicate the interpretation of movement patterns in relation to typical or pathological development. For instance, infant S2 was flagged by both clinicians for exhibiting concerning traits. During both the three- and four-month sessions, S2 was observed to be slumped to the right, likely causing asymmetric movement. This was reflected in the low cross-correlation between the left and right wrist accelerations, which was among the lowest within the typical range at three months. In contrast, infant S6 showed the highest cross-correlation between left/right wrist accelerations at three months but dropped to the lowest in the cohort without concerns at four months and recorded the lowest cross-correlation at five months. Clinician notes indicated that this infant spent much of the recording time with a finger in his mouth, restricting spontaneous movements. This constraint should ideally be addressed, and the assessment repeated under more controlled conditions to provide a more accurate depiction of the infant's movement characteristics.

3.6 Case study 2: Assessment of quantitative metrics for spontaneous movements analysis on Twins with Divergent Health Profiles

As outlined in the previous sections, we have developed a MS protocol for analyzing infant upper limb movements using a single RGB-D camera at home. However, a significant issue of this protocol is the extensive time required during the transfer learning phase of the DLC for operators to manually identify the PoIs, which complicates the clinical applicability of this method. As previously described, DLC was trained by manually labeling PoIs in 10% of the video frames, selected using k-means clustering (Mathis et al., 2018). It has been observed that while k-means clustering effectively selects images under varying lighting conditions, it often continues to include, in the training set, frames where the subject assumes similar postures. This selection does not consider the relative position of the subject's limbs but focuses only on pixel brightness. Furthermore, the identification of PoIs on the RGB images depends on how the PoI area is captured by the camera. Depending on the RGB frame, a single PoI might be marked on different parts of the infant's body surface (**Figure 74**). It is beneficial for the DLC to receive as input various frames representing different joint configurations during training phase.

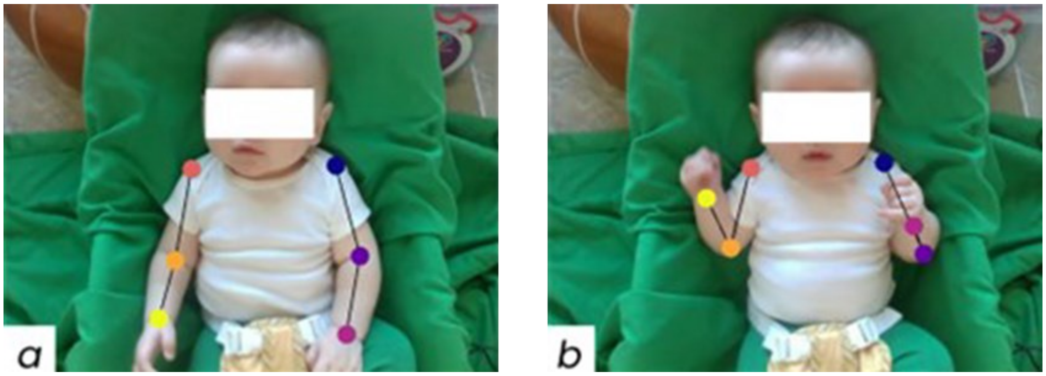


Figure 74. (a) RE is located midway between the epicondyles. (b) RE is positioned on the medial epicondyle.

Moreover, the proposed method does not account for specific issues that could arise during recordings conducted in the subjects' homes. In particular, the reconstruction of movement could be compromised by accidental movements of the camera or the baby seat if the environment is not controlled. These unintended movements could introduce motion artifacts and potentially lead to misinterpretations of the infant's actual movements.

For all these reasons, two important steps have been added to the previous method:

a. a video processing technique to enhance the robustness of the movement reconstruction in case of accidental movements.

b. a training set construction method that enhances DLC’s efficiency in monitoring the general movements of infants;

The proposed refined method was applied to a critical unique real-world scenario on a pair of twins where one is TD, and the other has been diagnosed as AR for CP.

The primary objective of this study was to investigate the potential of these metrics as reliable indicators of developmental abnormalities.

3.6.1 Method description

Figure 75 shows the block diagram of the method, for simplicity, only the two additional steps will be explained and validated.

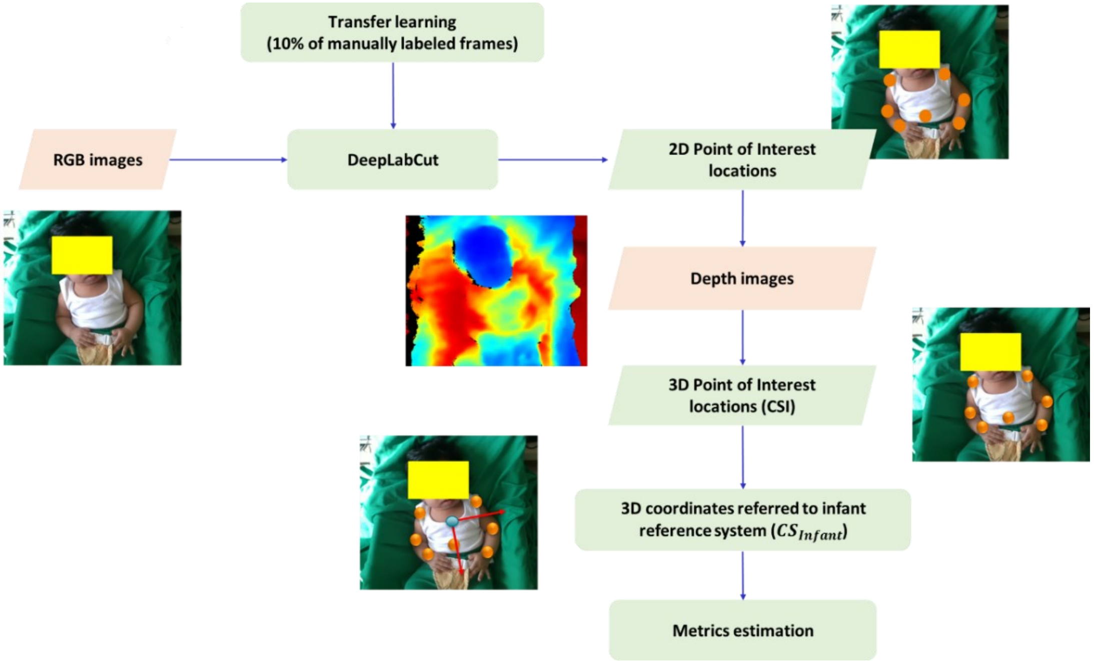

Transfer learning
(10% of manually labeled frames)
RGB images
DeepLabCut
2D Point of Interest locations
Depth images
3D Point of Interest locations (CSI)
3D coordinates referred to infant reference system (CS_{Infant})
Metrics estimation

Figure 75. Block diagram of the improved method.

1) Referring the 3D coordinates to an infant (local) reference system

In order to limit the effects of accidental motions of the camera, a new local reference system relative to the infant's trunk, CS_{Infant}, was introduced.

First of all, DLC was trained to identify an additional point of interest, the belly button (B). CS_{Infant} is centered at the centroid of the 3D coordinates of LS, RS, and B. It features a medio-lateral axis connecting LS and RS, an antero-posterior axis orthogonal to the plane formed by LS, RS, and B, and a vertical axis derived from the cross-product of the antero-posterior and medio-lateral axes. The 3D positions of PoIs were referred to this new infant reference system, and their coordinates were subsequently low pass filtered (**Figure 76**).

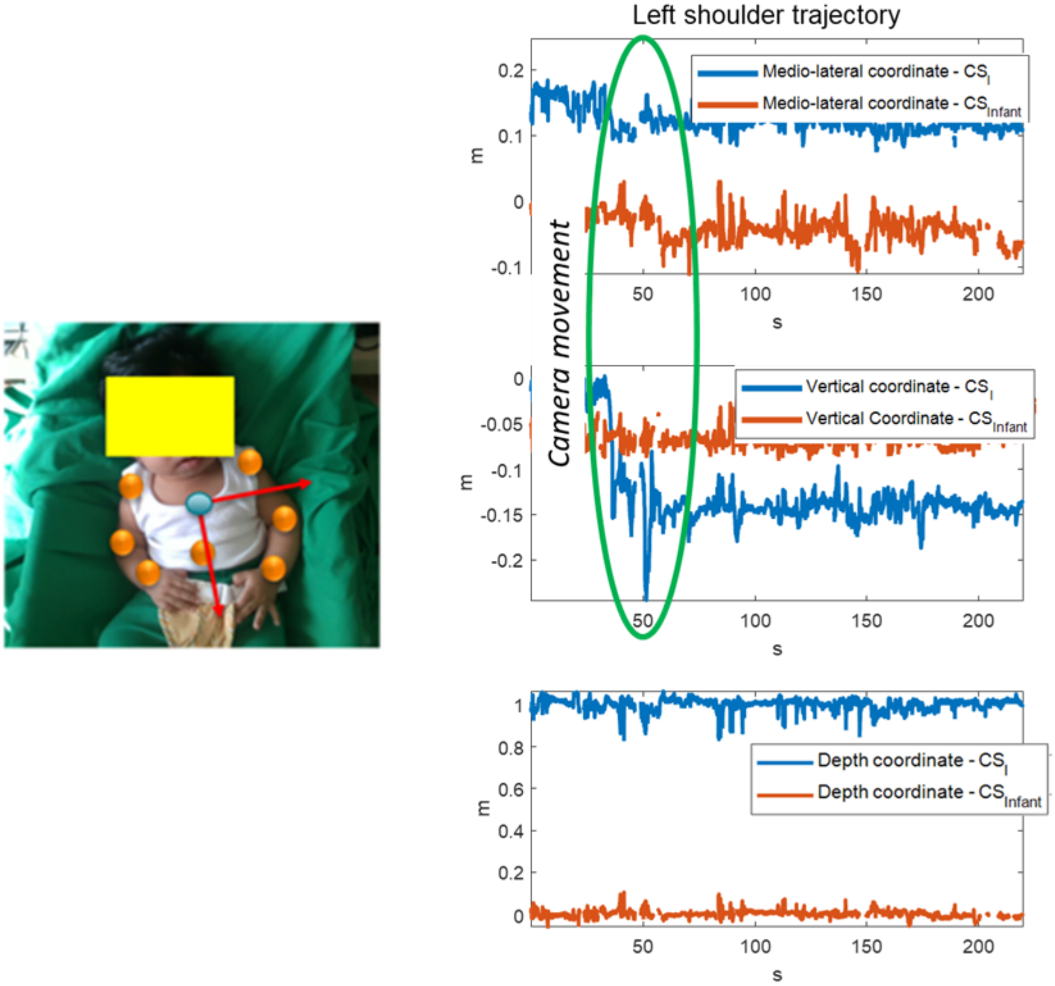


Figure 76. 3D coordinates of left shoulder in both reference systems. The blue line represents the image reference system while the red one represents the new infant reference system. The effects on the shoulder trajectory of the accidental camera movement occurring between frames 40 and 50 are circled in green.

2) An improved training set construction

A new method for constructing the training set was implemented by using the following steps:

a. The video sections containing only the infant fully in the FoV were extracted and the k-means was applied to generate a training set with 10% of video frames (S1).

b. Frames where the infant assumes similar poses were then removed since they do not include additional information in the training set (S2a);

c. Frames with less than 60% of the PoIs visible and frames where left and right shoulder and belly button were not fully visible (S2b) were removed (**Figure 77**);

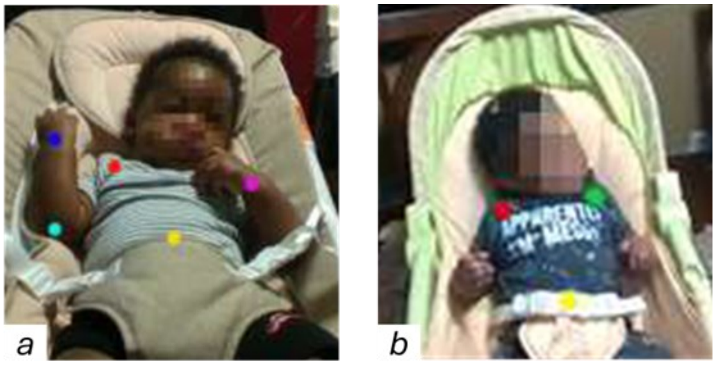


Figure 77. a) LS is not visible due to the occlusion from LW. b) less than 60% (4 out 7) of POIs is fully visible.

d. Since a single PoI may be seen by the camera with different angles (e.g. elbow joint center could be tracked on the medial or on the lateral epicondyle depending on the side of the arm facing the camera), the presence in the training set of all PoI views contained in the video acquisition was checked (S3).

For sake of clarity, the following figures are examples of joint configurations/views that may occur during the video acquisition for each POI (**Figure 78**, **Figure 79**, **Figure 80**).

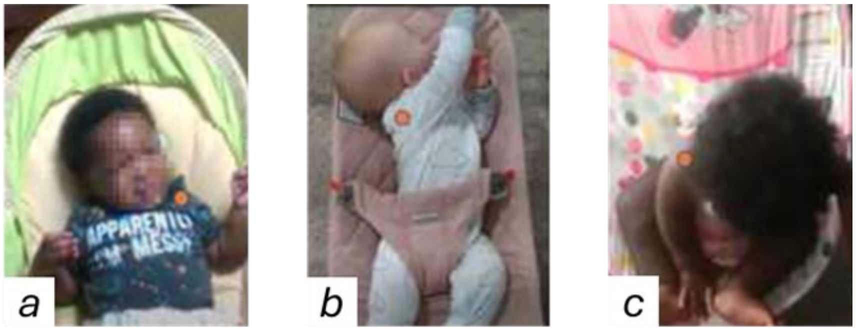


Figure 78. Shoulder configurations: a) frontal position, b) lateral position, c) on the top of the shoulder**.**

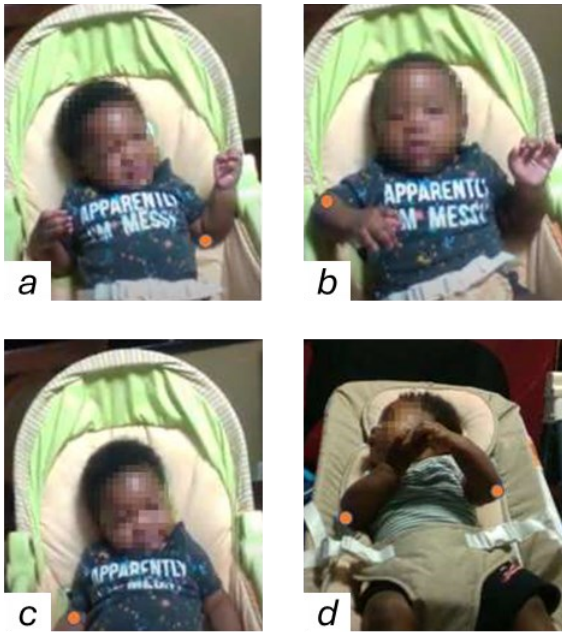


Figure 79. Elbow configurations: a) medial epicondyle, b) lateral epicondyle, c) in the middle between the two epicondyles. d) on the olecranon.

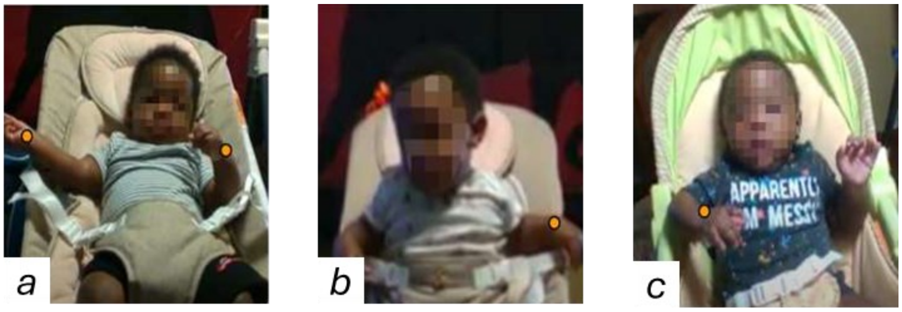


Figure 80. Wrist configurations: a) lateral styloid b) medial styloid, c) in the middle between the two styloids.

The similarity between frames and the completeness of angle views in the training set were assessed subjectively by two researchers.

3.6.2 Validation test

An infant sitting in a baby seat was positioned in front of an RGB-Depth camera and recorded for three minutes.

- DLC was then trained as in "2D Tracking Algorithm" (1), and using S1, S2 and S3. Segment lengths in 3D (UA and FA lengths) were then estimated as in "Validation tests". We assessed DLC's effectiveness by comparing the calculated lengths of the infants' UA and FA against manually measured reference values (**Table 19**).

Table 19. Duration, time needed to label PoIs and body segment lengths estimations and relevant errors at each step of the proposed method.

Method	***Training set length***	***Labeling time***	***FA length*** *(mm)*			***UA length*** *(mm)*		
	(frames)	*(hours)*	*Ref*	*Mean±sd*	*MAE*	*Ref*	*Mean±sd*	*MAE*
(1)	536	7.4	*95*	113 ± 10	18	119	103 ± 8	16
S1	480	6.6		93 ± 9	2		102 ± 9	16
S2 (a+b)	235	3.2		97 ± 10	2		108 ± 5	11
S3	235	3.2		97 ± 10	2		108 ± 5	11

Discussions

The findings from this study demonstrate a notable improvement in the accuracy of measuring the lengths of upper body segments compared to previous work referenced in (Balta, et al., 2022). This enhancement suggests that the refined training set, which excludes frames with PoIs occlusions, provides more informative data for analysis. By eliminating repetitive frames from the training set, the time required for PoI labeling by approximately 50% was reduced, thus significantly enhancing the practicality and broader applicability of the proposed MS protocol. Furthermore, the consistent results observed between steps S2 and S3 suggest that the k-means clustering approach is effectively creating a comprehensive training set including all the joint configuration present in the video acquisition. This set appears to capture all the potential PoI positions that an infant might assume, which is critical for the robustness and reliability of the training process.

3.6.3 Experimental setup and protocol

A pair of twins aged 4 months was recruited and positioned on a baby seat on the floor in front of the Intel RealSense D435. One of them is TD and the other has been diagnosed as AR for CP. Video recordings of up to three minutes for each twin were conducted at seven timepoints, spaced at about 30 days. 3D coordinates of seven PoIs, (left and right shoulders, elbows, wrists and belly button) on each twin were tracked using DLC as explained in "*An improved training set construction*". From the 3D coordinates of each PoI obtained as described in "Referring the 3D coordinates to an infant (local) reference system", metrics #1, #2, #3, #4, #5, #6, #7, #9, #11, #12, #13, #14, #15, explained in the paragraph "*Kinematic parameters and metrics estimation*" were estimated.

3.6.4 Results

Figure 81 and **Figure 83** highlight the areas where the wrist trajectories and velocities deviated from their moving average, as well as the regions where these trajectories and velocities fell outside the standard deviation for AR and TD at each timepoint. **Figure 82** shows periodicity index (PI) of wrist trajectories and velocities of AR and TD for each timepoint. **Figure 84** reports the values of skewness of wrists velocity of AR and TD for each timepoint. **Figure 85** shows the average range of motion of the elbow angle of AR and TD for each timepoint. **Figure 86** shows stereotypy score, jerk index and the kurtosis of wrists acceleration of AR and TD for each timepoint. Finally, **Figure 87** shows the correlation between wrists velocity and the wrists average tangential velocity of AR and TD for each timepoint. Timepoints 5 to 7 (8 months to 10 months) are represented in a lighter shade because the children at these ages are older than those analyzed in the reference articles (3-5 months).

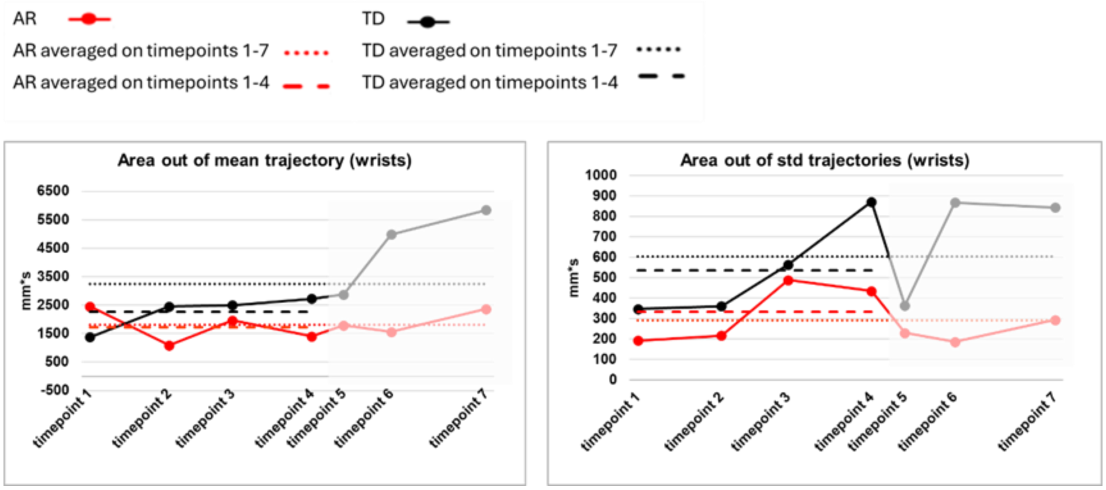


Figure 81. Area where the wrist trajectories deviated from the moving average and area where the wrists trajectories fell outside of the standard deviation of AR (red line) and TD (black line) for each timepoint. The dashed line represents the average value from the 1st to the 4th timepoints while the dotted line represents the average value from the 1st to the 7th timepoints.

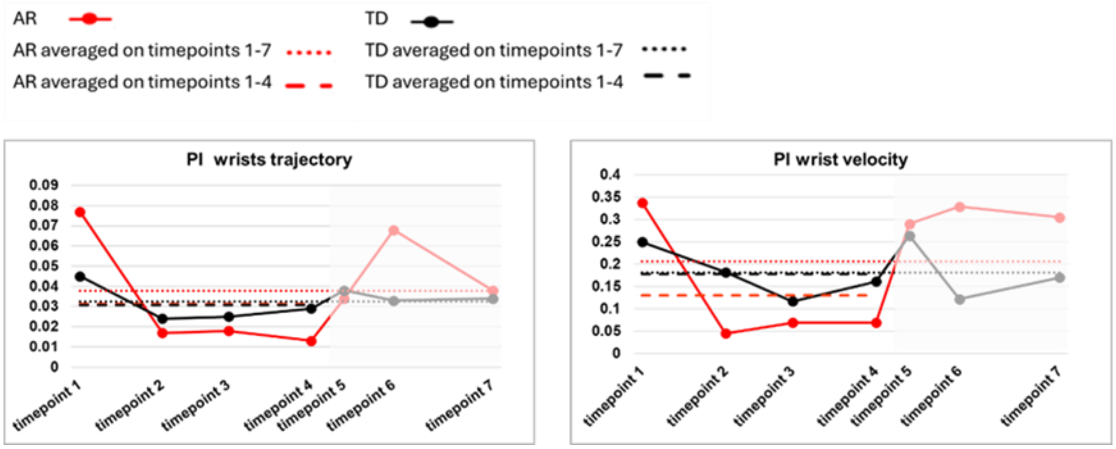


Figure 82. Periodicity index of wrist trajectories and velocities of AR (red line) and TD (black line) for each time point. The dashed line represents the average value from the 1st to the 4th timepoints while the dotted line represents the average value from the 1st to the 7th timepoints.

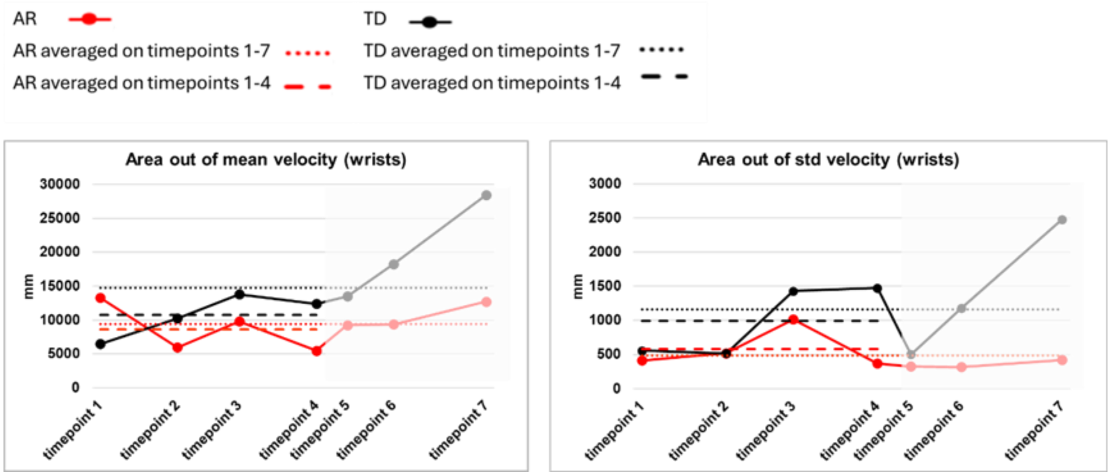


Figure 83. Area where the wrist velocities deviated from the moving average and area where the wrists velocities fell outside of the standard deviation of AR (red line) and TD (black line) for each timepoint. The dashed line represents the average value from the 1st to the 4th timepoints while the dotted line represents the average value from the 1st to the 7th timepoints.

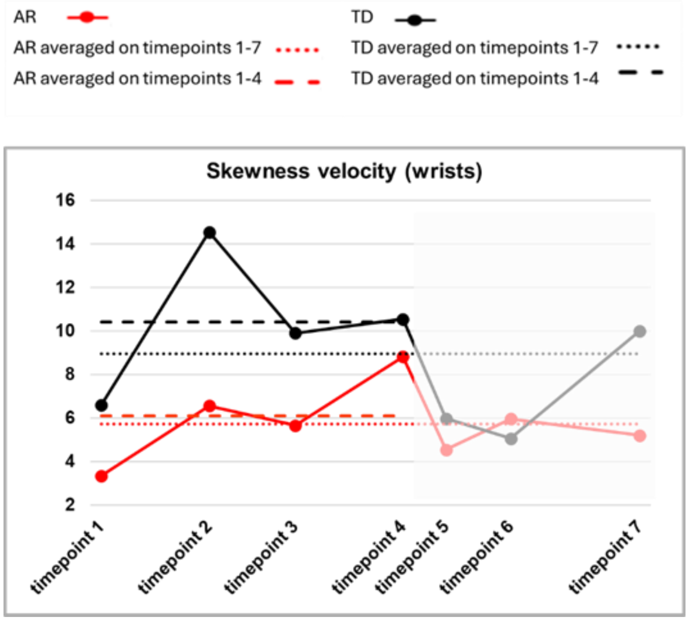


Figure 84. Skewness of wrist velocities of AR (red line) and TD (black line) for each time point**.** The dashed line represents the average value from the 1st to the 4th timepoints while the dotted line represents the average value from the 1st to the 7th timepoints.

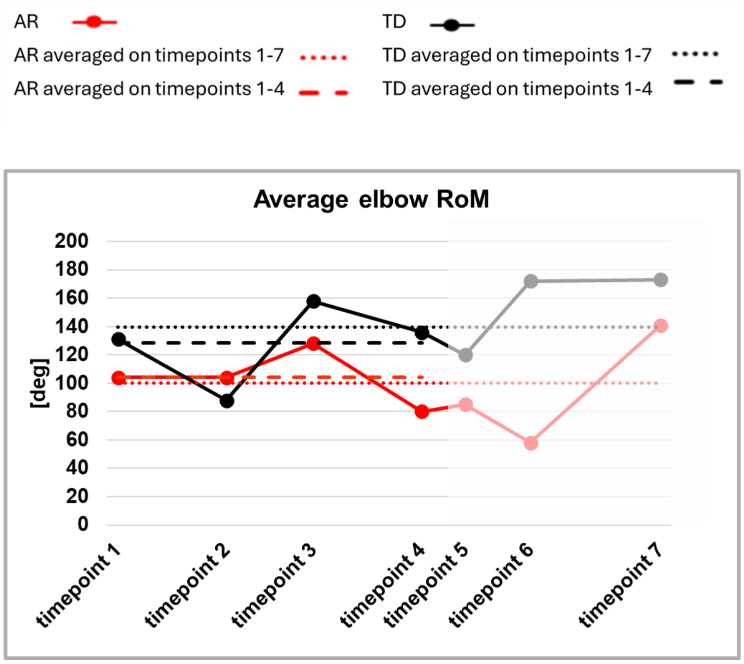


Figure 85. Average ROM of the elbow angle of AR (red line) and TD (black line) for each time point**.** The dashed line represents the average value from the 1st to the 4th timepoints while the dotted line represents the average value from the 1st to the 7th timepoints

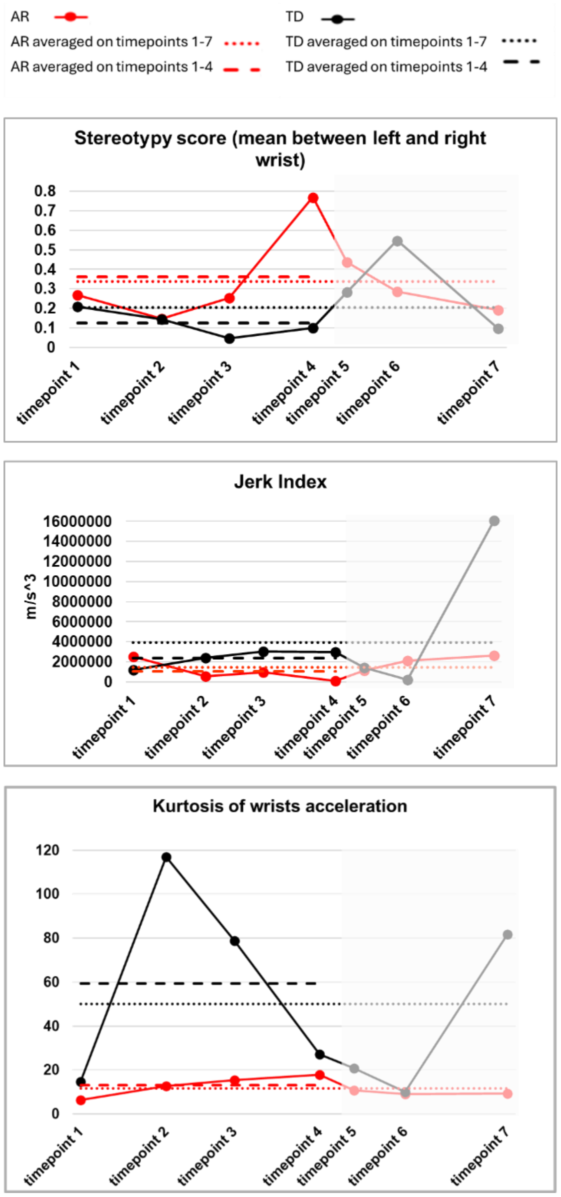


Figure 86. Stereotypy score, Jerk index and Kurtosis of wrist acceleration of AR (red line) and TD (black line) for each time point**.** The dashed line represents the average value from the 1st to the 4th timepoints while the dotted line represents the average value from the 1st to the 7th timepoints.

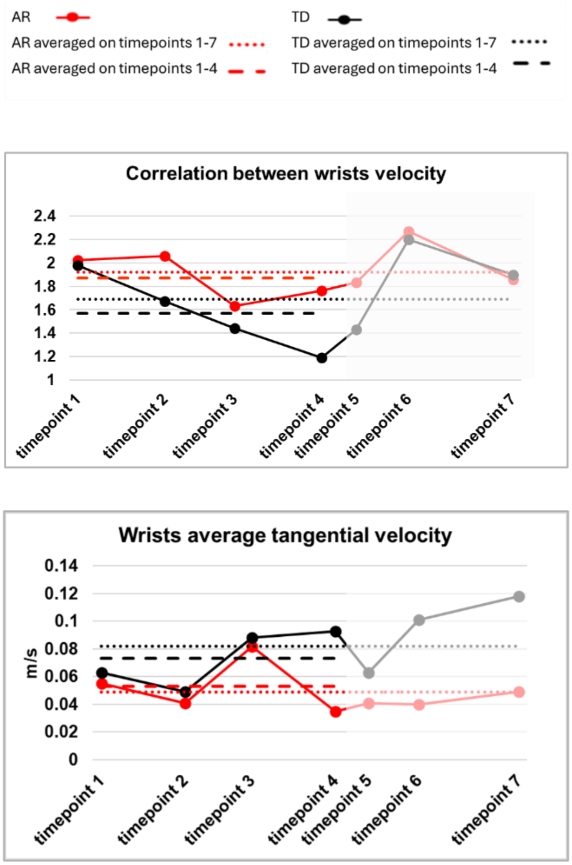


Figure 87. Correlation between wrists velocity and Wrists average tangential velocity of AR (red line) and TD (black line) for each time point. The dashed line represents the average value from the 1st to the 4th timepoints while the dotted line represents the average value from the 1st to the 7th timepoints.

3.6.5 Discussions

It is important to highlight that the objective of this study diverges from the reference articles from which quantitative metrics were selected (Disselhorst-Klug et al., 2012; Kanemaru et al., 2013a; Meinecke et al., 2006); in fact the primary aim of this study is not to conduct GMA but to provide a robust tool for monitoring movements by delivering quantitative metrics also in real-world scenario (home setting) where the acquisitions are not controlled from a clinical. Unlike general movement assessments, where children are typically placed on the ground in supine position, in this study, children were placed on seats that may change between timepoints. Additionally, the age of the twins in this study is slightly older than the 3-5 month range considered in the reference articles (Disselhorst-Klug et al., 2012; Kanemaru et al., 2013a; Meinecke et al., 2006).

1) Area out of mean and the standard deviation of wrist trajectories and velocities

Previous findings (wrist average tangential velocity and range of motion of elbow angle) are further supported by the analysis of both the area out of mean and the standard deviation of wrist trajectories and velocities, as depicted in **Figure 81** and in **Figure 83**. These metrics also show that as age increases, the difference between AR and TD becomes more pronounced. This suggests that TD individual exhibits more diversification in their movements due to the presence of intentional movements. The higher the value of these parameters, the more variants can be found in their movements (Disselhorst-Klug et al., 2012). Values found in this study are in line with those in (Disselhorst-Klug et al., 2012).

2) Periodicity index

As shown in **Figure 82**, the periodicity index for wrist trajectories and velocities in the first 4 time points (from 4 months to 7 months) is higher in TD compared to AR. This is influenced by the presence of predictable, repetitive, and intentional voluntary movements observed in TD subject, which can appear from the fifth month (Disselhorst-Klug et al., 2012), such as clapping hands and hitting the table. This is in line with previous findings (Disselhorst-Klug et al., 2012) indicating that infants begin to perform simple voluntary movements between the third and fifth months of life. From 8 to 10 months (5th to 7th timepoints), TD subject' s movements become faster, wider and less predictable, whereas AR subject becomes more still remaining in similar positions throughout the video recordings (as confirmed by elbow's ROM and average wrist velocity). Consequently, the periodicity index of the AR subject becomes higher than that of the TD subject.

3) Skewness of wrist velocities

As expected, in all measurements of this study, skewness was positive (right-skewed) due to the dominance of movements with lower velocities. As explained in (Meinecke et al., 2006), for the affected newborns (such as those with certain neurological or developmental conditions), movements often involve unusually high velocities. This causes the velocities distribution to shift toward the right (positive skewness). This shift happens because high velocities are outliers compared to the majority of slower movement velocities, thus stretching the distribution to the right. However, contrary to what is reported in (Meinecke et al., 2006), in our acquisition, TD subject shows slow movements along with repetitive, and intentional high-speed movements. For this reason, the velocity distribution could become positively skewed (higher skewness) reflecting the presence of higher and unequally distributed velocity values compared to those from AR who show slower and equally distributed movements (lower skewness) (**Figure 84**).

4) Range of motion of elbow angle

It is evident from **Figure 85** that the average elbow's ROM follows a similar trend to the wrist's average tangential velocity. As age increases, TD tends to exhibit wider movements compared to those with AR. This demonstrates that this metric effectively indicates the differences observed in the video recording between AR and TD.

5) Stereotypy score

The lack of variation in AR’s movements at all time points is demonstrated by the stereotypy score. This score measures the repetitiveness and predictability of movements. Higher stereotypy scores indicate more repetitive and less variable movements, which is characteristic of AR individuals (Karch et al., 2008). In contrast, TD tends to have lower stereotypy scores, reflecting more varied and adaptive movements. The stereotypy score effectively highlights the differences in movement patterns between AR and TD, showing that AR individual exhibits consistent, repetitive movements with little variation across different time points (**Figure 86**). This further highlights the distinct motor development trajectories between AR and TD.

6) Jerk Index

There is no evident trend in the jerk index for both TD and AR individuals, as this measure is significantly affected by noise in the positioning of PoIs, given that it is calculated as the third derivative of the trajectories (**Figure 86**). Moreover, the reference article (Kanemaru et al., 2013a) did not highlight any statistically significant differences between TD and AR.

7) Kurtosis of wrist acceleration

In the reference article (Kanemaru et al., 2013a), wrist acceleration kurtosis was found to be higher in AR compared to TD. In contrast, the current study finds the opposite trend **Figure 86**. One of the reasons could be that the kurtosis values of wrist acceleration are influenced by the presence of repetitive and intentional movements, such as clapping hands, hitting the table, and bending forward. These activities were noted in the TD subject at timepoints 2, 3, and 7 and they may lead to a more leptokurtic distribution (higher kurtosis). Such a distribution features heavier tails because these extreme movements occur more frequently than would be expected in a normal distribution.

8) Correlation between wrists velocity

Values found in this study (**Figure 87**) are in line with those in the reference article since the AR subject exhibited less variability in their movements compared to TD subjects.

9) Wrist average tangential velocity

Values found in this study are in line with those in the reference article (Kanemaru et al., 2013a). Interestingly, the wrist average tangential velocity shows an inverse trend between TD and AR subjects across the timepoints, demonstrating that as age increases, the movements of the subjects tend to be divergent as shown in **Figure 87**.

It is worth noticing that for the majority of parameters , the values obtained from the proposed MS protocol are of the same order of magnitude as those reported in the reference studies. However, it is not possible to make the same comparison for the jerk index because, in the reference article (Kanemaru et al., 2013a), the PoIs coordinates are normalized with respect to the trunk length. Similarly, it is not possible to compare the areas out of the standard deviation or out of the moving average for trajectories/velocities because the reference article (Disselhorst-Klug et al., 2012) does not clearly specify the units of measurement used.

In conclusion, in the estimates extracted from this study, significant variability was observed due to the fact that the data collection was conducted in the subject's home, an environment that is extremely uncontrolled. For instance, variability can arise from factors such as the choice of seating. It must be highlighted that also the presence of other people in the room can further encourage the intentional movements of the child during data recording.

References

Adde, L., Helbostad, J. L., Jensenius, A. R., Taraldsen, G., Grunewaldt, K. H., & StØen, R. (2010). Early prediction of cerebral palsy by computer-based video analysis of general movements: A feasibility study. *Developmental Medicine and Child Neurology*, *52*(8), 773–778. https://doi.org/10.1111/j.1469-8749.2010.03629.x

Balta, D., Figari, G., Paolini, G., Pantzar-Castilla, E., Riad, J., Croce, U. D., & Cereatti, A. (2023). 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*, 1. https://doi.org/10.1109/ACCESS.2023.3340622

Balta, D., Kuo, H. H., Wang, J., Porco, I. G., Morozova, O., Schladen, M. M., Cereatti, A., Lum, P. S., & Della Croce, U. (2022). Characterization of Infants' General Movements Using a Commercial RGB-Depth Sensor and a Deep Neural Network Tracking Processing Tool: An Exploratory Study. *Sensors*, *22*(19). https://doi.org/10.3390/s22197426

Balta, D., Kuo, H., Wang, J., Porco, I., Schladen, M., Cereatti, A., Cereatti, A., & Della Croce, U. (2022). Infant upper body 3D kinematics estimated using a commercial RGB-D sensor and a deep neural network tracking processing tool. *17th Edition of IEEE International Symposium on Medical Measurements and Applications (MeMeA)*.

Balta, D., Kuo, H., Wang, J., Porco, I., Schladen, M., Cereatti, A., Lum, P., & Della Croce, U. (2022). Estimating infant upper extremities motion with an RGB-D camera and markerless deep neural network tracking: a validation study. *44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*.

Balta, D., Salvi, M., Molinari, F., Figari, G., Paolini, G., Croce, U. Della, & Cereatti, A. (2020). A two-dimensional clinical gait analysis protocol based on markerless recordings from a single RGB-Depth camera. *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 1–6. https://doi.org/10.1109/MeMeA49120.2020.9137183

Benedetti, M. G., Catani, F., Leardini, A., Pignotti, E., & Giannini, S. (1998). Data management in gait analysis for clinical applications. *Clinical Biomechanics*, *13*(3), 204–215. https://doi.org/https://doi.org/10.1016/S0268-0033(97)00041-7

Bergamini, E., Ligorio, G., Summa, A., Vannozzi, G., Cappozzo, A., & Sabatini, A. M. (2014). Estimating orientation using magnetic and inertial sensors and different sensor fusion approaches: Accuracy assessment in manual and locomotion tasks. *Sensors (Switzerland)*, *14*(10), 18625–18649. https://doi.org/10.3390/s141018625

Büker, L., Quinten, V., Hackbarth, M., Hellmers, S., Diekmann, R., & Hein, A. (2023). How the Processing Mode Influences Azure Kinect Body Tracking Results. *Sensors*, *23*(2). https://doi.org/10.3390/s23020878

Cappozzo, A., Della Croce, U., Leardini, A., & Chiari, L. (2005). Human movement analysis using stereophotogrammetry. Part 1: Theoretical background. In *Gait and Posture* (Vol. 21, Issue 2, pp. 186–196). Elsevier Ireland Ltd. https://doi.org/10.1016/j.gaitpost.2004.01.010

Castelli, A., Paolini, G., Cereatti, A., & Della Croce, U. (2015). A 2D markerless gait analysis methodology: Validation on healthy subjects. *Computational and Mathematical Methods in Medicine*, *2015*. https://doi.org/10.1155/2015/186780

Cereatti, A., Margheritini, F., Donati, M., & Cappozzo, A. (2010). Is the human acetabulofemoral joint spherical? *The Journal of Bone & Joint Surgery British Volume*, *92-B*(2), 311–314. https://doi.org/doi:10.1302/0301-620X.92B2.22625

Christensen, D., Van Naarden Braun, K., Doernberg, N. S., Maenner, M. J., Arneson, C. L., Durkin, M. S., Benedict, R. E., Kirby, R. S., Wingate, M. S., Fitzgerald, R., & Yeargin-Allsopp, M. (2014). Prevalence of cerebral palsy, co-occurring autism spectrum disorders, and motor functioning - Autism and Developmental Disabilities Monitoring Network, USA, 2008. *Developmental Medicine and Child Neurology*, *56*(1), 59–65. https://doi.org/10.1111/dmcn.12268

Clark, R. A., Bower, K. J., Mentiplay, B. F., Paterson, K., & Pua, Y. H. (2013). Concurrent validity of the Microsoft Kinect for assessment of spatiotemporal gait variables. *Journal of Biomechanics*, *46*(15), 2722–2725. https://doi.org/10.1016/j.jbiomech.2013.08.011

Clark, R. A., Mentiplay, B. F., Hough, E., & Pua, Y. H. (2019). Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and Kinect alternatives. In *Gait and Posture* (Vol. 68, pp. 193–200). Elsevier B.V. https://doi.org/10.1016/j.gaitpost.2018.11.029

D'Antonio, E., Taborri, J., Mileti, I., Rossi, S., & Patane, F. (2021). Validation of a 3D Markerless System for Gait Analysis Based on OpenPose and Two RGB Webcams. *IEEE Sensors Journal*, *21*(15), 17064–17075. https://doi.org/10.1109/JSEN.2021.3081188

Davis, R. B., Õunpuu, S., Tyburski, D., & Gage, J. R. (1991). A gait analysis data collection and reduction technique. *Human Movement Science*, *10*(5), 575–587. https://doi.org/https://doi.org/10.1016/0167-9457(91)90046-Z

Disselhorst-Klug, C., Heinze, F., Breitbach-Faller, N., Schmitz-Rode, T., & Rau, G. (2012). Introduction of a method for quantitative evaluation of spontaneous motor activity development with age in infants. *Experimental Brain Research*, *218*(2), 305–313. https://doi.org/10.1007/s00221-012-3015-x

Einspieler, C., & Prechtl, H. F. R. (2005). Prechtl's assessment of general movements: A diagnostic tool for the functional assessment of the young nervous system. *Mental Retardation and Developmental Disabilities Research Reviews*, *11*(1), 61–67. https://doi.org/https://doi.org/10.1002/mrdd.20051

El-Tallawy, H. N., Farghaly, W. M. A., Shehata, G. A., Rageh, T. A., Metwally, N. A., Badry, R., Sayed, M. A. M., El Hamed, M. A., Abd-Elwarth, A., & Kandil, M. R. (2014). Cerebral palsy in Al-Quseir City, Egypt: Prevalence, subtypes, and risk factors. *Neuropsychiatric Disease and Treatment*, *10*, 1267–1272. https://doi.org/10.2147/NDT.S59599

Graham, H. K., Rosenbaum, P. L., Paneth, N. S., Dan, B., Lin, J.-P., Damiano, D. L., Becher, J. G., Gaebler-Spira, D. J., Colver, A., Reddihough, D. S., Crompton, K. E., & Lieber, R. L. (2016). Cerebral palsy. *Nature Reviews Disease Primers*, *2*.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition*. http://arxiv.org/abs/1512.03385

Heinz, P., Prechtl, F. R., Prechtl, H. F. R., Einspieler, C., Cioni, G., Bos, F., Ferrari, F., & Sontheimer, D. (1997). An early marker for neurological deficits after perinatal brain lesions. In *THE LANCET* (Vol. 349).

Hekne, L., Montgomery, C., & Johansen, K. (2021). Early access to physiotherapy for infants with cerebral palsy: A retrospective chart review. In *PLoS ONE* (Vol. 16, Issue 6 June). Public Library of Science. https://doi.org/10.1371/journal.pone.0253846

Ihlen, E. A. F., Støen, R., Boswell, L., de Regnier, R. A., Fjørtoft, T., Gaebler-Spira, D., Labori, C., Loennecken, M. C., Msall, M. E., Möinichen, U. I., Peyton, C., Schreiber, M. D., Silberg, I. E., Songstad, N. T., Vågen, R. T., Øberg, G. K., & Adde, L. (2020). Machine learning of infant spontaneous movements for the early prediction of cerebral palsy: A multi-site cohort study. *Journal of Clinical Medicine*, *9*(1). https://doi.org/10.3390/jcm9010005

Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., & Schiele, B. (2016). *DeeperCut: A Deeper, Stronger, and Faster Multi-Person Pose Estimation Model*. http://arxiv.org/abs/1605.03170

Kanemaru, N., Watanabe, H., Kihara, H., Nakano, H., Takaya, R., Nakamura, T., Nakano, J., Taga, G., & Konishi, Y. (2013a). Specific characteristics of spontaneous movements in preterm infants at term age are associated with developmental delays at age 3 years. *Developmental Medicine and Child Neurology*, *55*(8), 713–721. https://doi.org/10.1111/dmcn.12156

Kanemaru, N., Watanabe, H., Kihara, H., Nakano, H., Takaya, R., Nakamura, T., Nakano, J., Taga, G., & Konishi, Y. (2013b). Specific characteristics of spontaneous movements in preterm infants at term age are associated with developmental delays at age 3 years. *Developmental Medicine and Child Neurology*, *55*(8), 713–721. https://doi.org/10.1111/dmcn.12156

Kanko, R. M., Laende, E. K., Davis, E. M., Selbie, W. S., & Deluzio, K. J. (2021). Concurrent assessment of gait kinematics using marker-based and markerless motion capture. *Journal of Biomechanics*, *127*. https://doi.org/10.1016/j.jbiomech.2021.110665

Karch, D., Kang, K. S., Wochner, K., Philippi, H., Hadders-Algra, M., Pietz, J., & Dickhaus, H. (2012a). Kinematic assessment of stereotypy in spontaneous movements in infants. *Gait and Posture*, *36*(2), 307–311. https://doi.org/10.1016/j.gaitpost.2012.03.017

Karch, D., Kang, K. S., Wochner, K., Philippi, H., Hadders-Algra, M., Pietz, J., & Dickhaus, H. (2012b). Kinematic assessment of stereotypy in spontaneous movements in infants. *Gait and Posture*, *36*(2), 307–311. https://doi.org/10.1016/j.gaitpost.2012.03.017

Karch, D., Kim, K. S., Wochner, K., Pietz, J., Dickhaus, H., & Philippi, H. (2008). Quantification of the segmental kinematics of spontaneous infant movements. *Journal of Biomechanics*, *41*(13), 2860–2867. https://doi.org/10.1016/j.jbiomech.2008.06.033

Lv, X., Wang, S., Chen, T., Zhao, J., Chen, D., Xiao, M., Zhao, X., & Wei, H. (2021). Human Gait Analysis Method Based on Sample Entropy Fusion AlphaPose Algorithm. *Proceedings of the 33rd Chinese Control and Decision Conference, CCDC 2021*, 1543–1547. https://doi.org/10.1109/CCDC52312.2021.9602427

Ma, Y., Mithraratne, K., Wilson, N. C., Wang, X., Ma, Y., & Zhang, Y. (2019). The validity and reliability of a kinect v2-based gait analysis system for children with cerebral palsy. *Sensors (Switzerland)*, *19*(7). https://doi.org/10.3390/s19071660

Marcroft, C., Khan, A., Embleton, N. D., Trenell, M., & Plötz, T. (2015). Movement recognition technology as a method of assessing spontaneous general movements in high risk infants. In *Frontiers in Neurology* (Vol. 6, Issue JAN, p. 284). Frontiers Media S.A. https://doi.org/10.3389/fneur.2014.00284

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

Mazzarella, J., McNally, M., Richie, D., Chaudhari, A. M. W., Buford, J. A., Pan, X., & Heathcock, J. C. (2020). 3d motion capture may detect spatiotemporal changes in pre-reaching upper extremity movements with and without a real-time constraint condition in infants with perinatal stroke and cerebral palsy: A longitudinal case series. *Sensors (Switzerland)*, *20*(24), 1–17. https://doi.org/10.3390/s20247312

McIntyre, S., Morgan, C., Walker, K., & Novak, I. (2011). Cerebral palsy-Don't delay. *Developmental Disabilities Research Reviews*, *17*(2), 114–129. https://doi.org/10.1002/ddrr.1106

Meinecke, L., Breitbach-Faller, N., Bartz, C., Damen, R., Rau, G., & Disselhorst-Klug, C. (2006). Movement analysis in the early detection of newborns at risk for developing spasticity due to infantile cerebral palsy. *Human Movement Science*, *25*(2), 125–144. https://doi.org/10.1016/j.humov.2005.09.012

Metz, C., Jaster, M., Walch, E., Sarpong-Bengelsdorf, A., Kaindl, A. M., & Schneider, J. (2022). Clinical Phenotype of Cerebral Palsy Depends on the Cause: Is It Really Cerebral Palsy? A Retrospective Study. *Journal of Child Neurology*, *37*(2), 112–118. https://doi.org/10.1177/08830738211059686

Moro, M., Marchesi, G., Hesse, F., Odone, F., & Casadio, M. (2022). Markerless vs. Marker-Based Gait Analysis: A Proof of Concept Study. *Sensors*, *22*(5). https://doi.org/10.3390/s22052011

Moro, M., Marchesi, G., Odone, F., & Casadio, M. (2020). Markerless gait analysis in stroke survivors based on computer vision and deep learning: A pilot study. *Proceedings of the ACM Symposium on Applied Computing*, 2097–2104. https://doi.org/10.1145/3341105.3373963

Moro, M., Pastore, V. P., Tacchino, C., Durand, P., Blanchi, I., Moretti, P., Odone, F., & Casadio, M. (2022). A markerless pipeline to analyze spontaneous movements of preterm infants. *Computer Methods and Programs in Biomedicine*, *226*. https://doi.org/10.1016/j.cmpb.2022.107119

Novak, I., Morgan, C., Adde, L., Blackman, J., Boyd, R. N., Brunstrom-Hernandez, J., Cioni, G., Damiano, D., Darrah, J., Eliasson, A.-C., de Vries, L. S., Einspieler, C., Fahey, M., Fehlings, D., Ferriero, D. M., Fetters, L., Fiori, S., Forssberg, H., Gordon, A. M., … Badawi, N. (2017). Early, Accurate Diagnosis and Early Intervention in Cerebral Palsy: Advances in Diagnosis and Treatment. *JAMA Pediatrics*, *171*(9), 897–907. https://doi.org/10.1001/jamapediatrics.2017.1689

Oskoui, M., Coutinho, F., Dykeman, J., Jetté, N., & Pringsheim, T. (2013). An update on the prevalence of cerebral palsy: A systematic review and meta-analysis. In *Developmental Medicine and Child Neurology* (Vol. 55, Issue 6, pp. 509–519). https://doi.org/10.1111/dmcn.12080

Prechtl, H. F. R. (1990). Qualitative changes of spontaneous movements in fetus and preterm infant are a marker of neurological dysfunction. *Early Human Development*, *23*(3), 151–158. https://doi.org/https://doi.org/10.1016/0378-3782(90)90011-7

Rajchl, M., Lee, M. C. H., Oktay, O., Kamnitsas, K., Passerat-Palmbach, J., Bai, W., Damodaram, M., Rutherford, M. A., Hajnal, J. V., Kainz, B., & Rueckert, D. (2016). *DeepCut: Object Segmentation from Bounding Box Annotations using Convolutional Neural Networks*. http://arxiv.org/abs/1605.07866

Romeo, L., Marani, R., Malosio, M., Perri, A. G., & D'Orazio, T. (2021). Performance analysis of body tracking with the microsoft azure kinect. *2021 29th Mediterranean Conference on Control and Automation, MED 2021*, 572–577. https://doi.org/10.1109/MED51440.2021.9480177

Ruescas-Nicolau, A. V., Medina-Ripoll, E., de Rosario, H., Sanchiz Navarro, J., Parrilla, E., & Juan Lizandra, M. C. (2024). A Deep Learning Model for Markerless Pose Estimation Based on Keypoint Augmentation: What Factors Influence Errors in Biomechanical Applications? *Sensors*, *24*(6). https://doi.org/10.3390/s24061923

Sakoe, H. (1978). Dynamic Programming Algorithm Optimization for Spoken Word Recognition. In *IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING* (Issue 1).

Schindelin, J., Rueden, C. T., Hiner, M. C., & Eliceiri, K. W. (2015). The ImageJ ecosystem: An open platform for biomedical image analysis. In *Molecular Reproduction and Development* (Vol. 82, Issues 7–8, pp. 518–529). John Wiley and Sons Inc. https://doi.org/10.1002/mrd.22489

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., & Blake, A. (n.d.). *Real-Time Human Pose Recognition in Parts from Single Depth Images*.

Silva, N., Zhang, D., Kulvicius, T., Gail, A., Barreiros, C., Lindstaedt, S., Kraft, M., Bölte, S., Poustka, L., Nielsen-Saines, K., Wörgötter, F., Einspieler, C., & Marschik, P. B. (2021). The future of General Movement Assessment: The role of computer vision and machine learning – A scoping review. *Research in Developmental Disabilities*, *110*. https://doi.org/10.1016/j.ridd.2021.103854

Stagni, R., Doto, T., Tomadin, A., Sansavini, A., Aceti, A., Corvaglia, L. T., & Bisi, M. C. (2023). General movements automatic assessment: Methodological issues for pose estimation. *Gait & Posture*, *106*, S195–S196. https://doi.org/https://doi.org/10.1016/j.gaitpost.2023.07.236

Stenum, J., Rossi, C., & Roemmich, R. T. (2021). Two-dimensional video-based analysis of human gait using pose estimation. *PLoS Computational Biology*, *17*(4). https://doi.org/10.1371/journal.pcbi.1008935

Surer, E., Cereatti, A., Grosso, E., & Croce, U. Della. (2011). A markerless estimation of the ankle-foot complex 2D kinematics during stance. *Gait and Posture*, *33*(4), 532–537. https://doi.org/10.1016/j.gaitpost.2011.01.003

Te Velde, A., Morgan, C., Novak, I., Tantsis, E., & Badawi, N. (2019). Early diagnosis and classification of cerebral palsy: An historical perspective and barriers to an early diagnosis. *Journal of Clinical Medicine*, *8*(10). https://doi.org/10.3390/jcm8101599

Wang, H. H., Hwang, Y. S., Ho, C. H., Lai, M. C., Chen, Y. C., & Tsai, W. H. (2021). Prevalence and initial diagnosis of cerebral palsy in preterm and term-born children in taiwan: A nationwide, population-based cohort study. *International Journal of Environmental Research and Public Health*, *18*(17). https://doi.org/10.3390/ijerph18178984

Wu, G., Siegler, S., Allard, P., Kirtley, C., Leardini, A., Rosenbaum, D., Whittle, M., D'Lima, D. D., Cristofolini, L., Witte, H., Schmid, O., & Stokes, I. (2002). ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—part I: ankle, hip, and spine. *Journal of Biomechanics*, *35*(4), 543–548. https://doi.org/https://doi.org/10.1016/S0021-9290(01)00222-6

Yamamoto, M., Shimatani, K., Hasegawa, M., Kurita, Y., Ishige, Y., & Takemura, H. (2021). Accuracy of Temporo-Spatial and Lower Limb Joint Kinematics Parameters Using OpenPose for Various Gait Patterns with Orthosis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 2666–2675. https://doi.org/10.1109/TNSRE.2021.3135879

Chapter 4

4. Conclusions

Movement analysis is used to assess and diagnose movement disorders. The gold standard for movement analysis typically involves marker-based (MB) systems, which, although accurate, can be expensive and cumbersome. Recent advancements in computer vision algorithms have enabled more accessible and user-friendly markerless (MS) approaches. However, the accuracy and validity of MS methods for biomechanical and clinical applications remain an open issue (Lam et al., 2023; Wade et al., 2023), limiting their use in clinical settings. For screening and evaluating treatment purposes, portability, affordability, and user-friendliness are essential. Methods based on a single camera with minimal setup time are preferred. Inexpensive tracking systems featuring an RGB camera integrated with an infrared depth sensor (RGB-depth) have been developed, combining RGB images with depth data to generate depth color images (2D+) without requiring a multi-camera setup. Algorithms for estimating human motion from single video data are typically divided into deterministic and AI-based approaches. Deterministic methods rely on formulas and clear rules for defining joint centers and use predefined kinematic models to track or match against the video data or human anatomical proportions applied directly on video data. In contrast, AI-based approaches identify motion characteristics directly from the data using deep learning algorithms that automatically learn features from large datasets. Those methods are generally model-free (AI model-free approach) but can also include predetermined model to enhance joint centers estimates (AI model-based approach).

Several studies, belonging to both categories, have been proposed regarding both markerless gait analysis and upper-limb movement analysis.

Clinical gait analysis is essential for understanding and interpreting the physio-pathological characteristics of human locomotion, and its diagnostic value is well-established. In this context, many of these single camera methods, however, have not been validated against clinically accepted standards on pathological populations, creating uncertainties about their accuracy in measuring joint movements for clinical applications (Amprimo et al., 2021; Balta et al., 2020; Castelli et al., 2015; Ferraris, Amprimo, Masi, et al., 2022; Latorre et al., 2018, 2019; Leu et al., 2011;Goffredo Michela and Carter, 2009) or the validation has been conducted only on a single joint (Hatamzadeh et al., 2022; Leu et al., 2011; Pantzar-Castilla et al., 2018; Surer et al., 2011). Some MS approaches focus on validating joint center positions (Hesse et al., 2023) or classifying motor activities and detecting gait abnormalities, rather than investigating kinematic and spatial-temporal parameters (Chen et al., 2011; Clark et al., 2015; Ferraris, Amprimo, Pettiti, et al., 2022; Kojovic et al., 2021; Li et al., 2018; Stricker et al., 2021). Methodological limitations, such as dependence on color filters and uniform backgrounds (Castelli et al., 2015; Pantzar-Castilla et al., 2018), also constrain their practical use in clinical settings.

This thesis aimed to fill these gaps in clinical gait analysis by (*1*) proposing and validating against a MB system original deterministic MS protocols based on a single RGB-D camera in patients with cerebral palsy and foot deformities and (*2*) by exploring the clinical validity of AI-based algorithms on healthy subjects.

Moreover, upper-limb movement analysis is particularly useful for the early detection of movement disorders in preterm infants. The General Movement Assessment, proposed by Heinz Prechtl, is considered the gold standard for identifying motor disorders early on, involving a qualitative and visual analysis of video recordings by a clinician. However, this method demands extensive training and a significant time to execute. While 3D MB analysis is highly accurate, it is not ideal because the markers placed on the infant's skin can interfere with their natural movements. Consequently, many studies have focused on using 2D video analysis through MS system (Adde et al., 2010; Ihlen et al., 2020; Moro et al., 2022; Stagni et al., 2023). However, a 3D analysis using a single RGB-D camera could be more useful and accurate, considering the inherently 3D nature of movement.

This thesis work provided the responses to the research questions outlined in Chapter 1:

1. This thesis developed a 2D MS protocol utilizing a single RGB-D camera, which was validated against a stereophotogrammetric system on subjects with cerebral palsy and foot deformities. The protocol showed errors between 2 to 5 degrees in joint kinematics estimation which are acceptable but have to be considered carefully for clinical interpretation.
2. An automatic segmentation algorithm was created in this thesis to identify subjects within the camera's field of view using a grayscale histogram approach. This method facilitates MS data acquisition without requiring a uniform background, simplifying experimental setups and making the process more adaptable to different environments.
3. The thesis introduced a 3D MS protocol based on a two-segment foot model to estimate the ankle and metatarsophalangeal joint angles. This represents a significant improvement over traditional MS gait analysis techniques that typically model the foot as a single rigid segment, thereby providing a more detailed and accurate evaluation of foot mechanics during gait.
4. A comparison was made between joint angles obtained from the Azure Kinect and those from a stereophotogrammetric system. The results showed substantial errors, particularly for the ankle joint (Mean Absolute Difference = 33 deg), indicating that the Azure Kinect SDK does not deliver sufficiently accurate estimates for clinical gait analysis, especially for critical joints like the ankle.
5. The thesis developed a 3D markerless protocol that compensates for out-of-plane movements using a single SMPL model. By incorporating a 3D subject-specific lower-limb model, this protocol advances beyond conventional 2D gait analysis, providing a more comprehensive assessment of gait. The protocol showed errors between 2 to 5 degrees in joint kinematics

estimation which are acceptable but have to be considered carefully for clinical interpretation.

6. Unlike existing studies that focus on general movements using 2D data, this thesis developed a markerless protocol to estimate 3D kinematics of the upper limb by identifying joint centers with DeepLabCut. The approach includes an algorithm to reconstruct the Z-coordinate of each point of interest, effectively managing body occlusions. This method allows the study of infant movements directly in their home environment, making assessments more accessible and less stressful for patients and their families.

7. The thesis also estimated a subset of metrics from the literature on a pair of twins with different health profiles, providing preliminary evidence that these metrics can effectively highlight differences between the subjects.

All details regarding the different research questions are provided in the following paragraphs. For simplicity, the main findings of this thesis are divided into MS protocols for gait analysis and for the analysis of upper limb movements in preterm infants.

4.1 Gait analysis protocols

Both 2D and 3D MS gait analysis protocols based on deterministic and AI-based model-based methods for gait analysis using a single RGB-D camera were validated against the gold standard (MB system), emphasizing their application in clinical settings for evaluating individuals with pathological gait patterns.

4.1.1 2D deterministic model-based protocol

In the proposed 2D deterministic model-based MS protocol, the subject's movement was represented using a 2D multi-segmental model derived from a single 2D RGB image. Depth sensor data were solely exploited to isolate foreground limb. 2D anatomical landmark coordinates were determined by independently matching the model's thigh, shank, and foot segments to their closest points in the 2D RGB images through the Iterative Closest Point Technique. Projecting 3D body motion into 2D introduces inherent errors and ambiguities, which were only partially mitigated by multiple anatomical calibrations. One significant issue was the RGB image quality captured by the specific RGB-D camera employed. The Kinect v2's automatic exposure settings often resulted in blurred images when recording rapid movements, negatively affecting the accuracy of iterative closest point algorithm. This problem could be addressed by selecting a camera that provides manual adjustment of exposure settings. The protocol is not entirely automated, as it requires the initial identification of external anatomical landmarks to create a static subject-specific model. However, since clinical gait analysis in CP patients typically follows a clinical examination where joint motion and spasticity are assessed, identifying a few anatomical landmarks is generally feasible for clinicians. This research validates the technical feasibility of the MS single-camera protocol for clinical gait analysis in individuals with CP. The results indicate high accuracy in estimating joint kinematics and demonstrate good to excellent reliability in computing a comprehensive set of clinically significant gait features.

4.1.2 Comparison between 2D deterministic model-based protocol and Azure Kinect body tracking SDK

Moreover, the 2D deterministic model-based protocol was compared to the Azure Kinect DK's body tracking SDK (SDK) alongside its potential for clinical applications. Lower limb joint kinematics were evaluated on five healthy subjects across ten gait trials with both MB system and MS protocols. However, data collection from both MS and MB systems was not simultaneous due to infrared interference. In order to compare the MS protocols against the MB system, the study extracted seven key gait variables from the sagittal lower-limb joint angles.

2D MS deterministic model-based protocol requires a manual input from an expert operator to create subject-specific models for each gait trial. Despite being less automatic, the 2D MS deterministic model-based protocol allowed for controlled outcomes to enhance clinical application accuracy. The SDK method includes a critical issue associated with left-right confusion (Nguyen et al., 2022). In a clinical setting, such errors can lead to significantly misleading information, potentially compromising the diagnostic and treatment processes. Both SDK and 2D MS deterministic model-based methods performed well in estimating knee and hip angles compared to the MB method, considered the gold standard. However, the SDK method faced significant limitations in evaluating the ankle angle, with 2D MS deterministic model-based protocol providing visibly better performance. The study concluded that the SDK method, as implemented, is not suitable for clinical gait analysis due to performance limitations particularly for the ankle angle computation. Despite its rapid implementation and versatility, the SDK method's major drawback is its "black box" nature, offering no control over the output.

4.1.3 Comparison between 2D and 3D deterministic model-based protocols for estimating sagittal lower limb joint kinematics

To improve the accuracy in estimating the sagittal lower limb kinematics using the 2D MS deterministic model-based protocol, a 3D MS deterministic protocol based on a single RGB-Depth camera with the benefit of a 3D statistical lower-limb model was proposed. It should be noted that a 3D subject-specific model was reconstructed using a single RGB-D camera by merging only three views, aiming to minimize patient discomfort, and maximizing the system’s portability. This approach is preferred over a multi-camera or 3D scanner system, which increases complexity, incurs higher costs, and requires longer setup times confining them to specialized laboratories. The 3D approach offers several advantages over the 2D protocol. It eliminates the need for manual input from the operator to identify the 2D locations of anatomical landmarks during the calibration of three models (static, loading, and swing models). In addition, given the use of a 3D model, there is no longer a need for three separate 2D models for partially compensate for the movements outside the sagittal plane. However, a primary limitation remains the use of a single view, which, while allowing for portability, does not provide information on internal-external and abduction-adduction rotations. The current method performs 3D fitting between the 3D model and a 3D point cloud for each frame of the gait cycle using the Articulated ICP algorithm, contrasting with the 2D approach that involved fitting the kinematic model on 2D images through the ICP algorithm. This allows for directly estimating the 3D joint centers' trajectories unlike the previous method that tracked anatomical landmarks on 2D images. The identification of joint centers is entirely automatic; the fixed morphology of the model allows for an easy extraction of their positions for each dynamic frame using the joint regression matrix provided by the Max Planck Institute.

In comparing the two methods, it is important to note that the 2D MS method was validated on eighteen subjects with cerebral palsy, while the 3D MS method was validated on ten subjects, including 6 with cerebral palsy and 4 with clubfeet. The 2D method was validated by synchronous comparison with MB system, whereas the 3D method did not involve simultaneous acquisitions due to infrared interference between IR sensors.

The 3D method demonstrates comparable performance to the 2D MS protocol in terms of mean absolute error against MB system for gait features related to the hip (4.2 deg vs 3.7 deg), ankle (3.8 deg vs 3.5 deg), and knee joint (4.0 deg vs 4.3 deg). Furthermore, this 3D method showed high reliability (ICC >0.75) for each feature extracted from 3D MS protocol, comparable to the values found using the MB system except for A5 which is highly affected by the quality of the depth images during the highest velocities. This indicates that the 3D method can provide automated estimates that are less sensitive to out-of-plane movements without the need to include three models and a manual identification of specific anatomical landmarks as done in the 2D MS protocol. It is important to highlight that this method has several areas for improvement. Firstly, depth sensing technology is undergoing rapid advancements. Further improvements could include a proper metrological characterization of the depth sensor under dynamic conditions and the inclusion of depth completion techniques to enhance model fitting, especially during the swing phase. Additionally, a proper standardization of the protocol for static acquisitions could improve the reconstruction of the 3D model. Finally, this method could be enriched by including the estimation of volumetric parameters to evaluate and monitor asymmetries, going beyond the traditional gait analysis.

4.1.4 3D deterministic model-based protocol for estimating sagittal foot kinematics using a 3D two-segments model

Finally, the study of foot kinematics using a single RGB-D camera was explored. MS alternatives based on a single RGB-Depth camera (e.g., Azure Kinect Body tracking SDK, OpenPose) allow modeling the foot as a single segment without articulating the metatarsophalangeal (MTP) joint, which is crucial for effective foot loading and correct progression (Dobbs & Gurnett, 2017). This study aimed to design a 3D MS method based on a single RGB-Depth camera for estimating sagittal ankle and MTP kinematics using a 3D foot model composed by two segments (mid rear foot and forefoot) connected by a revolute joint (metatarsophalangeal joint) and explore its clinical applicability on children with foot deformities. The proposed 3D protocol is structured in two parts: creation of a 2-segment 3D foot template by merging four static views of the foot (Frontal, Lateral, Medial, Posterior) aligning three common points identified on the foot sole of each view. Then, a 2D shank template was obtained as in (Balta et al., 2023). Both foot and shank templates were calibrated by manually identifying the lateral epicondyle, lateral malleolus, the fifth metatarsophalangeal joint, and toe on the RGB static image. Estimation of the joint kinematics during a gait cycle was achieved by implementing a depth completion technique to reconstruct missing depth information during the gait trials by exploiting RGB information. The positions of the ankle and fifth metatarsophalangeal joint were reconstructed by matching the 3D foot template to the dynamic point clouds applying the ICP algorithm (Besl & McKay, 1992), while the knee joint center position was extracted by implementing the 2D markerless protocol proposed by (Balta et al., 2020, 2023).

The experimental session was conducted on ten subjects affected by clubfoot who were asked to walk straight at a self-selected speed for six trials (three for left and three for right) in front of the camera, which was placed laterally to the walkway. The computed joint angles were validated by comparing them with those obtained from manually labeled anatomical landmarks on RGB images. Considering the limited number of views required for the creation of the 3D model, the reconstruction accuracy showed promising results, represented by a mean percentage error of 5.2% for the right foot and 5.7% for the left foot. The acquisition protocol is specifically designed to minimize any discomfort that the subject may experience during the upright static position. However, it is important to note that incorrect positioning of the subject during the static acquisition phase can introduce errors and challenges in aligning the different views. Therefore, ensuring proper subject positioning is crucial to mitigate such issues and enhance the accuracy of the reconstruction process. Specifically, as regards joint kinematics, the average RMSE for the MTP joint is 5 deg while for the ankle, it is 4.8 deg. The reported errors are mostly associated with the technological limitations of the RGB-Depth device employed. In particular, during high-speed movements, especially during the swing phase, the depth sensor of RGB-Depth cameras may fail to accurately reconstruct depth values due to motion blurs, leading to artifacts such as holes or fake boundaries. This limitation could be improved by implementing a depth sensor characterization in dynamic conditions to develop more suitable depth completion techniques (e.g., inpainting-based, or deep learning-based models). The presence of missing points during the fitting process affects the accuracy of ICP, especially in the forefoot segments (Forefoot to Mid-Rear-foot points ratio = 0.26), leading to errors in estimating the position of the fifth MTP joint. Residual errors in estimating joint kinematics are due to the assumption of rigid segments in the ICP algorithm. From a clinical perspective, a limitation of this method is that the acquisition was performed with the camera laterally to the walkway, not considering the kinematics of the first metatarsophalangeal joint, which is a significant indicator of gait quality compared to the fifth joint (Allan et al., 2020). In conclusion, considering the rapid technological advancement in depth sensing, the proposed approach is a very promising solution, in terms of preparation and acquisition time and effective cost, to evaluate the gait of subjects with foot deformities.

Overall, advancements in 3D modeling signify a substantial step forward in achieving accurate, automated gait analysis. Future work should focus on addressing current technological limitations, optimizing depth quality images, and expanding the protocols to include comprehensive volumetric features for enhanced clinical applicability.

4.2 Upper-limb movement analysis protocols

This study demonstrates the feasibility of measuring general movements (GM) metrics using a single, inexpensive RGB-D sensor. The simplicity and portability of the markerless protocol enables its use as a screening tool in homes or other familiar settings, avoiding clinical environments that can complicate the assessment of genuine neurodevelopmental performance.

Unlike previous research, this study describes a method to characterize GMs without attaching markers to the infant's skin, which could affect natural movements and the infant's behavioral state. This MS approach provides 3D coordinates for each point of interest (PoI) using DeepLabCut algorithm, which represents an improvement over traditional 2D motion capture systems that struggle with out-of-plane rotations, thus offering a more accurate depiction of GMs. The depth information captured by the RGB-D sensor allows addressing PoI occlusions often encountered in single-camera motion analyses. A proper training set construction was proposed to reduce computational time for manually labeling the PoIs, even including biomechanical domain knowledge. To improve the robustness of the algorithm, the 3D PoIs coordinates have been referenced to a local coordinate system to compensate for accidental movements of the seat or the camera. This approach allows for the extraction of metrics, selected from the literature, capable of describing the infants' spontaneous movements. The system is particularly suited for home environments, potentially improving the screening for neurodevelopmental disorders, especially for infants and families in rural and remote areas with limited access to healthcare services.

The first case study explained in paragraph “Case study 1: Exploratory Analysis of General Movements in 3-5 Month Old Infants Using a single RGB-Depth sensor” involved recording the GM of eight infants at home at 3, 4, and 5 months old. Eight GM metrics from the literature (Disselhorst-Klug et al., 2012; Meinecke et al., 2006), along with a novel metric (range of motion of the elbow angle), were estimated from the 3D PoI trajectories at each timepoint. A pediatric neurologist and a physiatrist provided an overall clinical evaluation based on the infants' videos. Subsequently, a comparison between the metrics and the clinical evaluations was performed. Due to the small sample size, it was not possible to conduct meaningful statistical analyses. Additionally, environmental factors and the infants' postures during recording influenced the results. The clinical evaluation was not a formal GMA; rather, each clinician was asked to indicate if they noticed any developmental concerns in the infants. Despite these issues, the results demonstrated that GM metrics can be meaningfully estimated and potentially used for the early identification of movement disorders.

The second case study explained in paragraph “Case study 2: Assessment of quantitative metrics for spontaneous movements analysis on Twins with Divergent Health Profiles” involved a pair of twins, one at-risk (AR) and the other typically developing (TD), with data collected at seven different time points from 4 to 10 months of age. The study showed that the majority of the metrics (i.e. wrist average tangential velocity, range of motion of elbow angle, stereotypy score, area out of mean and the standard deviation of wrist trajectories and velocities) proposed in the literature can effectively describe differences between AR and TD and align with previous research findings (Disselhorst-Klug et al., 2012) . However, other metrics are significantly affected by noise in the estimation of PoI coordinates and are influenced by intentional movements, which are particularly common in infants older than five months (kurtosis of wrist acceleration, skewness of wrist velocities and periodicity index).

Even in this case study on twins, the limitations of the study include significant variability in the estimates due to environmental factors, such as the use of different baby chairs or the presence of other people in the scene. Nonetheless, these aspects are real-world factors likely encountered when data is collected in an uncontrolled environment such as home. Therefore, to conduct this analysis at home, future studies will focus on further strengthening the MS algorithms to consider these factors.

The overall study proposed in this thesis has demonstrated that both deterministic and AI-based approaches can, when properly configured, extract parameters of interest for clinical movement analysis. The proposed deterministic model-based gait analysis methods offer errors ranging between 2 deg and 5 deg compared to MB system, which are considered acceptable but must be considered for clinical analyses (McGinley et al., 2009).

This thesis has shown that Azure Kinect body tracking SDK is difficult to control as it is not properly tuned and can introduce errors that might compromise clinical estimates. Open-source AI-based methods, such as DeepLabCut, have proven useful for being adapted to challenging cases like studying general movements in preterm infants at home.

In conclusion, this thesis demonstrates that markerless approaches are increasingly gaining momentum in clinical movement analysis thanks to advancements in technology, computer vision techniques, and machine learning. It is worth noticing that the codes developed in this work have been made available online. It is important to mention that the clinical applicability of the deterministic model-based methods for gait analysis developed in this thesis is currently being evaluated at both Skaraborg Hospital in Skövde and the outpatient clinics of ASL TO5 in Turin. The ultimate goal is to integrate markerless gait analysis into routine clinical practice

References

Adde, L., Helbostad, J. L., Jensenius, A. R., Taraldsen, G., Grunewaldt, K. H., & StØen, R. (2010). Early prediction of cerebral palsy by computer-based video analysis of general movements: A feasibility study. *Developmental Medicine and Child Neurology*, *52*(8), 773–778. https://doi.org/10.1111/j.1469-8749.2010.03629.x

Allan, J. J., McClelland, J. A., Munteanu, S. E., Buldt, A. K., Landorf, K. B., Roddy, E., Auhl, M., & Menz, H. B. (2020). First metatarsophalangeal joint range of motion is associated with lower limb kinematics in individuals with first metatarsophalangeal joint osteoarthritis. *Journal of Foot and Ankle Research*, *13*(1). https://doi.org/10.1186/s13047-020-00404-0

Amprimo, G., Pettiti, G., Priano, L., Mauro, A., & Ferraris, C. (2021). *Kinect-based solution for the home monitoring of gait and balance in elderly people with and without neurological diseases*.

Balta, D., Figari, G., Paolini, G., Pantzar-Castilla, E., Riad, J., Croce, U. D., & Cereatti, A. (2023). 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*, 1. https://doi.org/10.1109/ACCESS.2023.3340622

Balta, D., Salvi, M., Molinari, F., Figari, G., Paolini, G., Croce, U. Della, & Cereatti, A. (2020). A two-dimensional clinical gait analysis protocol based on markerless recordings from a single RGB-Depth camera. *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 1–6. https://doi.org/10.1109/MeMeA49120.2020.9137183

Castelli, A., Paolini, G., Cereatti, A., & Della Croce, U. (2015). A 2D markerless gait analysis methodology: Validation on healthy subjects. *Computational and Mathematical Methods in Medicine*, *2015*. https://doi.org/10.1155/2015/186780

Chen, S. W., Lin, S. H., Liao, L. De, Lai, H. Y., Pei, Y. C., Kuo, T. S., Lin, C. T., Chang, J. Y., Chen, Y. Y., Lo, Y. C., Chen, S. Y., Wu, R., & Tsang, S. (2011). Quantification and recognition of parkinsonian gait from monocular video imaging using kernel-based principal component analysis. *BioMedical Engineering Online*, *10*. https://doi.org/10.1186/1475-925X-10-99

Clark, R. A., Vernon, S., Mentiplay, B. F., Miller, K. J., McGinley, J. L., Pua, Y. H., Paterson, K., & Bower, K. J. (2015). Instrumenting gait assessment using the Kinect in people living with stroke: reliability and association with balance tests. *Journal of Neuroengineering and Rehabilitation*, *12*, 15. https://doi.org/10.1186/s12984-015-0006-8

Disselhorst-Klug, C., Heinze, F., Breitbach-Faller, N., Schmitz-Rode, T., & Rau, G. (2012). Introduction of a method for quantitative evaluation of spontaneous motor activity development with age in infants. *Experimental Brain Research*, *218*(2), 305–313. https://doi.org/10.1007/s00221-012-3015-x

Dobbs, M. B., & Gurnett, C. A. (2017). The 2017 ABJS Nicolas Andry Award: Advancing Personalized Medicine for Clubfoot Through Translational Research. *Clinical Orthopaedics and Related Research*, *475*(6), 1716–1725. https://doi.org/10.1007/s11999-017-5290-0

Ferraris, C., Amprimo, G., Masi, G., Vismara, L., Cremascoli, R., Sinagra, S., Pettiti, G., Mauro, A., & Priano, L. (2022). Evaluation of Arm Swing Features and Asymmetry during Gait in Parkinson's Disease Using the Azure Kinect Sensor. *Sensors*, *22*(16). https://doi.org/10.3390/s22166282

Ferraris, C., Amprimo, G., Pettiti, G., Masi, G., & Priano, L. (2022). Automatic Detector of Gait Alterations using RGB-D sensor and supervised classifiers: a preliminary study. *Proceedings - IEEE Symposium on Computers and Communications*, *2022-June*. https://doi.org/10.1109/ISCC55528.2022.9912923

Goffredo Michela and Carter, J. N. and N. M. S. (2009). 2D Markerless Gait Analysis. In P. and N. M. and H. J. Vander Sloten Jos and Verdonck (Ed.), *4th European Conference of the International Federation for Medical and Biological Engineering* (pp. 67–71). Springer Berlin Heidelberg.

Hatamzadeh, M., Busé, L., Chorin, F., Alliez, P., Favreau, J. D., & Zory, R. (2022). A kinematic-geometric model based on ankles' depth trajectory in frontal plane for gait analysis using a single RGB-D camera. *Journal of Biomechanics*, *145*. https://doi.org/10.1016/j.jbiomech.2022.111358

Hesse, N., Baumgartner, S., Gut, A., & Van Hedel, H. J. A. (2023). Concurrent Validity of a Custom Method for Markerless 3D Full-Body Motion Tracking of Children and Young Adults based on a Single RGB-D Camera. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. https://doi.org/10.1109/TNSRE.2023.3251440

Ihlen, E. A. F., Støen, R., Boswell, L., de Regnier, R. A., Fjørtoft, T., Gaebler-Spira, D., Labori, C., Loennecken, M. C., Msall, M. E., Möinichen, U. I., Peyton, C., Schreiber, M. D., Silberg, I. E., Songstad, N. T., Vågen, R. T., Øberg, G. K., & Adde, L. (2020). Machine learning of infant spontaneous movements for the early prediction of cerebral palsy: A multi-site cohort study. *Journal of Clinical Medicine*, *9*(1). https://doi.org/10.3390/jcm9010005

Kojovic, N., Natraj, S., Mohanty, S. P., Maillart, T., & Schaer, M. (2021). Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-94378-z

Lam, W. W. T., Tang, Y. M., & Fong, K. N. K. (2023). A systematic review of the applications of markerless motion capture (MMC) technology for clinical measurement in rehabilitation. In *Journal of NeuroEngineering and Rehabilitation* (Vol. 20, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12984-023-01186-9

Latorre, J., Colomer, C., Alcañiz, M., & Llorens, R. (2019). Gait analysis with the Kinect v2: Normative study with healthy individuals and comprehensive study of its sensitivity, validity, and reliability in individuals with stroke. *Journal of NeuroEngineering and Rehabilitation*, *16*(1). https://doi.org/10.1186/s12984-019-0568-y

Latorre, J., Llorens, R., Colomer, C., & Alcañiz, M. (2018). Reliability and comparison of Kinect-based methods for estimating spatiotemporal gait parameters of healthy and post-stroke individuals. *Journal of Biomechanics*, *72*, 268–273. https://doi.org/10.1016/j.jbiomech.2018.03.008

Leu, A., Ristic-Durrant, D., & Graser, A. (2011). A robust markerless vision-based human gait analysis system. *SACI 2011 - 6th IEEE International Symposium on Applied Computational Intelligence and Informatics, Proceedings*, 415–420. https://doi.org/10.1109/SACI.2011.5873039

Li, T., Chen, J., Hu, C., Ma, Y., Wu, Z., Wan, W., Huang, Y., Jia, F., Gong, C., Wan, S., & Li, L. (2018). Automatic timed up-and-go sub-task segmentation for Parkinson's disease patients using video-based activity classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(11), 2189–2199. https://doi.org/10.1109/TNSRE.2018.2875738

McGinley, J. L., Baker, R., Wolfe, R., & Morris, M. E. (2009). The reliability of three-dimensional kinematic gait measurements: A systematic review. In *Gait and Posture* (Vol. 29, Issue 3, pp. 360–369). https://doi.org/10.1016/j.gaitpost.2008.09.003

Meinecke, L., Breitbach-Faller, N., Bartz, C., Damen, R., Rau, G., & Disselhorst-Klug, C. (2006). Movement analysis in the early detection of newborns at risk for developing spasticity due to infantile cerebral palsy. *Human Movement Science*, *25*(2), 125–144. https://doi.org/10.1016/j.humov.2005.09.012

Moro, M., Pastore, V. P., Tacchino, C., Durand, P., Blanchi, I., Moretti, P., Odone, F., & Casadio, M. (2022). A markerless pipeline to analyze spontaneous movements of preterm infants. *Computer Methods and Programs in Biomedicine*, *226*. https://doi.org/10.1016/j.cmpb.2022.107119

Nguyen, M.-H., Hsiao, C.-C., Cheng, W.-H., & Huang, C.-C. (2022). Practical 3D Human Skeleton Tracking Based on Multi-View and Multi-Kinect Fusion. *Multimedia Syst.*, *28*(2), 529–552. https://doi.org/10.1007/s00530-021-00846-x

Pantzar-Castilla, E., Cereatti, A., Figari, G., Valeri, N., Paolini, G., Della Croce, U., Magnuson, A., & Riad, J. (2018). Knee joint sagittal plane movement in cerebral palsy: a comparative study of 2-dimensional markerless video and 3-dimensional gait analysis. *Acta Orthopaedica*, *89*(6), 656–661. https://doi.org/10.1080/17453674.2018.1525195

Stagni, R., Doto, T., Tomadin, A., Sansavini, A., Aceti, A., Corvaglia, L. T., & Bisi, M. C. (2023). General movements automatic assessment: Methodological issues for pose estimation. *Gait & Posture*, *106*, S195–S196. https://doi.org/https://doi.org/10.1016/j.gaitpost.2023.07.236

Stricker, M., Hinde, D., Rolland, A., Salzman, N., Watson, A., & Almonroeder, T. G. (2021). Quantifying step length using two-dimensional video in individuals with Parkinson’s disease. *Physiotherapy Theory and Practice*, *37*(1), 252–255. https://doi.org/10.1080/09593985.2019.1594472

Surer, E., Cereatti, A., Grosso, E., & Croce, U. Della. (2011). A markerless estimation of the ankle-foot complex 2D kinematics during stance. *Gait and Posture*, *33*(4), 532–537. https://doi.org/10.1016/j.gaitpost.2011.01.003

Wade, L., Needham, L., Evans, M., McGuigan, P., Colyer, S., Cosker, D., & Bilzon, J. (2023). Examination of 2D frontal and sagittal markerless motion capture: Implications for markerless applications. *PLoS ONE*, *18*(11 NOVEMBER). https://doi.org/10.1371/journal.pone.0293917

List of publications

Peer-review paper

D. Balta *et al*., "A Model-Based Markerless Protocol for Clinical Gait Analysis Based on a Single RGB-Depth Camera: Concurrent Validation on Patients With Cerebral Palsy," in *IEEE Access*, vol. 11, pp. 144377-144393, 2023, doi: 10.1109/ACCESS.2023.3340622.

Balta, D.; Kuo, H.; Wang, J.; Porco, I.G.; Morozova, O.; Schladen, M.M.; Cereatti, A.; Lum, P.S.; Della Croce, U. Characterization of Infants' General Movements Using a Commercial RGB-Depth Sensor and a Deep Neural Network Tracking Processing Tool: An Exploratory Study. Sensors 2022, 22, 7426. https://doi.org/10.3390/s22197426

Ashley Polhemus, Laura Delgado Ortiz, Gavin Brittain, Nikolaos Chynkiamis, Francesca Salis, Heiko Gaßner, Michaela Gross, Cameron Kirk, Rachele Rossanigo, Kristin Taraldsen, **Diletta Balta**, Sofie Breuls, Sara Buttery, Gabriela Cardenas, Christoph Endress, Julia Gugenhan, Alison Keogh, Felix Kluge, Sarah Koch, M. Encarna Micó-Amigo, Corinna Nerz, Chloé Sieber, Parris Williams, Ronny Bergquist, Magda Bosch de Basea, Ellen Buckley, Clint Hansen, A. Stefanie Mikolaizak, Lars Schwickert, Kirsty Scott, Sabine Stallforth, Janet van Uem, Beatrix Vereijken, Andrea Cereatti, Heleen Demeyer, Nicholas Hopkinson, Walter Maetzler, Thierry Troosters, Ioannis Vogiatzis, Alison Yarnall, Clemens Becker, Judith Garcia-Aymerich, Letizia Leocani, Claudia Mazzà, Lynn Rochester, Basil Sharrack, Anja Frei, Milo Puhan & Mobilise-D. Walking on common ground: a cross-disciplinary scoping review on the clinical utility of digital mobility outcomes. npj Digit. Med. 4, 149 (2021). https://doi.org/10.1038/s41746-021-00513-5

Conference Proceedings published on International Journals

D. Balta, H. Kuo, D.Tran, M. Schladen, P. Lum, U. Della Croce (2021) Recording infants' motion with a single 3D camera and a markerless tracking algorithm: evaluation of an occlusion recovery method. In proceedings of XXI Annual Congress of the Italian Society for analysis of movement in the clinic (SIAMOC 2021).

D. Balta et al., "Characterization of infants' general movements based on a single RGB-Depth camera: a feasibility study," Gait Posture, vol. 97, pp. 32–33, 2022, doi: https://doi.org/10.1016/j.gaitpost.2022.09.055. (SIAMOC 2022)

D. Balta et al., "Infant upper body 3D kinematics estimated using a commercial RGB-D sensor and a deep neural network tracking processing tool," 2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA), 2022, pp. 1-6, doi: 10.1109/MeMeA54994.2022.9856585.

D. Balta et al., "Estimating infant upper extremities motion with an RGB-D camera and markerless deep neural network tracking: A validation study," 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2022, pp. 2548-2551, doi: 10.1109/EMBC48229.2022.9871928.

D. Balta, E. Pantzar-Castilla, J. Riad, M. Salvi, F. Molinari, G. Paolini, U. Della Croce, A. Cereatti (2021) Validation of a 2D RGB-depth method for gait analysis in children with cerebral palsy. In proceedings of XXI Annual Congress of the Italian Society for analysis of movement in the clinic (SIAMOC 2021).

D. Balta, H. Kuo, D.Tran, M. Schladen, P. Lum, U. Della Croce (2021) Recording infants' motion with a single 3D camera and a markerless tracking algorithm: evaluation of an occlusion recovery method. In proceedings of XXI Annual Congress of the Italian Society for analysis of movement in the clinic (SIAMOC 2021).

M. Caruso, S. Bertuletti, **D. Balta**, A. Zedda, E. Gusai, S. Spanu, A. Pibiri, M. Monticone, D. Pani, A. Cereatti (2021) Real-time kinematics estimation with a scalable IMU body-sensor network in tele-rehabilitation. In proceedings of XXI Annual Congress of the Italian Society for analysis of movement in the clinic (SIAMOC 2021). *(Ownership 50%)*

Balta, D.; Salvi, M.; Molinari, F.; Figari, G.; Paolini, G.; Croce, U. D.; Cereatti, A. (2020) A two-dimensional clinical gait analysis protocol based on markerless recordings from a single RGB-Depth camera. In: 15th IEEE Int2.ernational Symposium on Medical Measurements and Applications, MeMeA 2020, ita, 2020, pp. 1-6. ISBN: 978-1-7281-5386-5

Balta, D.; Salvi, M.; Molinari, F.; Figari, G.; Paolini, G.; Croce, U. D.; Cereatti, A. (2020) A markerless gait analysis protocol based on a single RGB-Depth camera: sensitivity to background changes. In proceedings of Gruppo Nazionale di Bioingegneria, GNB 2020D. Balta et al., "Characterization of infants' general movements based on a single RGB-Depth camera: a feasibility study," Gait Posture, vol. 97, pp. 32–33, 2022, doi: https://doi.org/10.1016/j.gaitpost.2022.09.055. (SIAMOC 2022)