TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Peeter Maran, Anton Osvald Kuusk, Karl Mihkel Seenmaa 206688IAIB, 206159IAIB, 193937IAIB

MARKERLESS COMPUTER VISION BASED FRAMEWORK FOR CEREBRAL PALSY GAIT ANALYSIS FOR HAAPSALU NEUROLOGICAL REHABILITATION CENTER

Bachelor's Thesis

Supervisor: Gert Kanter
PhD

Co-supervisor: Elli Valla
MSc

Tallinn 2023

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Peeter Maran, Anton Osvald Kuusk, Karl Mihkel Seenmaa 206688IAIB, 206159IAIB, 193937IAIB

MASINNÄGEMISEL BASEERUV TSEREBRAALPARALÜÜSI KÕNNIANALÜÜSI MARKERITETA SÜSTEEM HAAPSALU NEUROLOOGILISE REHABILITATSIOONIKESKUSE JAOKS

Bakalaureusetöö

Juhendaja: Gert Kanter
PhD

Kaasjuhendaja: Elli Valla
MSc

Tallinn 2023

Author's Declaration of Originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Peeter Maran, Anton Osvald Kuusk, Karl Mihkel Seenmaa

22.05.2023

Abstract

The primary objective of this project is to conduct a proof-of-concept study on an alternative solution to marker-based gait analysis, which is currently employed at Haapsalu Neurological Rehabilitation Centre (HNRC) for cerebral palsy gait analysis. By utilizing videos from two cameras situated in the lab, a model of the patient's movement will be created, enabling the measurement of various gait and kinematic parameters.

For the purpose of this thesis, two RGB video cameras are utilised to create a 3D model of the subject's walking. Open-source pose estimation frameworks are utilised to obtain 2D models from both cameras, which are later combined into a single 3D model.

Using Automated Machine Learning (AutoML) techniques, gait cycles are detected from this model. The trained neural networks demonstrated good results, achieving an accuracy of 94% in detecting gait cycles. This level of accuracy renders the system promising for gait cycle detection and analysis.

For larger angles, such as knee and hip flexion and extension, the system produced very strong results, with Pearson correlations as high as 0.969 and Spearman's rank correlation coefficient as high as 0.945. However, for angles with smaller ranges, the results were much poorer, with Pearson correlations as low as 0.222 and Spearman's rank correlation coefficient as low as 0.218.

The result of this thesis is to serve as a proof-of-concept study demonstrating the feasibility and accuracy of analysing patients' movements using only two cameras. By establishing the effectiveness of this approach, this study seeks to lay the groundwork for future research and development in the field of automated gait analysis.

The thesis is written in English and is thirty-six pages long, including six chapters, nine figures and seven tables.

Annotatsioon

Masinnägemisel baseeruv tserebraalparalüüsi kõnnianalüüsi markeriteta süsteem Haapsalu Neuroloogilise Rehabilitatsioonikeskuse jaoks

Selle projekti peamine eesmärk on teostada alternatiivse lahenduse kontseptsiooni tõestuse uuring markeri põhise sammu analüüsi süsteemi vastu, mis on hetkel kasutuses Haapsalu Neurolooglise Rehabilitatsioonikeskuses (HNRK) tserebraalparalüüsi analüüsimiseks. Kasutades ära HNRK kõnnilabori kahte kaamerat, luuakse patsiendi liikumisest mudel, mis võimaldab mõõta erinevaid kõnni- ja kinemaatilisi parameetreid.

Lõputöö eesmärkide täitmiseks kasutati kahte RGB kaamerat, mille abil luuakse 3D mudel patsiendi kõnnakust. Vabavaralisi poosituvastuse raamistike abiga saadakse mõlemast kaamerast 2D mudel patisendist, mis hiljem kombineeritakse kokku üheks 3D mudeliks.

Kasutades automatiseeritud masinõpet (AutoML) tuvastatakse kõnnitsüklid. Treenitud närvivõrgud demonstreerisid häid tulemusi, saavutades 94% täpsuse kõnnitsüklite tuvastamisel. Selline tulemus võimaldab kõnnitsüklite tuvastamise ja analüüsi.

Suuremate nurkade, näiteks põlve- ja puusafleksiooni/-ekstensiooni korral, andis süsteem väga tugevaid tulemusi, mille Pearsoni korrelatsioonikordajad ulatusid kuni 0,969 ja Spearmani järjestuskorrelatsioonikordajad kuni 0,945ni. Kuid väiksemate vahemikega nurkade puhul olid tulemused palju kehvemad, Pearsoni korrelatsioonikordajad said kõige väiksemaks tulemuseks 0,222 ja Spearmani järjestuskorrelatsioonikordajad 0,218.

Selle lõputöö tulemusena loodi *proof-of-concept*, mis näitab, et on võimalik analüüsida patsientide liikumist ainult kahe kaamera abil. See paneb aluse tulevikus järgmiste veel paermate ilma markeriteta süsteemide arendamiseks.

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti kolmkümne kuuel leheküljel, kuus peatükki, üheksa joonist, seitse tabelit.

List of Abbreviations and Terms

2D	Two-Dimensional
3D	Three-Dimensional
AutoML	Automated Machine Learning - frameworks that provides methods and processes to make machine learning available for non-machine learning experts
CP	Cerebral Palsy
DNN	Deep Neural Network
FPS	Frames Per Second
GRF	Ground Reaction Force
HNRC	Haapsalu Neurological Rehabilitation Centre
IMU	Inertial Measurement Unit
RGB	Red Green Blue
RMSE	Root-Mean-Square Deviation

Table of Contents

List of Figures

List of Tables

1. Introduction

Cerebral palsy (CP) is a group of disorders that affect a person's ability to move, maintain balance and posture [1]. It is the most common motor disability in childhood. Based on a study conducted in the United States, CP affects approximately 2.9 in every 1000 8-year-old children. [2]

In recent years, motion capture systems have become increasingly popular for analysing human movement in various fields, such as sports, healthcare, and entertainment [3]. Traditionally, marker-based motion capture systems such as Vicon [4] have been widely used due to their high accuracy and precision. However, these systems are often limited by the requirement of markers placed on the subject. The process of marker placement is time-consuming and can limit the range of motion that can be captured in a gait lab.

Cerebral palsy gait analysis using the traditional marker based method is also done at Haapsalu Neurological Rehabilitation Center (HNRC). The current solution at HNRC has a downside - it is extremely time-consuming. The process of placing markers on the patient and recording data while the patient is moving can take up to 1.5 to 2.5 hours [5]. Precision is critical for marker placement, so it is done by highly trained physicians. Once data is collected, it takes up to 3 to 16 hours to analyse the patient's kinematic, kinetic, temporal, and spatial parameters. This process also requires a specialist and involves little automation. Clinicians manually select gait cycles from the generated 3D model and assess each aspect of the patient's gait parameters separately, which is a time-consuming process. Patients, who are predominantly children, find the process draining as it requires them to remain still during marker placement. Prior to marker placement, patients' body lengths are extensively measured using different devices, which may be frightening for younger patients. The current walking procedure involves patients walking while eight infrared cameras track the markers, which can cause the room to heat up during longer sessions, causing further discomfort to patients.

1.1 Problem Statement

The traditional marker-based gait analysis process can be quite time-consuming and tedious, leading to long queues and wait times for the clinician's visit. Unfortunately, this can make the overall experience to be much more uncomfortable for the patients. Additionally, the process can be costly in terms of hours of clinicians' work required.

This thesis aims to investigate the potential of a markerless motion capture system for gait analysis, as an alternative to the current marker-based approach. Specifically, our goal is to build our own markerless system with the two cameras at HNRC and to compare its accuracy with the current Vicon [4] infrared camera system. Furthermore, our objective is to assess the potential advantages of our markerless system, including increased efficiency, time savings, and ease of use for hospital staff. This will involve automating the data collection and analysis process, allowing for faster and more efficient diagnosis and treatment of patients with cerebral palsy. Therefore, the goals of this thesis are to:

- 3D model creation
- gait-cycle phase detection
- calculation of different gait and kinematic variables for gait analysis

The 3D model will be created by combining the data from 2D pose estimation frameworks such as MediaPipe (See Section 3.2.1). This markerless solution would be significantly less time-consuming and less inconvenient for patients. In addition, utilizing pre-existing pose detection frameworks makes the system significantly easier to construct and set up, thus reducing costs and potentially increasing accessibility for hospitals.

The gait cycle detection is done by using automated machine learning (see Section 3.4.1). This will mostly help the physicians, as selecting the gait cycle from walking trials is often still done by hand. Furthermore, this will set the ground for automatically calculating different gait and kinematic variables (see Sections 3.5.1 and 3.5.2) from the created 3D model for gait analysis (Section 3.5).

By accomplishing these steps, the authors of this thesis aim to reduce the time consumption and needed hours clinicians' work of cerebral palsy gait analysis and to set the grounds for future developments and research for markerless gait laboratory systems.

2. Background

2.1 Cerebral Palsy

CP is a neurological disorder that affects movement and muscle coordination. It is caused by damage to the brain before, during, or shortly after birth, and can result in a wide range of physical and cognitive impairments[1]. According to the Autism and Developmental Disabilities Monitoring Network, CP affects approximately 2.9 in every 1000 8-year-old children in the United States [2]. Symptoms of CP can vary widely, ranging from mild to severe, and can include muscle stiffness, spasticity, and difficulty with balance and coordination.

Gait analysis is a critical aspect of clinical assessments for individuals with CP. The ability to walk and move is often affected in individuals with CP, and gait analysis can help identify the underlying causes of gait abnormalities and track the effectiveness of interventions. For example, gait analysis is used to measure walking speed, step length, stride duration, and joint angles during walking. This information can be used to develop individualized treatment plans to improve gait and mobility in individuals with CP.[6]

2.2 Gait Analysis

Gait analysis is the studies of human walking and other forms of locomotion, such as running and jumping. It involves measuring and analysing the body's movement patterns during these activities. Gait analysis is used in various fields, including clinical medicine, rehabilitation, sports science, and biomechanics.[7] [8]

There are several methods for gait analysis, including visual observation, pressure-sensitive walkways, force plates, and motion capture systems [9]. The most commonly used method is motion capture, which involves recording the movement of the body using cameras and markers. Capture data can be used to analyse joint angles, movement patterns, and other kinematic variables[7].

Gait analysis can provide valuable information for assessing and treating various conditions affecting walking and mobility. For example, gait analysis can be used to identify the causes of abnormal gait patterns in individuals with neurological disorders like cerebral palsy, stroke, and multiple sclerosis. It can also be used to evaluate the effectiveness of

interventions like orthotics, prosthetics, and physical therapy.[10]

Traditionally, gait analysis has been performed using marker-based motion capture systems like Vicon [4]. However, these systems have limitations in terms of setup time, invasiveness, and range of motion that can be captured. Markerless motion capture systems using RGB cameras have the potential to provide a more flexible and non-invasive solution for gait analysis. Evaluating the accuracy and precision of markerless systems compared to marker-based systems, can help identify the most effective and efficient methods for gait analysis in different populations.

Overall, gait analysis is an important tool for understanding human locomotion and developing interventions to improve mobility and quality of life for individuals with movement disorders.

2.2.1 Haapsalu Neurological Rehabilitation Center

The patients' specific symptoms are analysed in Haapsalu neurological rehabilitation center's (HNRC) clinical motion and gait analysis laboratory [5]. HNRC gait lab is equipped with Vicon motion and gait analysis systems: 8 MX T-Series infra-red cameras [11], 2 Basler pilot piA640-210gc video cameras [12] (one records patient from the side and the other from the front). HNRC also has 2 AMTI force plates at the lab. All these technologies are supported by Vicon software: Vicon Polygon [13], Vicon Nexus [14], and other such programs that are designed for data recording, processing, and reporting. The main part of the process at HNRC involves infrared markers that are attached to the patient's body and the aforementioned 8 MX T-Series infra-red cameras which are used to track the markers.

The main problem with the current solution at HNRC is that the process of analysing and setting up the patient is time-consuming. The process of placing markers on the patient and recording the data while the patient is moving around takes already up to 1.5 to 2.5 hours in total [5].

This process is time-consuming because the markers must be placed on the patient with precision. For some markers, a laser sensor is used for measurements, as the markers have to be in very precise positions. All of that is done by well-trained physicians. After the data is gathered, it is analysed. According to the physicians at HNRC, the time to analyse the patient after data recording takes up to 3 to 16 hours. During the analysis, the physicians assess the patients' kinematic, kinetic, temporal and spatial parameters. Just as marker placement, this part also requires a well-trained specialist. Currently, there is very

little automation — clinicians need to manually select gait cycles from the generated 3D model and assess each aspect of the patient's gait parameters separately. The gait cycles are analysed thoroughly and as the final result of the analysis, the clinician gathers all findings to a single document by hand.

Besides being time-consuming, the process can be very tiring for the patients, who are predominantly children. The medical procedure can be quite exhausting, as the placement of markers on their bodies requires accuracy which will be affected by patient's motion. Additionally, patients' body lengths are extensively measured prior to the placement of markers, using different measuring devices that may initially frighten younger patients. The measuring procedure may even tickle some patients, further lengthening the already tedious process. Once the patients' body lengths have been measured and the markers have been placed, it is time for the actual walking component of the process. The clinicians at HNRC have noted that the current walking procedure, which involves having the patient walk. At the same time, eight infrared cameras track the markers, can cause the room to noticeably heat up during lengthier sessions, further causing distress to the patients.

In summary, the process is time-consuming and expensive (requires many clinical hours). Additionally, considering the manual nature of the analysis, it can be prone to the subjectivity of the clinician[15].

2.3 Literature Review

Earlier studies, such as [16] and [17] have found that at least 8 cameras are necessary to track human movements with the precision comparable to the marker based system. Moro, Marchesi, Hesse, Odone, Casadio [18], on the other hand, recently conducted a comparison between markerless and marker-based systems. Testing was done with three cameras, and they managed to achieve comparable results without markers.

For the past few decades the challenge of markerless human motion tracking has needed at least 8 cameras for precise results. However, the developments in machine learning technologies and results from Moro, Marchesi, Hesse, Odone, Casadio [18] hint at a very real possibility of using as few as three cameras to precisely track joint and limb movements.

Gait analysis systems, such as the marker-based one at HNRC [15], have been used for a long time. However, markerless systems with few cameras have just recently been tested. Ma, Mithraratne, Wilson, Wang, Ma, Zhang [19] evaluated Kinect, a camera developed by Microsoft, for gait analysis. They conclude that the promising results "support the

Kinect's clinical capacity in becoming a potential gait analysis tool" admitting that further improvements might be necessary. Additionally, Nieto-Hidalgo, Caballero-Gil, Ferrández-Pastor, Valdivieso-Sarabia, Mora-Pascual, García-Chamizo [20] used only a smartphone with cloud computing assistance to extract features claiming 95% accuracy for the side and 80% for frontal view.

In 2015, Castelli Andrea, Paolini Gabriele, Cereatti Andrea, and Corce Ugo Della used a single video camera, meaning 2D data, to perform lower body sagittal plane kinematic analysis [21]. They showed strong correlations as high as 0.99, which are very promising results for this thesis. However, since this thesis aims to perform kinematic analysis in all anatomic planes, not just the sagittal plane, at least two cameras are needed for 3D data.

3. Methodology

For this thesis, two RGB video cameras (further discussed in Section 3.1.3) are utilised to create a 3D model of the subject's walking. Open-source pose estimation frameworks (discussed in Section 3.2.1 and 3.2) are utilised to obtain 2D models from both cameras, which are later combined into a single 3D model (as described in Section 3.3). Using machine learning techniques (discussed in Section 3.4.1), gait cycles (described in Section 3.4) are detected from this model. Gait analysis can then be conducted using the created model and the detected gait cycles (further discussed in Section 3.5).

3.1 Sensors

3.1.1 Overview

Currently, the system used at HNRC (See section 2.2.1) consists of 8 MX T-Series infrared cameras [11], 2 Basler pilot piA640-210gc video cameras (one records patient from the side and other from the front)[12] working on 50 Hz. This thesis aims to use the two video cameras to achieve the same outcome as the 8 infrared cameras. In addition to video cameras (also known as RGB cameras) and infrared cameras, depth cameras, which measure the depth of an image, and inertial sensors are two other types of sensors sometimes used for motion capture in gait laboratories.

3.1.2 Infrared Cameras

Infrared cameras can detect and measure infrared energy not visible to the naked eye [22]. In motion tracking applications, infrared cameras are used in conjunction with reflective markers that reflect infrared light, as their name suggests. The infrared cameras emit infrared light, which is then reflected by the markers and detected by the cameras. It is possible to effectively track these reflective markers in 3D space by using multiple infrared cameras [23]. Therefore, by placing markers on all major body parts, it is possible to track the complete movement of a person using infrared cameras, which is why infrared cameras are often used in gait laboratories.

3.1.3 RGB Cameras

RGB cameras (also known as "regular" video cameras) are utilised in motion tracking for gait analysis, but they do not carry out the tracking themselves. Instead, complex computer algorithms are built on top of the camera feed to perform the actual motion tracking. Using RGB cameras in gait analysis is relatively new, but there has been quite some work done on this, for example, the work by Martin Sandau and Henrik Koblauch in 2014, who used 8 RGB cameras and were able to provide a reliable 3D gait kinematics in the sagittal and frontal plane (visualized in Appendix 2) [24]. This thesis aims to utilise different pose estimation frameworks, which are discussed further in Sections 3.2 and 3.2.1, to capture 2D pose from two RGB cameras which will later be combined into a 3D model (as described in Section 3.3).

3.1.4 Depth Cameras

Depth cameras use different sensing technologies to measure the distance, or depth, of points in a scene from the camera. This measurement can be calculated using various techniques, such as time-of-flight or stereo vision [25][26]. In gait analysis, depth cameras can provide real-time distance measurements of moving objects and enable body volume reconstruction, which allows for the creation of a model of the patient's movement. For example, Auvinet, Edouard and Multon, Franck, and Meunier, Jean demonstrated this in 2011 [27], and Auvinet, Edouard and Meunier, Jean and Multon, Franck in 2012 [28].

3.1.5 Inertial Measurement Unit

An IMU (Inertial Measurement Unit) is a sensor that is typically attached to a patient's body or clothing and used to measure their movements using accelerometers and gyroscopes. This provides information about both the linear and angular motion of the body segments, which can be used for gait analysis. In addition to improving the accuracy of motion capture systems, IMUs can also be used as standalone devices for gait analysis, allowing for the measurement of gait parameters such as step length, stride length, cadence, and walking speed. Furthermore, IMUs can be used in combination with other sensors to provide a more comprehensive analysis of gait and movement.[29]

3.1.6 Ground Reaction Force Plate

A GRF (Ground Reaction Force) plate, also known as a pressure plate, is a type of sensor system used in gait analysis to measure the force exerted by the ground on a person's feet

during walking, running, or other movements. The plate is typically placed on the floor and contains embedded sensors that detect changes in pressure as a person walks across it. These sensors can measure the magnitude and direction of the forces exerted by the feet on the ground, providing valuable information about gait and movement patterns [30].

HRNC uses 2 AMTI GRF plates to detect and measure the moment, power, and force of a patient's walk during the gait cycle. This provides further valuable data and can be used as extra parameters to define the gait cycle phase.[5]

3.1.7 Sensor Choice and Conclusion

There are numerous methods available for motion capture in gait analysis. For the purpose of this thesis, we utilised the two RGB cameras that were already present at HNRC, as mentioned previously. This approach allows for easy comparison with the Vicon infrared camera system since the cameras are already in place. Moreover, HNRC had already collected a considerable amount of data using these cameras, providing us with ample training data if necessary.

3.2 Pose Estimation

Pose estimation is the process of determining the position and orientation of an object or a human body in a given space based on sensor data such as images or video. Specifically, human pose estimation involves identifying the locations of various key points on the human body, such as the head, torso, arms, and legs.

In computer vision, pose estimation typically involves using machine learning algorithms such as deep neural networks to analyse images or video frames and predict the locations of the key points. This can be done using various techniques such as heatmaps, regression, and graphical models. [31]

Pose estimation has a wide range of applications, including motion capture for animation and gaming, surveillance and security, human-computer interaction, and medical diagnosis and treatment. For example, in sports science, pose estimation can be used to analyse and improve athletic performance, while in healthcare, it can be used to diagnose and monitor movement disorders such as Parkinson's disease. [3]

Overall, pose estimation is an important and rapidly evolving field with numerous applications in various industries and domains.

3.2.1 Frameworks Overview

MediaPipe

MediaPipe is an open-source framework developed by Google for building real-time multimedia processing pipelines. One of its key features is its ability to perform accurate and efficient pose estimation on human bodies using deep learning models. MediaPipe's pose estimation algorithm works by first detecting key body landmarks, such as the shoulders, elbows, wrists, hips, knees, and ankles, using a convolutional neural network (CNN). It then uses these landmarks to estimate the orientation and position of the body.[32]

It is important to note that MediaPipe detects the 3D coordinates of various landmarks, with the z-coordinate (depth) provided relative to the subject's hips. However, our testing discovered that depth measurements were highly unreliable. As a result, the system presented in this thesis utilised only the x and y coordinates (2D data) obtained from MediaPipe.[32]

AlphaPose

AlphaPose is a real-time and accurate multi-person pose estimation system, working on COCO and MPII datasets. It is based on a fully convolutional network and combines a bottom-up and top-down approach to detect human body keypoints.

AlphaPose also incorporates an online tracking algorithm to track individuals and maintain consistency across video frames. The system has achieved state-of-the-art performance on several benchmark datasets, making it a valuable tool for a wide range of applications such as human motion analysis, sports training, and surveillance.[33, 34, 35]

AlphaPose is freely available for free non-commercial use and may be redistributed under these conditions.

OpenPose

OpenPose is a real-time multi-person 2D pose detection system that uses deep neural networks to detect human body keypoints, up to 27. It is based on a bottom-up approach that first detects individual body parts and then combines them to estimate the full body pose. Pose estimation requires great computational resources.[36, 37, 38, 39]

Metrabs

Metrabs (short for "Metric Learning-Based Multi-Object Tracking") is a multi-object tracking system that uses metric learning to improve object tracking accuracy. It is based on a deep neural network architecture that can track multiple objects in real-time.

Advanced models can only be used for non-commercial purposes due to the licensing of the used training datasets. [40, 41]

Tensorflow Movenet

TensorFlow Pose Estimation is a deep learning-based system that uses convolutional neural networks (CNNs) to estimate human body keypoints in images or videos. It is built on the TensorFlow framework, which is a popular open-source platform for building and training machine learning models.[42]

HRNET

HRNet (short for "High-Resolution Network") is a deep neural network architecture designed for computer vision tasks such as image classification, object detection, and semantic segmentation. It is characterized by its ability to maintain high-resolution feature maps throughout the network, improving visual recognition tasks' accuracy.

HRNet can also be used for human pose estimation, which involves predicting various body keypoint locations in an image or video. HRNet-based pose estimation systems typically consist of a multi-stage architecture that includes a feature extraction module, a multi-resolution fusion module, and a keypoint regression module. [43, 44, 45]

Detectron2

Detectron2 is an open-source computer vision library developed by Facebook AI Research. It is built on top of PyTorch and is designed to simplify the process of building and deploying object detection, instance segmentation, and keypoint detection models. Detectron2 is released under the Apache 2.0 license [46]

3.2.2 Frameworks Comparison

Conclusions Based on Overview

Table 1 gives a visual overview of frameworks and what their capabilities are. Some are either directly not available for commercial use while some are free to use, but the more

Table 1. Framework comparison table

Framework	Single	Multiple	Nvidia GPU	License
MediaPipe	+	-	-	+
Hrnet	+	+	+	+
OpenPose	+	+	+	-
AlphaPose	+	+	-	-
Metrabs	+	+	-	-
Tensorflow	+	-	-	+
Detectron2	+	+	-	+

advanced models are only for non-commercial use.

Another aspect that is irrelevant for us and can even be more problematic is multiple-person detection since there is only one patient in the lab at a time. The important thing for us are keypoints, how many specific the coverage of the body is and how accurate they are.

Most free-to-use models are quite limited on that front, having only the basic 17 keypoints leaving out the foot, which is a valuable datapoint for us. From this aspect, MediaPipe has the most comprohensive model that is available for use, containing 32 keypoints for the entire body.

Table 2. Test System Specifications

Component	Specification
CPU	Intel(R) Core(TM) i7-6820HQ CPU @ 2.70GHz
OS	Ubuntu 20.04
Memory	16 GiB
GPU	NVIDIA Quadro M1000M (CUDA compute 5.0)

Performance and Speed Test

An important aspect is the speed and performance of the framework. From the development side, it adds a lot of value the faster we can get results and work on problems. On the other hand, the faster the therapists can see results, preferably in real-time, the better their overview of potential problems. Figure 1 shows the results for speed performance tested with previously given system specifications. this test was conducted with the same video of a person moving and fed into every framework. The length of the original file was 30 seconds, with MediaPipe being the clear winner with almost realtime result of 34.2 seconds.

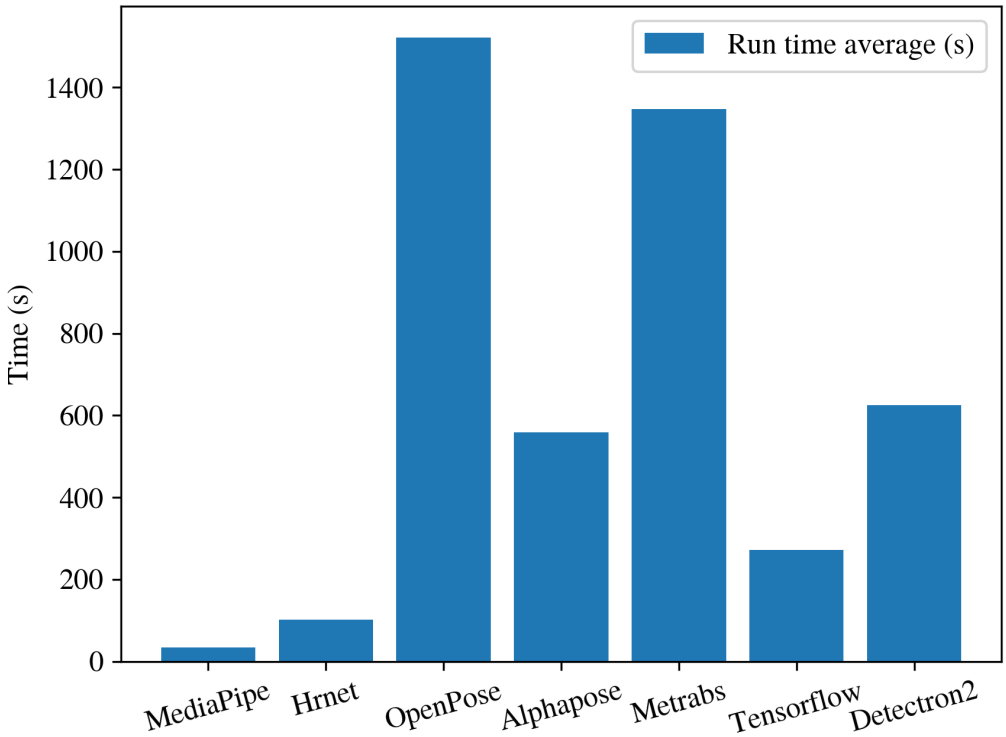


Figure 1. *Average time of processing a test footage with given framework, with 640x480 resolution and 30 FPS*

Stability Test

Testing was done on still videos to check the stability of our available frameworks, this was to check to see how consistent they are. Accuracy on a moving subject is highly dependent on picture angles and the frame rate output of the camera. But it is also important to see that the markers are moving along not replaced on every new picture. This was conducted on four of the frameworks that are suitable for our purposes. The test was done with a single cutout picture being turned into a 768 frame video. This was run through each system and the location coordinates of keypoints were stored. Table 3 shows the results of standard deviation in terms of placement of the keypoint in the video. This is shown in pixels, but an approximate real world measurement scale is 2 pixels for 1 centimeter.

Table 3. Standard deviation for different models

Model	**x (pixels)**	**y (pixels)**
MediaPipe	0.2459	0.4226
Detectron2	0.5519	0.4298
TensorFlow Movenet	2.1118	2.9386
HRNet	26.7547	28.8701

Framework Conclusion

Based on the data and test results, MediaPipe was the most suitable framework for this thesis. It has good accuracy and the most comprehensive model that is freely available. It is also the fastest working being capable of almost realtime performance. This means that when using higher framerates in cameras for more detail and smoother motion it wont slow down the process. Picture quality can be fairly low and still provide good detection results. Finally it has a simple to use API that is very easy to modify.

3.3 3D Model Creation

Our model combines two 2D models obtained from MediaPipe results mentioned in Section 3.2.1 to create a 3D model. This is achieved using OpenCV, an open-source computer vision framework [47]. The model creation process involves three main steps: camera calibration, determination of camera location and rotation, and triangulation to obtain the final 3D model.

3.3.1 Theoretical Background of Stereo Vision

At HNRC currently, two cameras are capturing the scene, one from the front of the walking trajectory and the other from the side. Using MediaPipe, 2D points are obtained on both images that correspond to human body landmark locations, providing correspondences between the two camera images. After determining the distance and rotation between the two cameras, the corresponding 3D point for each pair of 2D points on the images can be computed[48]. As mentioned earlier, this entire process consists of three parts: camera calibration, determination of camera location and rotation, and triangulation. Camera calibration and determination of the cameras' location and rotation is only done once during the initial setup, triangulation is done for each captured video frame.

Calibration

Camera calibration is the process of estimating the intrinsic parameters of a camera, which are necessary for accurately interpreting the 3D world from the 2D images captured by the camera[49]. More precisely, for the purpose of this thesis, it is the process of obtaining the cameras' intrinsic matrix seen in equation 3.1, that gives the camera's focal lengths f_x and f_y, camera's skew factor s (value by which camera's sensor is off from the center) that in modern cameras is almost always 0 and the principal point $(c_x; c_y)$ which is the point where the optical axis (imaginary line that passes through the center of the lens and is perpendicular to the sensor) intersects the image plane. From calibration, a distortion

model of the camera's lens is also obtained, which is used to correct for the distortions. [48]

$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3.1}$$

Calibration involves comparing points on the image to corresponding points in the real world. For this, in principle, any clearly distinguishable object could be used as a calibration object, yet the practical choice is a regular pattern such as a chessboard. Many images are used as described in Figure 2, which was originally created by Bradski, Kaehler in 2008 and was taken from their book [48]. In our testing, the best results were obtained with around 30 images. The complete theory behind camera calibration and, in particular, distortion removal is a complex and extensive process that is not covered in depth due to calibration not being the main focus of this thesis. However, a detailed process description can be found in the aforementioned [48]. It should be noted that the calibration process is only performed once during the system's initial setup.

By knowing the size of the chessboard, we can also find the chessboards position relative to the camera, which will be used to determine the cameras' relative positions in the next section.

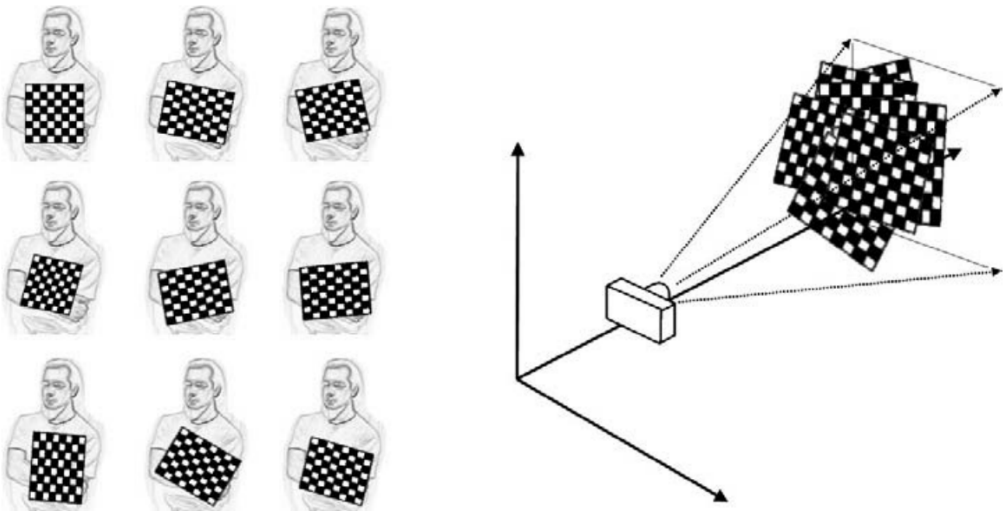


Figure 2. *Images of a chessboard being held at various orientations (left) provide enough information to completely solve for the camera's intrinsic and relative positions of the chessboard. Figure was originally created by Bradski, Kaehler in 2008 [48]*

Camera Location and Rotation

Once both cameras' intrinsic matrices and distortion coefficients are obtained, it is possible to calculate the cameras' positions (translation vector T and rotation matrix R) relative to each other as seen on Figure 3, which was also originally created by Bradski, Kaehler 2008 in their book [48]. For this, images of the calibration device (chessboard) in the same scene are needed for both cameras to get corresponding 2D image and 3D object points for both cameras. From those corresponding 3D points and 2D point pairs, one camera's location and rotation relative to the other can be calculated. For this process we found that around 5 images works well, greater number of images will cause the compute time to go up significantly without any real gain in accuracy.

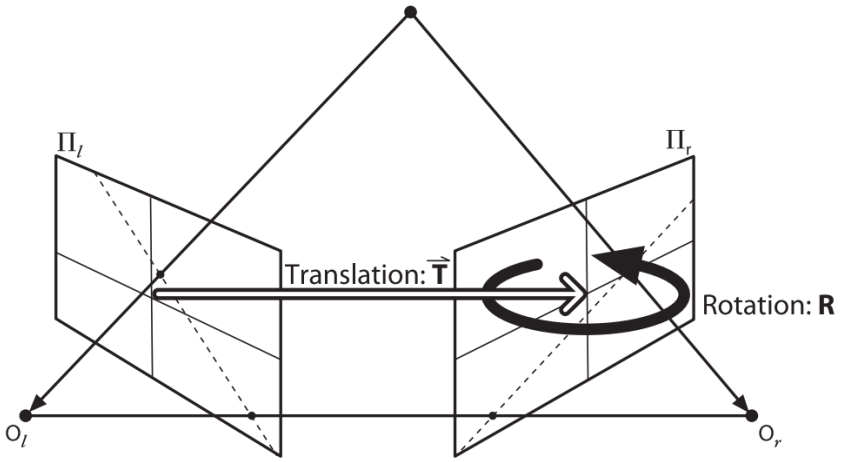


Figure 3. *Translation T and the rotation R describe the position of the second camera relative to the first in global coordinates. Figure is taken from [48]*

Lastly, the projection matrices are calculated after obtaining the second camera's pose relative to the first one. A projection matrix describes the mapping between 3D points in a scene, and their 2D image coordinates on an image plane [50], and it is necessary for triangulation.

As with camera calibration, determining the location of one camera relative to another is a fairly complex process that is very well described by Bradski, Kaehler 2008 in their book [48]. Determining one camera's position relative to the other one is also done only once during the system's initial setup.

Triangulation

Triangulation is the process of estimating the 3D location of a point in space using its projections in multiple images [51]. In stereo vision, this involves finding the 3D point from its two projections onto the two 2D camera planes. The 2D points are also processed

before to remove the effects of lens distortion.

However, in real-life scenarios, the 2D projections of points are often imprecise and do not correspond to the 3D point with absolute accuracy, making accurate triangulation quite complex[52]. The proposed system in this thesis encounters imprecision with 2D projections, primarily due to the complexity involved in accurately distilling a human body part, such as the knee or heel, down to a single point. As a result, pose estimation frameworks discussed in Section 3.2.1 may not provide identical positions for a point captured by two cameras. Fortunately, a significant amount of prior work has been conducted in this field, such as the work of Richard I. Hartley and Peter Stur in 1997 [51], Richard Szeliski and Sing Bing Kang in 2003 [53] and Richard Hartley in 2010 [52], due to which minor imprecision in 2D projections don't affect the results significantly.

In the system proposed in this thesis, triangulation is performed for every frame of every video captured of the patient's walking.

3.3.2 OpenCV

Of course, all mathematical concepts and calculations described in Section 3.3.1 are not performed manually, since there are numerous frameworks available that provide the needed functionality. For our system, we have opted for OpenCV, an open-source computer vision framework[47].

Functionality

OpenCV provides a wide range of computer vision-related functionality, from low-level image-processing functions to more advanced topics such as face detection, pedestrian detection, feature matching, and tracking [54]. Specifically, for this thesis, OpenCV's well-documented functionality for camera calibration, stereo vision, and point triangulation is particularly significant as it addresses all the necessary requirements to achieve the research objectives.

Documentation

One of the significant aspects of OpenCV is its extremely thorough documentation. In addition to a detailed documentation page [55] that provides explanations for all of its functions, there is also an excellent book by Bradski, Gary, and Kaehler, Adrian, published in 2008, that explains the core concepts of computer vision and OpenCV functions [48]. The concepts presented in this book were instrumental in the creation of the 3D model described in this thesis.

3.4 Gait Cycle

The gait cycle describes the cyclic pattern of movement that occurs while locomotion. A gait cycle starts with the heel of one foot striking the ground and ends when that same heel touches the ground again [56]. The gait cycle is usually divided into two main phases - stance and swing phase. The stance phase accounts for approximately 60 percent of one whole gait cycle and swing phase accounts for 40 percent [57].

The stance phase begins with the foot striking the ground and ends when the toes of the same foot are pushing into the ground and the ankle plantar flexes (foot point downwards of the ankle). During that phase, the foot is on the ground and is bearing the weight of the body. The stance phase consists of five sub-phases. [56]

1. **Heel strike** - Initial contact with ground. It requires the body's weight to be accepted by the leg.
2. **Foot flat** - Foot rolls forward until the entire surface of the foot is in contact with the ground.
3. **Mid-stance** - Weight of the body is directed forward so that the entire weight is being balanced over one leg.
4. **Heel-off** - Includes lifting the heel off the ground and shifting the body weight to the opposite leg.
5. **Toe-off** - Pushing toes into the ground while the ankle plantar flexes, creating forward movement.

The swing phase is the second phase of gait. During that phase, the leg is free to move forwards. This phase is the period between toe-off and new heel strike and contains 3 sub-phases. [56]

1. **Early swing** (acceleration phase) - foot is lifted from the ground. The ankle bends backwards (foot towards leg) and the knee bends to move the foot and toes from the ground. The hip bends to bring the leg forward, moving it directly under the body.
2. **Mid-swing phase** - the leg that is in the swing phase moves past the weight-bearing leg. The weight of the body is only on one grounded leg.
3. **Late swing** - The foot is moved onwards to a position in front of the body and the knee is extended straight, making the body prepare for heel strike with the ground.

Walking requires healthy functioning and cooperation of multiple body systems. It includes the musculoskeletal, nervous, cardiovascular and respiratory systems. These systems provide humans with stability, mobility and control of the body [56]. Thus, gait analysis

is essential for evaluating and diagnosing patients with cerebral palsy, which relies on accurate gait cycle detection. Amen, ElGebeily, El-Mikkawy, Yousry, El-Sobky in [58] and Park, Chung, Lee, Choi, Cho, Yoo, Lee in [59] both used gait cycle parameters to assess the effectiveness of single-event multilevel surgery for patients with cerebral palsy.

3.4.1 Detection

For the purpose of this thesis, the term 'detection' refers to the process of determining whether, during a certain frame, the patient has both feet on the ground, or if the left or right foot is moving while the other foot remains on the ground.

There are many different methods and sensors to detect gait cycles. One of the most common ways is kinematic analysis, which involves placing markers on the body and tracking the movement of each marker to determine different phases of the gait cycle. On the other hand, kinetic analysis is used to measure the forces acting on ground to identify different cycles. Another common way to detect gait cycles is by using accelerometers, to measure rotation and accelerations. [60]

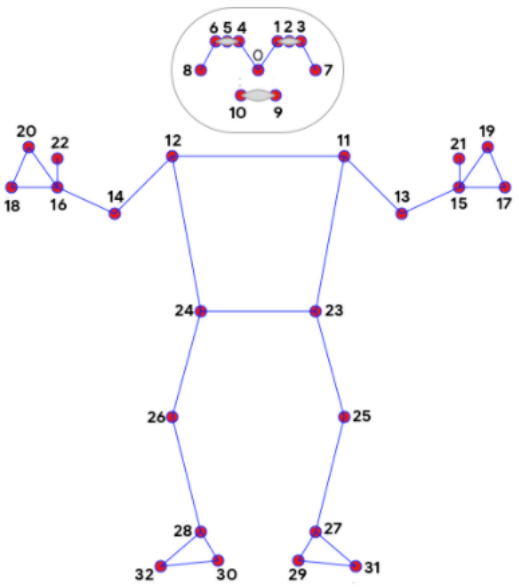
The methods used at HNRC are kinematic and kinetic analysis. By placing markers on the patients' body and tracking them with Vicon Motion Capture Systems [4] they can perform kinematic analysis. They also use pressure plates to measure the forces acting on the ground, that way they also implement kinetic analysis.

To classify gait cycle phases based on estimated poses (see section 3.2), we chose to use a machine learning approach. Specifically, we adopted an Automated Machine Learning (AutoML) [61] approach for this project, as our primary aim was not to determine the optimal parameters and architecture for the neural network. Instead, we focused on leveraging AutoML techniques to develop an accurate and efficient predictive model. While there are multiple AutoML systems to choose from, such as AutoWEKA [62] and Auto-sklearn [63], we selected Auto-PyTorch [64]. This is because it is the only AutoML framework that optimizes both the neural network architecture and the training hyperparameters, enabling fully automated deep learning.

Since the output of the pose estimation process (see section 3.2) is a set of body landmarks, we implemented a tabular classification approach. This allowed us to efficiently process the landmark data and classify the corresponding gait cycle phases.

3.4.2 Training Data

In this study, the training data is utilised for gait cycle phase classification and was limited to a sub-set of lower-body landmarks. Specifically, only a maximum of 10 points were selected, including the hip, knee, ankle, heel, and foot index for both the right and left sides (designated as points 23-32 on Figure 4). While this selection of landmarks may seem sparse, it was found to be sufficient for accurate classification of gait cycle phases based on pose estimation.



0. nose
1. left_eye_inner
2. left_eye
3. left_eye_outer
4. right_eye_inner
5. right_eye
6. right_eye_outer
7. left_ear
8. right_ear
9. mouth_left
10. mouth_right
11. left_shoulder
12. right_shoulder
13. left_elbow
14. right_elbow
15. left_wrist
16. right_wrist
17. left_pinky
18. right_pinky
19. left_index
20. right_index
21. left_thumb
22. right_thumb
23. left_hip
24. right_hip
25. left_knee
26. right_knee
27. left_ankle
28. right_ankle
29. left_heel
30. right_heel
31. left_foot_index
32. right_foot_index

Figure 4. *MediaPipe [32] body landmarks*

In addition to selecting a sub-set of landmarks, the training data was also subjected to modification and filtering. Since raw inputs can often contain a significant amount of noise, a median filter was implemented in order to reduce the noisiness of the data. This helped to remove any outliers or irregularities in the data that could potentially lead to inaccuracies in classification. Overall, the combination of selecting a sub-set of landmarks and filtering the data proved to be an effective approach for achieving accurate and consistent gait cycle phase classification results.

In order to extract more meaningful features from the pose data, we calculated the coordinate deltas using the Euclidean distance (see equation 3.2). Since the leg is in motion only during the swing phase and stays almost stationary during the stance phase, the coordinate deltas provide a measure of the displacement of the leg between frames. Specifically, we calculated the distance between the coordinates of each landmark point in the current frame and those of the same point in the previous frame. This allowed us to capture the direction and magnitude of movement for each landmark point over time.

$$d\left(p,q\right)=\sqrt{\left(q_x-p_x\right)^2+\left(q_y-p_y\right)^2+\left(q_z-p_z\right)^2} \tag{3.2}$$

We then used these coordinate deltas as input features for the gait cycle classification model. By tracking the movement of each landmark point over time, we were able to better differentiate between the various gait phases.

After estimating the poses in 3D, the next step was to annotate the data frame by frame. The purpose was to classify each frame into one of four different classes. The first class was used to describe the frames where both legs were in contact with the ground. The second and third classes were used to indicate whether the left or right leg was in the stance phase or swing phase, respectively. Finally, the fourth class was introduced to identify the frames where the estimated pose was inaccurate. This step was necessary to filter out inaccurate data and ensure the quality of the gait analysis.

During the annotation process, it was also decided to use a specific labelling system to represent a correct gait cycle. This labelling system used a combination of zeros, ones, and twos, where each number represented a specific phase of the gait cycle. A correct gait cycle would then have a pattern of 01020, with any number of zeros, ones, or twos in a row and with the number two and one possibly appearing in different orders. The labelling process can also be seen in Figure 5. This labelling system provided a useful reference point for later analysis and helped to ensure consistency across different data sets.

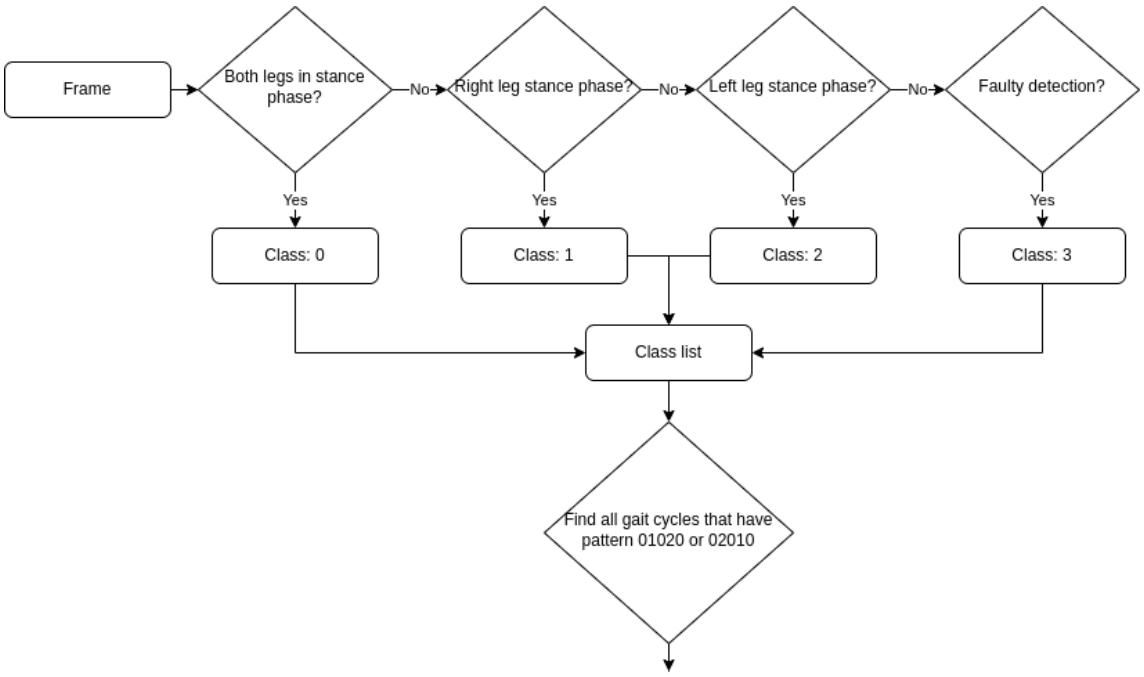


Figure 5. *Process of labelling frames.*

Overall, the annotation process was a crucial step in the development of an accurate and reliable gait analysis system. By carefully classifying each frame based on the estimated pose, we were able to filter out inaccurate data and ensure that our analysis was based on high-quality data.

3.5 Gait Analysis

Gait analysis is the systematic study of human walking that utilises instruments to measure body movements, mechanics, and muscle activity. This technique is particularly valuable for individuals with limited mobility, as it can aid in precise diagnoses and the development of personalized treatment plans. Notably, gait analysis is especially crucial for cerebral palsy patients, for whom it is an essential process.[65]

At HNRC [15], gait analysis is primarily conducted by analysing the kinematic variables of patients (discussed further in Section 3.5.1) and general gait parameters (discussed further in Section 3.5.2). This involves comparing numerical values and analysing various graphs that illustrate changes in different joint angles throughout a gait cycle. A treatment plan is developed based on the comparison of the left and right sides and the deviation of the data from the control group data. It is worth noting that in many cases, the exact values of kinematic variables are not crucial. Rather, the shape of the graph and its overall similarity to that of the control group are more important, as small fluctuations are normal and expected, according to physicians at HNRC [15]. These fluctuations are often caused by markers moving due to skin movement (in marker systems) and other minor factors.

In this thesis, we aim to evaluate the validity of markerless detection by comparing its accuracy with that of the Vicon marker motion capture system [4] currently utilised in the HNRC gait laboratory [5]. Our main focus will be on the accuracy of the measured kinematic variables.

3.5.1 Kinematic Variables

Kinematic variables (also known as gait kinematics) in gait analysis are measurements used to describe the motion of the body during the gait cycle (aforementioned in Section 3.4). The kinematic variables are divided into three anatomical planes corresponding to the human body and its movement. These are the sagittal, frontal (also known as coronal) and transverse plane[65]:

1. A sagittal plane divides the body into right and left portions.
2. The frontal plane divides the body part into front and back portions.

3. The transverse plane divides the body into upper and lower portions.

These anatomical planes are visualized in Appendix 2, which was taken from Michael W. Whittle's book on gait analysis, originally published in 1990 [65]. Most body joints can only move in one or two of those planes. The possible movements (also described on Figure 6) are [65]:

1. Flexion/extension takes place in the sagittal plane; in the ankle these movements are called dorsiflexion and plantar flexion, respectively.
2. Abduction/adduction takes place in the frontal plane.
3. Internal and external rotation occur in the transverse plane (also called medial and lateral rotation).

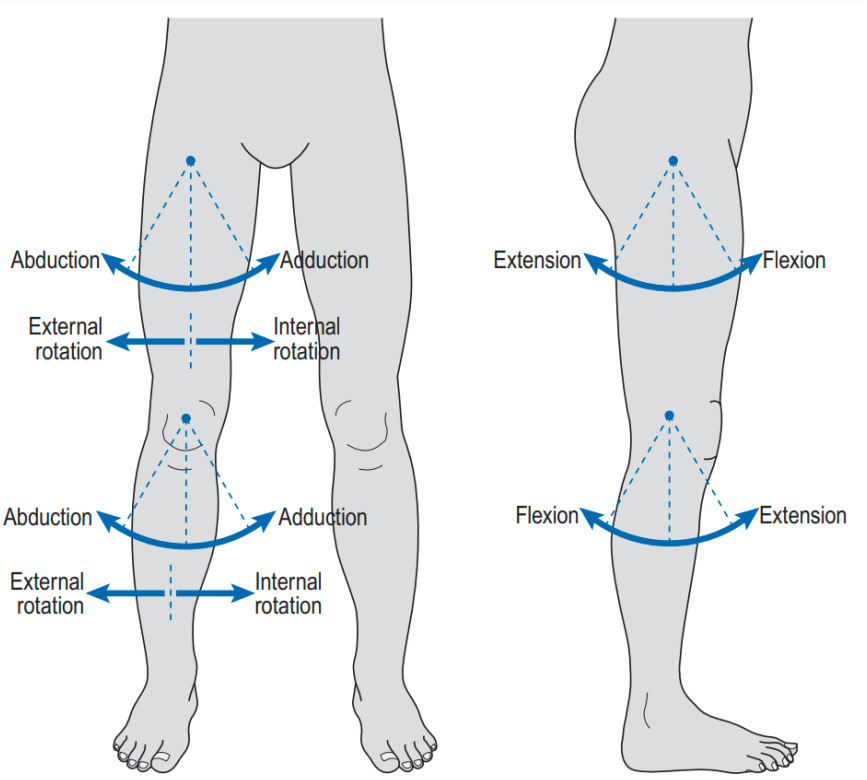


Figure 6. *Movements about the hip joint (above) and knee joint (below). Taken from [60]*

Kinematic variables for specific body parts are based on these movements, for example knee flexion/extension, hip abduction/adduction etc. The full list of kinematic variables currently measured at HNRC can be found in Appendix 3.

Calculations

The calculation steps for the Vicon system currently used at HNRC [15] can be found in the Vicon documentation [66] [4]. The documentation provided additional clarification on the description of each kinematic variable, but it was not particularly useful for creating

calculation algorithms. This is because all the information relied on markers, which are not utilised by our systems, of course.

To perform most calculations, the first step involves rotating the model (or just the needed part of it) by certain degrees in order to align the required body part with certain axes. Then, calculations involve measuring the angle of the joint with respect to a certain plane. For instance, to measure knee movement in the sagittal plane (flexion/extension), the model is rotated such that the thigh is aligned with the sagittal and frontal planes and perpendicular to the transverse plane. After this rotation, knee flexion is simply the angle between the leg (from knee to ankle) and the frontal plane.

3.5.2 General Gait Parameters

General gait parameters (also known as spatial and temporal gait parameters) constitute a small but important part of gait analysis. They describe the walking patterns of a patient in simple terms, such as step length, width, time spent in different phases, and more [67]. A complete list of general gait parameters used at HNRC is provided in Appendix 4.

These parameters were calculated mostly in consultation with physicians at HNRC [15], who provided guidance on their exact characteristics. Additionally, Vicon has comprehensive documentation on how different parameters are calculated on their website [4]. While this was helpful, the instructions were given for calculations using marker locations, which our system aims to avoid.

All calculations were performed by projecting foot movements onto the floor plane, as illustrated in Figure 7. The points were then rotated to align the walking trajectory with the x-axis. From this point, most calculations were straightforward.

3.5.3 Woltring Filter

All data of the 3D model proposed in this thesis are processed using a Woltring filter [68] to ensure smooth trajectories for calculating kinematic variables (see Section 3.5.1). The Woltring filter is commonly used in biomechanical analysis to minimize noise in marker trajectories [69]. It was originally proposed by Woltring H.J in 1986. The extent to which data is altered depends on the chosen window size of the Woltring filter. Determining the optimal windows size for such filters is not trivial and there is previous work done on this, such as the work by Roithner, Robin and Schwameder, Hermann, and Müller, Erich in 2000 [70]. However, for the purpose of this thesis, a window size of five was chosen since

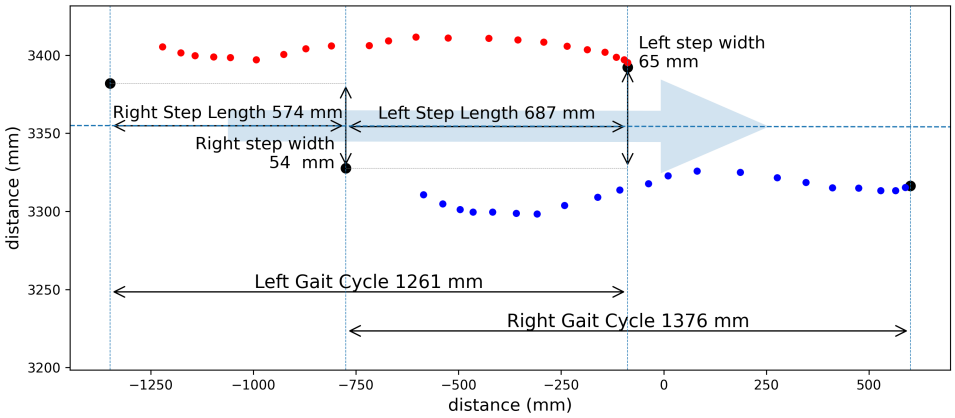


Figure 7. *Projection of heel movement onto the laboratory floor plane for both feet at the HNRC laboratory. The red dots represent left heel movement, the blue dots represent right heel movement, and the black dots indicate stationary heels. The arrow in the background indicates the direction of movement*

it is also the setting used in the current Vicon system at HNRC [15][66].

4. Results

4.1 Overview

As a result of this thesis, a 3D model was created using detections from MediaPipe (Section 3.2.1). This was a significant achievement in itself given the complexity of the task and the authors' lack of experience, and it was not absolutely certain that it could be accomplished. An example of the model can be seen in Appendix 5.

The gait cycle detection logic (Section 3.4.1) accomplishes predicting different gait cycle phases. Overall, the trained model predicts four different classes that are later used to detect gait-cycles (Section 3.4). This enables the calculation of different parameters needed for gait analysis (Section 3.5).

Additionally, a functional pipeline was built that automated and connected all the parts of the thesis. The end result is a report in PDF format that contains various kinematic variables (Section 3.5.1) and general gait parameters (Section 3.5.2). It is worth noting that the creation of the PDF was not within the scope of this thesis, and therefore, its creation is not described in detail. However, an example of the PDF can be found in Appendix 6.

4.1.1 Gait-Cycle Detection

As a result, a clear differentiation between two main gait cycle phases (see section 3.4) was achieved. This was accomplished by detecting four different classes. The four classes that we identified were crucial in providing a multi-class classification approach to gait cycle detection. These classes detected whether both feet were in contact with the ground, whether one foot was in the stance phase while the other was in the swing phase, and whether the 3D model was too inaccurate to make an accurate prediction. The four different classes allowed us to further automate the calculation of kinematic and gait parameters.

4.1.2 Gait Analysis Overview

Out of the 11 kinematic variables measured at HNRC (See list in Appendix 3) with the system proposed in this thesis right now 5 measurements are integrated (See list in Appendix 3). This is mostly due to the fact that the model has too few points on certain joints. For example, the model's pelvis consists of only two points, which makes it, in

all regards, a 2D object, meaning its rotation cannot be calculated in all three dimensions. To ensure the accuracy of the mathematical calculations and avoid any possibility of error, the authors of this thesis also did not include some kinematic variables that still had uncertainties about their exact meaning. Overall, achieving 5 out of 11 variables is a commendable accomplishment since it covers the most commonly used variables by clinicians at HNRC [15]. Additionally, it allows for a comprehensive analysis of the system's potential.

4.2 Analysis

4.2.1 Gait-Cycle Detection

Training with Auto-PyTorch

Auto-PyTorch uses automated ensemble selection. Starting with an empty set, the models which give the largest performance improvement are iteratively added to the ensemble until the set size reaches the predefined ensemble size.

One of the major advantages of this approach is that it allows for the inclusion of models other than deep neural networks (DNNs) in the ensemble. This means that Auto-PyTorch is able to leverage a wide range of models to achieve optimal predictive accuracy, regardless of the specific nature of the dataset or the task at hand. As stated in [64], the combination of different model classes is one of the key performance factors on tabular data.

The impact of training data and parameters on ensemble accuracy

The accuracy of an ensemble model is heavily influenced by the quality and quantity of the dataset used for training. This was particularly evident in an early training of an ensemble model that was done on a dataset that was recorded using only one camera. Instead of using a 3D dataset, the training was done on a 2D dataset with six different points placed on the foot (points 27-32 on Figure 4). The result was an accuracy of only 58%, which clearly indicated the limitations of the dataset. One of the major issues with this dataset was its noisiness. To address this issue, a new training was conducted with the implementation of a median filter [71] on the data. This helped to alleviate the effects of the noisy data and resulted in a significant improvement in the accuracy of the ensemble, which increased to 69%. Implementing the median filter was a crucial step in improving the dataset's quality and the ensemble model's accuracy. The filter allowed for the removal of outliers and irrelevant data points, which helped to reduce the noise in the dataset and improve the accuracy of the model.

The use of a 3D dataset for training had a significant impact on the accuracy of the ensemble model. In contrast to the 2D dataset, which only provided information in two dimensions, the 3D dataset allowed for the calculation of x-, y-, and z-coordinates, resulting in a more comprehensive representation of the data. This improvement was reflected in the accuracy of the ensemble model, which improved from 58% with the 2D dataset to 89% with the 3D dataset. This represents a remarkable 31 percentage points improvement in accuracy, demonstrating the clear advantages of using a 3D dataset for training.

In addition, another factor of the quality of the dataset is the environment. Although the research of the analysis of environments' impact on the accuracy of the ensemble was not in the scopes of this work, it was discovered that the model was better with dataset that had fewer objects and noise on the background during training. Training a new model with data recorded at HNRC resulted in higher model accuracy of 91%.

Another aspect that helped to achieve better accuracy of the model was training time. The Auto-PyTorch trains on multiple machine learning algorithms and chooses only the best of them to make an ensemble. Allowing for the model to train for a longer time allows it to choose the better model and weights. This longer training time resulted in accuracy of 94% (see Figure 8).

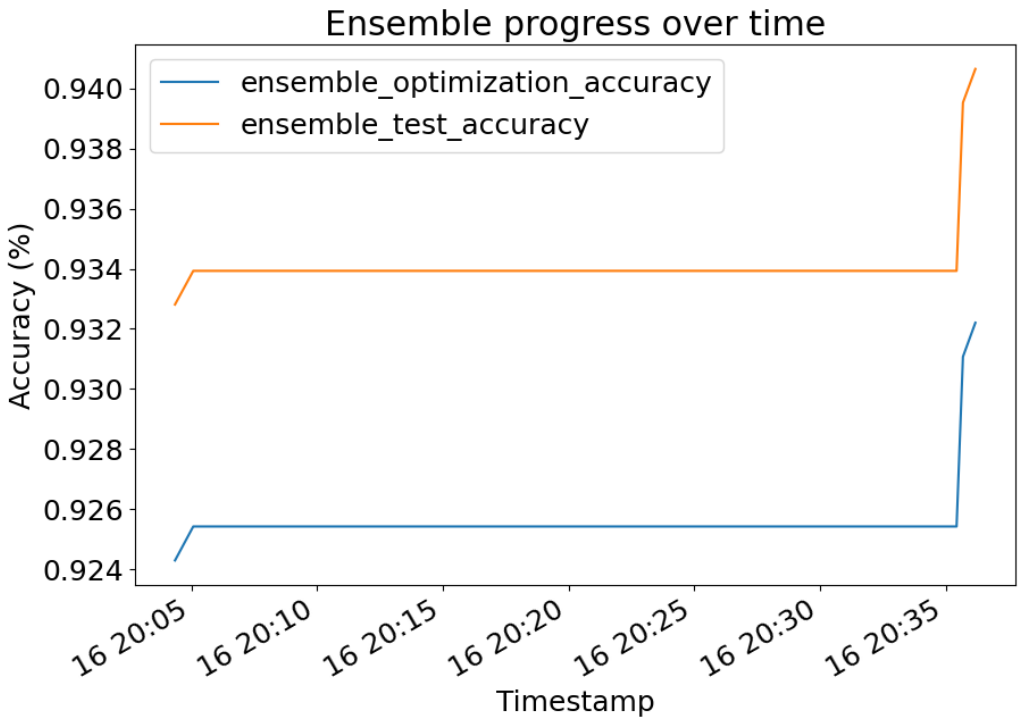


Figure 8. *Accuracy progress over time during training. Dataset with 6 body landmarks: foot, ankle and heel on both sides.*

Thus far, the model was trained with a dataset that consisted of 6 different features, them being landmarks of each foot. Since the knee and hip both have a crucial role in gait cycle phases, an additional 4 features were added to the dataset. The dataset with 4 additional features had an accuracy of 92.9% as seen in Figure 9.

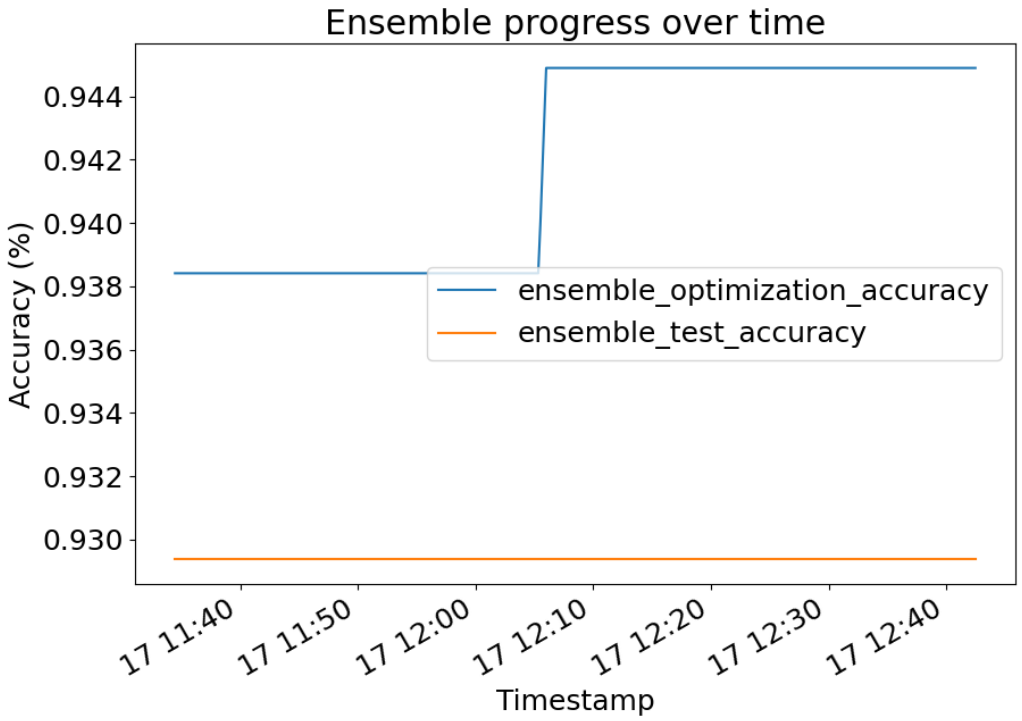


Figure 9. *Accuracy progress over time during training. Dataset with 10 body landmarks: foot, ankle, heel, knee and hip on both sides.*

4.2.2 Gait Analysis

Three gait cycles were selected for a detailed analysis, during which the results obtained from the proposed system in this thesis and those from the current solution at HNRC were visually compared for all five gait variables (See list in Appendix 3). A complete set of visual comparisons can be found in Appendices 8, 9, 10.

Visual Comparison Analysis

As observed from the graphs, the system proposed in this thesis performs relatively well with larger angles (such as knee flexion/extension), but not as well with smaller angle ranges (such as hip abduction/adduction). The most probable source of faulty results is the underlying detection mechanism of MediaPipe (See Section 3.2.1). This does not necessarily imply that MediaPipe is inaccurate, rather points to the complexity involved in accurately distilling a human body part, such as the knee or heel, down to a single point.

When dealing with larger angles, a slight variation in the position of the detected point may cause only a small change. However, with smaller angles, where we are dealing with a range of approximately 10 degrees, even a slight variation can significantly affect the results.

In terms of the actual graphs, clinicians at HNRC have indicated that the overall shape is more crucial than the specific values. This is because for diagnosis, they do not compare the actual numerical values but rather examine the shapes of the graphs overall, how the shapes compare between the left and right sides, and how they compare to control group data graph shapes. Therefore, for gait analysis purposes, the authors of this thesis believe that angles with bigger ranges, such as knee flexion/extension, hip flexion/extension, and ankle dorsiflexion/plantar flexion, can be used. However, it is important to note that the actual feedback from clinicians at HNRC is discussed in Section 4.3.2.

4.3 Validation

4.3.1 Gait-Cycle Detection

As we do not have any data to validate gait-cycle detection from HNRC, then mostly a visual validation was done for each trained model. The key areas to detect were the foot's initial contact with ground compared to annotated data. Furthermore, a good understanding of the model's performance can be understood by looking at the accuracy of the model. As stated above in Section 4.2.1 the best model had an accuracy of 94%.

In addition, the main criteria for the gait-cycle detection was its usability for gait analysis. At the moment, it is used to detect gait-cycles automatically. So by that means it also fulfils its criteria.

In conclusion, the trained model demonstrates good accuracy and effectively achieves its intended purpose of automatic and precise detection of gait-cycles.

4.3.2 Gait Analysis Validation

To validate the results, a thorough comparison of numerical values was conducted. Pearson's correlation coefficient, root-mean-square deviation (RMSE), and Spearman's rank correlation coefficient were used to determine the similarity between the results obtained by the two systems. In addition, clinicians from HNRC were consulted and asked for feedback.

Pearson Correlation

The Pearson correlation coefficient is the most common way of measuring a linear correlation, it gives a number between –1 and 1 that measures the strength and direction of the relationship between two data sets, where 1 indicates a perfect correlation, -1 a negative correlation and 0 no correlation[72]. For this thesis, as discussed at the end of Section 4.2.2, the overall shape and similarity between the results and test data is of great importance, making Pearson correlation an ideal tool for validating the results.

Table 4. Pearson correlation between the system proposed in this thesis and the current Vicon system at HNRC

Angle	Knee Flex/Ext	Hip Flex/Ext	Ankle Flex/Ext	Foot prog.	Hip Abd/Add
Trial 1.	0.939	0.969	0.825	0.723	0.284
Trial 2.	0.918	0.983	0.819	-0.068	0.374
Trial 3.	0.853	0.957	0.874	0.012	0.430
Avg	**0.903**	**0.969**	**0.839**	**0.222**	**0.362**

Table 4 provides an overview of the Pearson correlation coefficients for the three selected trials (discussed in Section 4.2.2) compared to the results obtained from the Vicon system used at HNRC. The Pearson correlation coefficient for the proposed system in this thesis shows a good correlation, particularly for gait kinematics with a larger range of angles. However, for gait kinematics with smaller angles, the correlation is poor. Similar information was derived from the visual comparisons presented in Section 4.2.2.

Spearman's Rank Correlation Coefficient

Spearman's rank correlation coefficient, is an alternative to Pearson. It uses the rankings of data from each variable, as opposed to the raw data itself. This is useful, one of the variables is on an ordinal level of measurement or when the data from one or both variables does not follow normal distributions. The Spearman correlation coefficient measures the monotonicity of relationships. In a monotonic relationship, each variable also always changes in only one direction, but not necessarily at the same rate. [73]

Table 5. Spearman's correlation between the system proposed in this thesis and the current Vicon system at HNRC

Angle	Knee Flex/Ext	Hip Flex/Ext	Ankle Flex/Ext	Foot prog.	Hip Abd/Add
Trial 1.	0.862	0.939	0.741	0.655	0.305
Trial 2.	0.881	0.956	0.711	0.025	0.229
Trial 3.	0.778	0.941	0.797	-0.026	0.194
Avg	**0.840**	**0.945**	**0.749**	**0.218**	**0.242**

This table 5 gives an overview of 3 trials that were conducted in HRNC. In this we looked

at the spearman's rank correlation between our output and that of the Vicon system on the same gait cycle. There is a similar trend, that angles in sagittal plane that are impacted by wider area of motion and bigger changes in angle, have better overall correlation to our validation results. As for the frontal and transverse angles, meaning the rotation of the foot and hip abduction adduction, since these are harder to measure and more susceptible to small changes. There is very little correlation to the overall trend and expected movement during a gait cycle.

RMSE

Root Mean Square Error (RMSE) is a standard way to measure the error of a model in predicting quantitative data. This is used to give an overview of actual difference between validation data. RMSE provides an overview of the average difference between the predicted values and the actual values of the data being analysed. It is particularly useful when dealing with continuous data, as it can measure the deviation between the predicted and actual values in a meaningful way.[74]

Table 6. RMSE (degrees) between the system proposed in this thesis and the current Vicon system at HNRC

Angle	Knee Flex/Ext	Hip Flex/Ext	Ankle Flex/Ext	Foot prog.	Hip Abd/Add
Trial 1.	7.703	7.978	3.959	3.179	4.267
Trial 2.	8.659	8.549	4.377	7.355	2.318
Trial 3.	10.702	8.725	3.331	7.560	2.363
Avg	**9.02**	**8.41**	**3.88**	**6.03**	**2.98**

The overall RMSE accuracy in table 6 is not as precise as the official data and still leaves a lot to be desired. It is important to note, that the bigger difference for flexion and extension, angles in sagittal plane are bigger due to an overall larger angle range that can be around 70 degrees. While hip abduction adduction takes place on smaller scale and thus even with overall bad correlation, the RMSE can give misleadingly good results. Table 7 displays RMSE normalised between 0 and 1 by the scale of the data to make RMSE results more understandable. There is still a dominant margin of error and our system cannot fully compete when it comes to accuracy. Overall, angles with a larger range of degrees (such as knee flexion/extension) clearly give a better result.

Interpretation and Analysis of Correlation and RMSE Results

Pearson correlation and Spearman's correlation provide an insight into the general shape of the data, while RMSE measures the deviation from the test set. Strong correlations and noticeable RMSE values are observed with kinematic variables that have a wider range of angles. This indicates that while the pattern of the data is similar to the test data, it is not

Table 7. Normalised Root Mean Square Error between the system proposed in this thesis and the current Vicon system at HNRC

Angle	Knee Flex/Ext	Hip Flex/Ext	Ankle Flex/Ext	Foot prog.	Hip Abd/Add
Trial 1.	0.128	0.174	0.160	0.258	0.429
Trial 2.	0.140	0.200	0.164	0.604	0.261
Trial 3.	0.178	0.201	0.132	0.728	0.250
Avg	**0.15**	**0.19**	**0.15**	**0.53**	**0.31**

in the exact position. Some part of this inaccuracy can be attributed to the way MediaPipe places body part landmark locations 'inside' the body part, as opposed to the Vicon system markers that are placed on the skin. Depending on the angle being measured and the placement of markers, this can cause the same pattern to appear in a slightly different range.

Feedback From HNRC clinicians

Clinicians at HNRC, including physiotherapists Killu Mägi and Leida Pikas, quality specialist Maire Nigul, and service manager Mari-Liis Ööpik-Loks, provided feedback to the authors of this thesis.

The physiotherapists concluded that the system would be useful for them in certain cases. Overall, the feedback was mostly positive, and everyone agreed that while this thesis does not achieve the same level of accuracy as the current infrared 8-camera system, it is still considerably useful. The HNRC clinicians agreed that the proposed system would be valuable in smaller hospitals without a gait laboratory. Since it requires only two inexpensive cameras and offers useful functionalities, they believe it would be a significant help. Of course, the current system has limitations, as discussed further in Section 5.1, which were also discussed with the clinicians. This discussion generated many ideas for potential future development and improvement, which are explored in Section 5.2.

5. Discussion

5.1 Current Limitations

5.1.1 Model Limitations

As discussed in Section 4.1.2, the model proposed in this thesis has some limitations due to the limited number of landmarks on the MediaPipe pose model (see Section 3.2.1 and 4). The limited number of landmarks makes some body parts two-dimensional, and their rotation cannot be calculated in all three dimensions. This makes some gait kinematics (see Section 3.5.1), such as knee/hip rotations, quite difficult, if not impossible.

The small number of pose landmarks also limits the shape that the model can represent. For example, having more landmarks on the legs/back would allow for a more detailed representation of a person with severe physical lower body abnormalities. As discussed further in Section 5.2 these problems can be overcome by implementing extra landmarks to the MediaPipe model. It is worth noting that MediaPipe is currently in its alpha version at v0.7 [32]. Therefore, significant development can be expected in the future, and the MediaPipe pose model could become much more precise and detailed.

5.1.2 Camera Limitations

As for the camera's picture quality, it is not a big issue since there is not much difference in picture quality and detection results. However, there is a distance factor, as most frameworks have a range in which they are most capable in terms of detection. Thus, a field of view is important to keep the cameras closer. This creates a problem, as bringing the camera closer limits the amount of space where the patient can be clearly seen. This means that the full gait cycle might not be captured for taller patients or patients with longer steps, and necessary data for gait analysis cannot be obtained.

The current system is very basic, as it only uses two cameras. This means that we have the minimum required systems to create a 3D model from pose estimation results. However, adding a third camera that provides a direct view of the other side of the body could significantly improve the 3D model, and it wouldn't drive the cost up significantly.

5.2 Future Work

The system has great potential for further enhancement, which can be achieved through two avenues: fine-tuning and expansion of the existing system/model and developing entirely new features.

5.2.1 Future Changes

Pose Detection Landmarks

As mentioned in Section 5.1.1, the current 3D model has limitations due to the limited number of landmarks in the MediaPipe pose model (see Section 3.2.1 and 4). To overcome this, additional custom landmark detections could be implemented in the future. Although this would require a significant amount of training data, it has the potential to greatly enhance the system's functionality and usefulness.

Different Systems Integrations

The addition of an extra camera could significantly improve the 3D model, and it wouldn't drive the cost up significantly. The extra camera could cover the side of the patient that is currently not captured by either of the cameras.

Moreover, there are various sensors that could further enhance the stability of our output. In particular, the IMU and GRF (as explained in Section 3.1) hold great promise in addressing the limitations of the current system by incorporating external data. GRF, in particular, could prove invaluable in analysing the kinetics of the patient by providing information on force, moment, and power.

5.2.2 New Features

Visualisation

Although visualisation of the 3D model was not the focus of this thesis, we currently utilise simple GIFs and plots to convey the general idea. However, in the future, third-party solutions could be integrated to provide a more user-friendly interface for visual inspection of the model from various angles. This was also recommended by physicians at HNRC (see Section 4.3.2).

Furthermore, we could incorporate a feature to visually compare multiple walking trials, which physicians at HNRC have strongly recommended. Such a system would enable

them to assess how much patients have improved and how their physical condition has changed over time between sessions. In addition, the dynamic time warping algorithm could be used for some interesting comparisons between two walking trials [75][76].The current solution at HNRC does not provide such features.

Analysis Automatisation

Another area that could be further improved is the post-procedure analysis currently performed by physiotherapists. With enough training data, this analysis could be significantly automated. By inputting measurements from walking trials (such as gait kinematics), the physician could receive a set of possible symptoms, solutions, and other relevant information. This would not be a final diagnosis but rather a helpful mechanism to speed up the work process for hospital staff and assist them in detecting key issues more effectively.

6. Summary

The primary aim of this thesis was to develop a markerless system with two RGB cameras for cerebral palsy gait analysis to reduce the time consumption, inconvenience and the needed hours of clinicians' work.

As a result of this thesis, a 3D model was created using detections from MediaPipe, an open-sourced pose estimation framework. The gait cycle detection logic accomplishes predicting different gait cycle phases. Overall, the trained model predicts four different classes that are later used to detect gait-cycles. This enables the calculation of different parameters needed for gait analysis.

While most open-source pose detection framework models are not specifically designed for gait analysis, this study demonstrates that they can provide valuable insights while requiring less specialized tools and being easier to work with and adapt to different needs. Based on this, the put-together 3D model from two 2D pose estimation results can already give a good visual aid to specialists.

The results of this study show that the system gait analysis provides decent results for common kinematic variables, with strong Pearson's correlations for knee (0.903), hip (0.969), and ankle (0.839) in the sagittal plane. The system performs better for larger and wider ranges of motion, which are easier to detect and measure. This leads to better overall correlation and distribution, but there are still not accurate enough with RMSE of the knee (9.02), hip (8.41), and ankle (3.88) degrees, respectively in the sagittal plane. However, angles with smaller changes in respect to movement of the body are less reliable and accurate since the detection does not use specific markers, and excessive movement is more prone to shifting or being insufficiently sharp to discern specific changes in rotation. Examples of this are the foot progression angle (Pearson 0.222 and RMSE 6.03) and the abduction adduction of the hip (Pearson 0.362 and RMSE 2.98)

This work serves as a valuable foundation for future research and testing in the field of gait analysis and cycle detection. The results demonstrate the potential for a new, cost-effective system that could offer an alternative to current implementations. With further development and refinement, this system could provide a more efficient and accessible solution for gait analysis, ultimately benefiting patients and healthcare providers alike.

References

[1] *What is Cerebral Palsy?* https://www.cdc.gov/ncbddd/cp/facts.html. Centers for Disease Control and Prevention, Accessed 2023.

[2] Maureen S Durkin et al. "Prevalence of Cerebral Palsy among 8-Year-Old Children in 2010 and Preliminary Evidence of Trends in Its Relationship to Low Birthweight". In: *Paediatric and perinatal epidemiology* (2016).

[3] Analytics Vidhya. *A Comprehensive Guide on Human Pose Estimation*. Accessed: 2023. URL: https://www.analyticsvidhya.com/blog/2022/01/a-comprehensive-guide-on-human-pose-estimation/.

[4] Vicon Motion Systems Ltd. *Vicon Motion Systems Ltd*. https://www.vicon.com/. Accessed: April 7, 2023.

[5] *Haapsalu Neurological Rehabilitation Centre Clinical motion and gait analysis laboratory*. Accessed: April 7, 2023. URL: https://www.hnrk.ee/services/clinical-motion-and-gait-analysis-laboratory/?lang=en.

[6] Stéphane Armand, Gérard Decoulon, and Alice Bonnefoy-Mazure. *Gait analysis in children with cerebral palsy*. 2016. DOI: 10.1302/2058-5241.1.000052.

[7] *Gait analysis*. https://www.orthomedctr.com/gait-analysis.php. Accessed: April 18, 2023.

[8] *GAIT ANALYSIS EXPLAINED*. https://www.runnersneed.com/expert-advice/gear-guides/gait-analysis.html. Accessed: April 18, 2023.

[9] Mendez-Zorrilla A. Muro-de-la-Herran A Garcia-Zapirain B. *Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications.* 2014. DOI: 10.3390/s140203362.

[10] Chris Kirtley. *Clinical Gait Analysis: Theory and Practice*. 2006. ISBN: 9780443100093.

[11] Vicon Motion Systems Limited. *Go Further with Vicon MX T-Series*. https://docs.vicon.com/download/attachments/71237793/T-Series_GoFurther_Rev1.3_2010Aug.pdf?version=1&modificationDate=1533912334000&api=v2. Accessed: April 18, 2023. 2010.

[12] *piA640-210gc datasheet*. http://www.altavision.com.br/Datasheets/Basler_EN/piA640-210gc.html. Accessed: April 18, 2023.

[13] *Vicon POLYGON*. https://www.vicon.com/software/polygon/. Accessed: April 18, 2023.

[14] *Vicon NEXUS*. https://www.vicon.com/software/nexus/. Accessed: April 18, 2023.

[15] *Haapsalu Neurological Rehabilitation Centre*. Accessed: April 7, 2023.

[16] Jennifer L McGinley et al. "The reliability of three-dimensional kinematic gait measurements: A systematic review". In: *Gait posture* (2009). DOI: 10.1016/j.gaitpost.2008.09.003.

[17] Lars Mündermann, Stefano Corazza, and Thomas P Andriacchi. "The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications". In: *Journal of NeuroEngineering and Rehabilitation* (2006). DOI: 10.1186/1743-0003-3-6. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1513229/.

[18] Matteo Moro et al. "Markerless vs. Marker-Based Gait Analysis: A Proof of Concept Study". In: *Sensors* 22.5 (2011). [Accessed: October 18, 2022]. DOI: 10.3390/s22052011. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8914751/.

[19] Y. Ma et al. *The Validity and Reliability of a Kinect v2-Based Gait Analysis System for Children with Cerebral Palsy*. Accessed: April 7, 2023. 2019. URL: https://doi.org/10.3390/s19071660.

[20] Mario Nieto-Hidalgo et al. "Gait Analysis Using Computer Vision Based on Cloud Platform and Mobile Device". In: *Mobile Information Systems* (2018). DOI: 10.1155/2018/7381264. URL: https://doi.org/10.1155/2018/7381264.

[21] Reinoud Maex et al. "A 2D Markerless Gait Analysis Methodology: Validation on Healthy Subjects". In: *BioMed Research International* 2015 (2015), p. 186780.

[22] *How Infrared Cameras Work*. https://www.fluke.com/en/learn/blog/thermal-imaging/how-infrared-cameras-work. Accessed: April 14, 2023.

[23] *Infrared marker-based motion capture*. https://www.futurelearn.com/info/courses/music-moves/0/steps/12692. Accessed: April 14, 2023.

[24] Martin Sandau et al. "Markerless motion capture can provide reliable 3D gait kinematics in the sagittal and frontal plane". In: *Medical Engineering Physics* 36.9 (2014), pp. 1168–1175. ISSN: 1350-4533. DOI: https://doi.org/10.1016/j.medengphy.2014.07.007. URL: https://www.sciencedirect.com/science/article/pii/S1350453314001775.

[25] Azmi Haider and Hagit Hel-Or. “What Can We Learn from Depth Camera Sensor Noise?” In: *Sensors* 22.14 (2022). ISSN: 1424-8220. DOI: 10.3390/s22145448. URL: https://www.mdpi.com/1424-8220/22/14/5448.

[26] Prabu Kumar. *What are depth-sensing cameras? How do they work?* https://www.e-consystems.com/blog/camera/technology/what-are-depth-sensing-cameras-how-do-they-work/. Sept. 2022.

[27] Edouard Auvinet, Franck Multon, and Jean Meunier. “Gait analysis with multiple depth cameras”. In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2011, pp. 6265–6268. DOI: 10.1109/IEMBS.2011.6091546.

[28] Edouard Auvinet, Jean Meunier, and Franck Multon. “Multiple depth cameras calibration and body volume reconstruction for gait analysis”. In: *2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*. 2012, pp. 478–483. DOI: 10.1109/ISSPA.2012.6310598.

[29] Thomas Seel, Jörg Raisch, and Thomas Schauer. “IMU-Based Joint Angle Measurement for Gait Analysis”. In: *Sensors* 14.4 (2014), pp. 6891–6909. ISSN: 1424-8220. DOI: 10.3390/s140406891. URL: https://www.mdpi.com/1424-8220/14/4/6891.

[30] Frantisek Vaverka et al. “System of gait analysis based on ground reaction force assessment”. In: *Acta Gymnica* 45.4 (2015), pp. 187–193. ISSN: 23364912. DOI: 10.5507/ag.2015.022. URL: https://gymnica.upol.cz/artkey/gym-201504-0004.php.

[31] Fritz AI. *Fritz AI*. Accessed: 2023. URL: https://www.fritz.ai/pose-estimation/.

[32] mediapipe Contributors. *mediapipe git Documentation*. Accessed: 2023. URL: https://github.com/google/mediapipe.

[33] Hao-Shu Fang et al. “AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.6 (2023), pp. 7157–7173. DOI: 10.1109/TPAMI.2022.3222784.

[34] Jiefeng Li et al. “Crowdpose: Efficient crowded scenes pose estimation and a new benchmark”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 10863–10872.

[35] Hao-Shu Fang et al. “RMPE: Regional Multi-person Pose Estimation”. In: *ICCV*. 2017.

[36] Zhe Cao et al. “Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields”. In: (July 2017).

[37] Tomas Simon et al. “Hand Keypoint Detection in Single Images Using Multiview Bootstrapping”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 2017.

[38] Zhe Cao et al. “Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 2017.

[39] Shih-En Wei et al. “Convolutional pose machines”. In: *CVPR*. 2016.

[40] István Sárándi et al. “MeTRAbs: Metric-Scale Truncation-Robust Heatmaps for Absolute 3D Human Pose Estimation”. In: *IEEE Transactions on Biometrics, Behavior, and Identity Science* 3.1 (2021), pp. 16–30. DOI: 10.1109/TBIOM.2020.3037257.

[41] István Sárándi et al. “Metric-Scale Truncation-Robust Heatmaps for 3D Human Pose Estimation”. In: *IEEE International Conference on Automatic Face and Gesture Recognition*. 2020, pp. 677–684.

[42] tensorflow movenet tutorial Contributors. *tensorflow movenet tutorial Documentation*. Accessed: 2023. URL: https://www.tensorflow.org/hub/tutorials/movenet.

[43] Bowen Cheng et al. “HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose Estimation”. In: *CVPR*. 2020.

[44] Ke Sun et al. “Deep High-Resolution Representation Learning for Human Pose Estimation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019.

[45] Jingdong Wang et al. “Deep High-Resolution Representation Learning for Visual Recognition”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.10 (2021), pp. 3349–3364. DOI: 10.1109/TPAMI.2020.2983686.

[46] Yuxin Wu et al. *Detectron2*. https://github.com/facebookresearch/detectron2. 2019.

[47] OpenCV.org. *OpenCV – Open Source Computer Vision Library*. Accessed on April 10, 2023. URL: https://opencv.org/.

[48] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O’Reilly, 2008.

[49] J. Heikkila and O. Silven. *A Four-Step Camera Calibration Procedure with Implicit Image Correction*. Conference on Computer Vision and Pattern Recognition, 1997. Proceedings CVPR '97. San Juan, PR, USA, June 1997. DOI: 10.1109/CVPR.1997.609468.

[50] Lior Wolf and Amnon Shashua. *On Projection Matrices $\mathcal{P}^k \rightarrow \mathcal{P}^2$, $k = 3, ..., 6$, and their Applications in Computer Vision*. ID: Wolf2002. 2002. DOI: 10.1023/A:1014855311993. URL: https://doi.org/10.1023/A:1014855311993.

[51] Richard I. Hartley and Peter Sturm. "Triangulation". In: *Computer Vision and Image Understanding* 68.2 (1997), pp. 146–157. DOI: https://doi.org/10.1006/cviu.1997.0547.

[52] Richard Hartley. *Triangulation made easy*. ISBN: 978-1-4244-6984-0, ISSN: 1063-6919, INSPEC Accession Number: 11500338. San Francisco, CA, USA, June 2010. DOI: 10.1109/CVPR.2010.5539785.

[53] Richard Szeliski and Sing Bing Kang. *Spacetime stereo: a unifying framework for depth from triangulation*. Madison, WI, USA: IEEE, June 2003. DOI: 10.1109/CVPR.2003.1211491.

[54] Kari Pulli et al. *Real-time computer vision with OpenCV*. Communications of the ACM, Volume 55, Number 6. 2012. DOI: 10.1145/2184319.2184337.

[55] OpenCV Contributors. *OpenCV Documentation*. Accessed: 2023. URL: https://docs.opencv.org/4.x/.

[56] Roberto Grujičić. *Gait cycle*. accessed April 10, 2023. URL: https://www.kenhub.com/en/library/anatomy/gait-cycle.

[57] Dave Thompson. *Muscle activity during the gait cycle*. accessed April 10, 2023. URL: https://ouhsc.edu/bserdac/dthompso/web/gait/kinetics/mmactsum.htm.

[58] Amen J et al. "Single-event multilevel surgery for crouching cerebral palsy children: Correlations with quality of life and functional mobility." In: *Journal of Musculoskeletal Surgery and Research* (2018). DOI: 10.4103/jmsr.jmsr_48_18.

[59] Park MS et al. "Issues of concern after a single-event multilevel surgery in ambulatory children with cerebral palsy." In: *J Pediatr Orthop.* (2009). DOI: 10.1097/BPO.0b013e3181b529e8.

[60] Michael W. Whittle 4th Edition. *Gait Analysis: An Introduction*. Elsevier, 2007.

[61] Freiburg Hannover Tübingen. *AutoML*. Accessed: 2023. URL: https://www.automl.org/automl/.

[62] Chris Thornton et al. “Auto-WEKA: Combined Selection and Hyperparameter Optimization of Classification Algorithms”. In: 2013. arXiv: 1208.3719 [cs.LG].

[63] Matthias Feurer et al. “Auto-Sklearn 2.0: Hands-free AutoML via Meta-Learning”. In: (2022). arXiv: 2007.04074 [cs.LG].

[64] Lucas Zimmer, Marius Lindauer, and Frank Hutter. “Auto-PyTorch Tabular: Multi-Fidelity MetaLearning for Efficient and Robust AutoDL”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021). also available under https://arxiv.org/abs/2006.13799, pp. 3079–3090.

[65] Michael W. Whittle. *Gait Analysis: An Introduction*. Ed. by Heidi Harrison and Siobhan Campbell. 4th. First published in 1990. Reprinted in 1997. Second edition published in 1996. Third edition published in 2002. Reprinted in 2003 (twice), 2004, and 2005. London, UK: Elsevier Ltd, 2007.

[66] Vicon. *Plug-in Gait kinematic variables*. https://docs.vicon.com/display/Nexus214/Plug-in+Gait+kinematic+variables. Accessed on April 13, 2023. 2023.

[67] C. Kirtley, M.W. Whittle, and R.J. Jefferson. *Influence of walking speed on gait parameters*. Journal of Biomedical Engineering, Volume 7, Issue 4. 1985. DOI: 10.1016/0141-5425(85)90055-X.

[68] Herman J. Woltring. “A Fortran package for generalized, cross-validatory spline smoothing and differentiation”. In: *Advances in Engineering Software (1978)* 8.2 (1986), pp. 104–113. ISSN: 0141-1195. DOI: https://doi.org/10.1016/0141-1195(86)90098-7. URL: https://www.sciencedirect.com/science/article/pii/0141119586900987.

[69] M. Molloy et al. “The effects of industry standard averaging and filtering techniques in kinematic gait analysis”. In: *Gait Posture* 28.4 (2008), pp. 559–562. ISSN: 0966-6362. DOI: https://doi.org/10.1016/j.gaitpost.2008.03.012. URL: https://www.sciencedirect.com/science/article/pii/S096663620800091X.

[70] Robin Roithner, Hermann Schwameder, and Erich Müller. “Determination of optimal filter parameters for filtering kinematic walking data using butterworth low pass filter”. In: July 2000.

[71] *Median Filter*. Accessed: 2023. URL: https://www.statistics.com/glossary/median-filter/.

[72] Shaun Turney. *Pearson Correlation Coefficient*. https://www.scribbr.com/statistics/pearson-correlation-coefficient/. Revised on December 5, 2022. May 2022.

[73] Pritha Bhandari. *Correlation Coefficient: Formula, Calculation & Interpretation*. https://www.scribbr.com/statistics/correlation-coefficient/#spearmans-rho. Revised on December 5, 2022. Aug. 2021.

[74] James Moody. "What does RMSE really mean?" In: *Towards Data Science* (Sept. 2019). URL: https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e.

[75] Esmaeil Alizadeh. *An Introduction to Dynamic Time Warping*. 2021. URL: https://builtin.com/data-science/dynamic-time-warping (visited on 04/14/2023).

[76] Adam Switonski, Henryk Josinski, and Konrad Wojciechowski. "Dynamic time warping in classification and selection of motion capture data". In: *Multidimensional Systems and Signal Processing* 30 (2019), pp. 1437–1468. DOI: 10.1007/s11045-018-0584-4.

Appendix 1 – Non-Exclusive License for Reproduction and Publication of a Graduation Thesis[1]

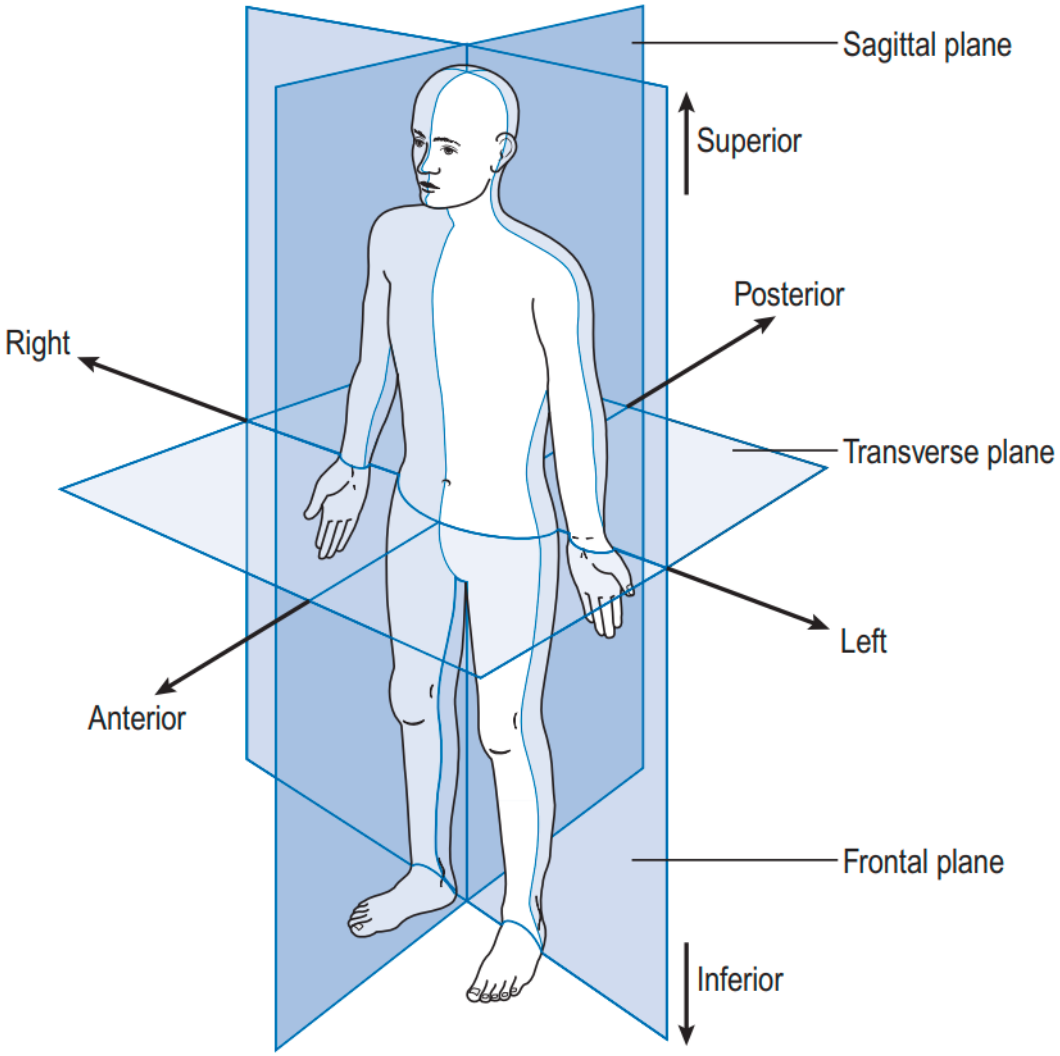
I Peeter Maran, Anton Osvald Kuusk, Karl Mihkel Seenmaa

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis "Markerless computer vision based framework for cerebral palsy gait analysis for Haapsalu Neurological Rehabilitation Center", supervised by Gert Kanter and Elli Valla
   1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;
   1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.
2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

22.05.2023

[1]The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.

Appendix 2 – Anatomical Planes



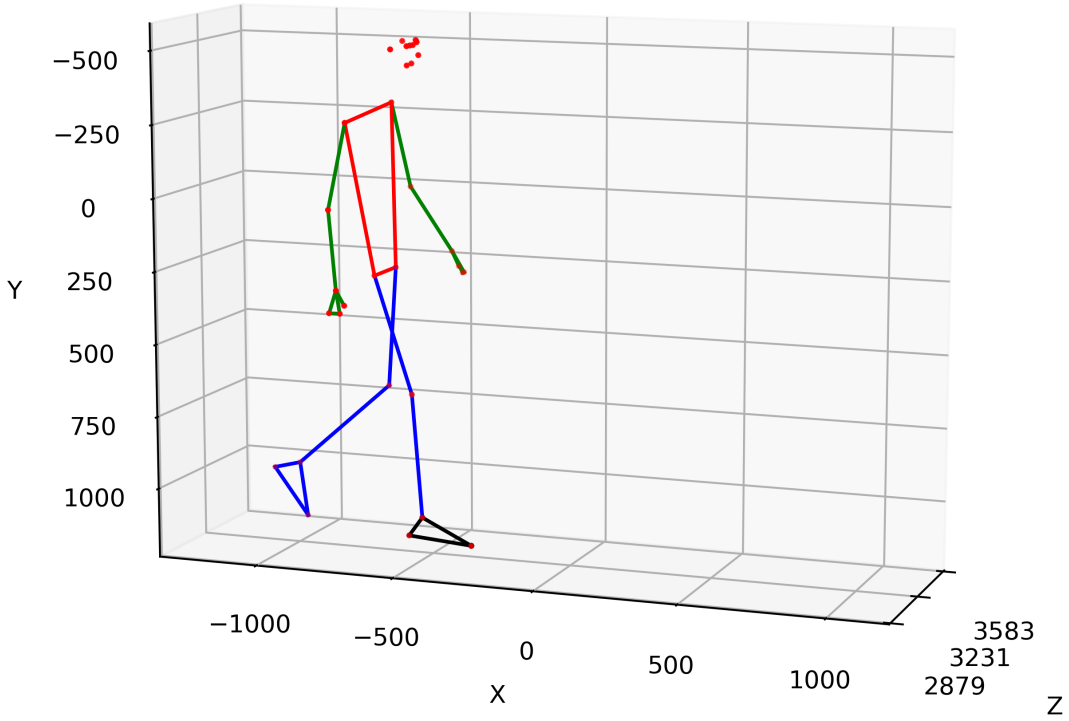
Appendix 3 – Gait Kinematics at HNRC

- Sagittal plane
  - Pelvis anterior/posterior
  - Hip flexion/extension
  - Knee flexion/extension
  - Ankle dorsiflexion/plantar flexion
- Frontal plane
  - Pelvis superior/inferior
  - Knee varus/valgus
  - Hip abduction/adduction
- Transverse plane
  - Foot progression angle
  - Pelvis transverse plane
  - Hip internal external rotation
  - Knee internal external rotation

Appendix 4 – General Gait Parameters Used at HNRC

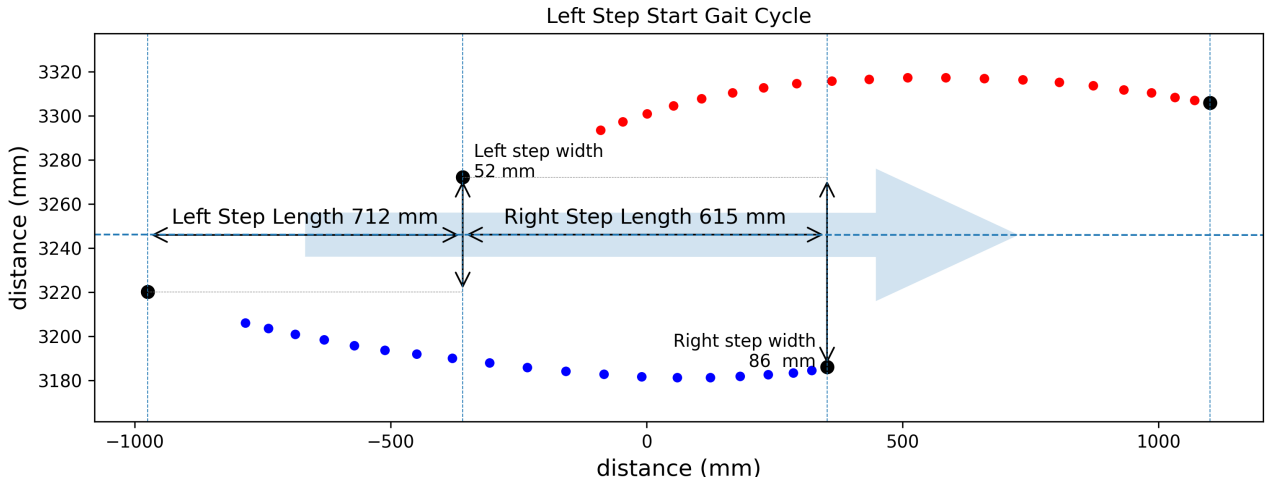
- Cadence
- Single Support
- Double support
- Final Contact
- Step Length
- Step Width
- Walking Speed
- Limp Index

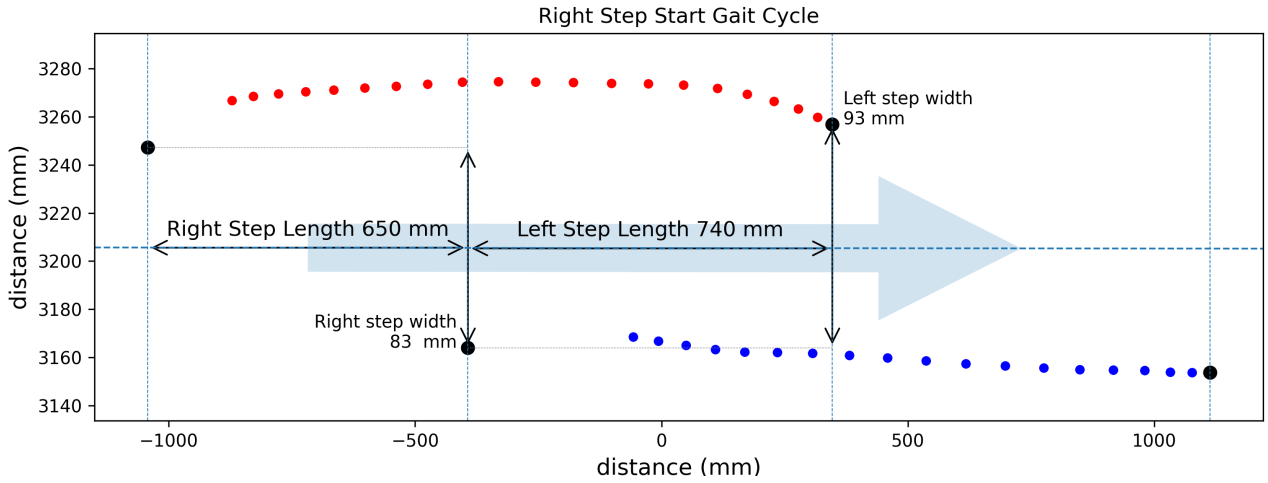
Appendix 5 – Model Example



Appendix 6 – PDF Example

Gait Analysis Report





- left heel position when foot isn't in contact with ground
- right heel position when foot isn't in contact with ground
- heel positions when both feet are in contact with ground

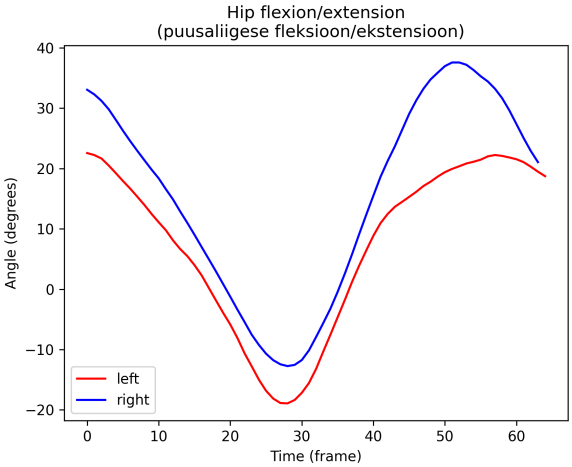
source file: New Session03 New Session09 date: 2023-04-13 time: 18:02:33.708297

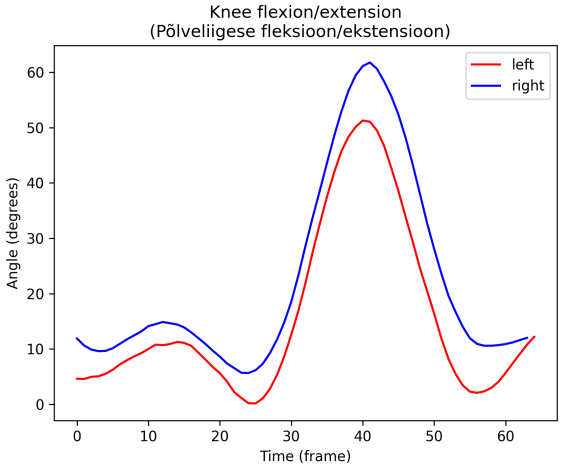
General Gait Parameters

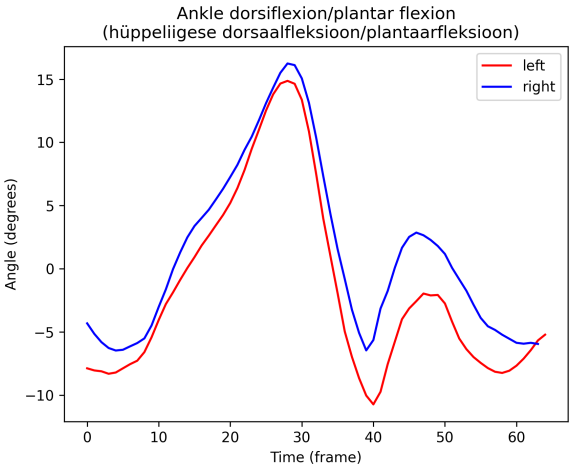
	CONTROL	RIGHT	LEFT
Step Length Sammu pikkus	0 mm	650 mm	712 mm
Step Width Sammu laius	0 mm	83 mm	52 mm
Single Support Üksiktoefaas	0 s	0.4 s	0.38 s
Double Support Kaksiktoefaas	0 s	0.16 s	0.18 s
Cadence Sammu sagedus	0 steps/min		108.12 steps / min
Walking Speed Sammu kiirus	0 mm	1369.09 mm/s	1303.57 mm/s
Final Contact Lõppkontakt	0 %	57.81 %	55.38 %
Limp Index Asümmeetria indeks	0	1.04	1.0

source file: New Session03 New Session09 date: 2023-04-13 time: 18:02:34.047297

Sagittal plane (sagitaaltasapind)

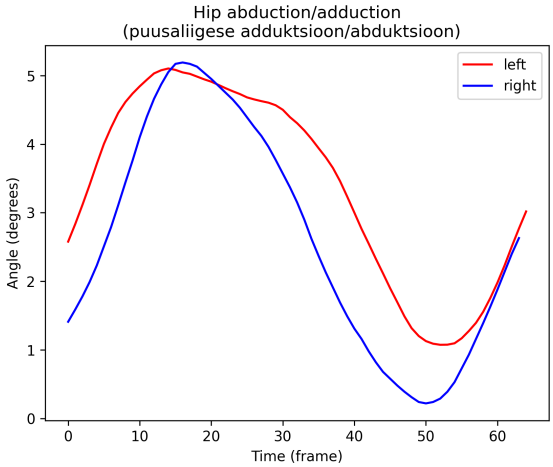


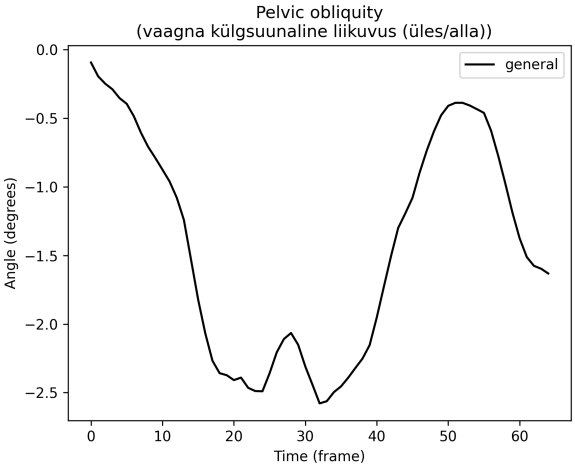




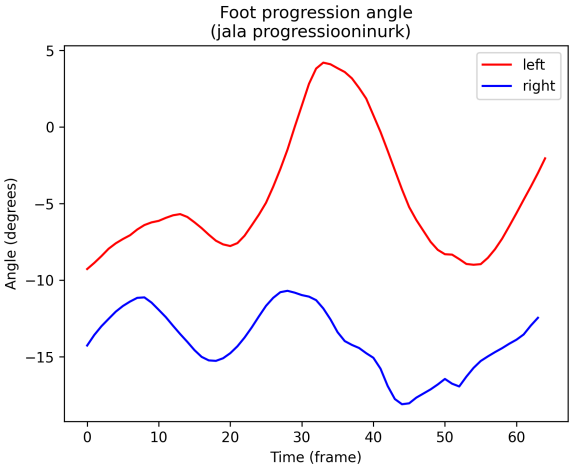
source file: New Session03 New Session09 date: 2023-04-13 time: 18:02:35.242301

Frontal plane (frontaaltasapind e. koronaaltasapind)





Transverse plane (transversaaltasapind)

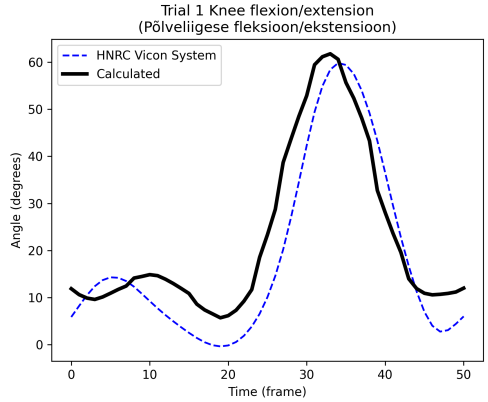


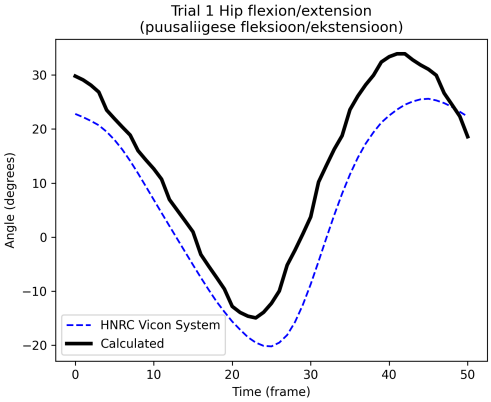
source file: New Session03 New Session09 date: 2023-04-13 time: 18:02:36.015296

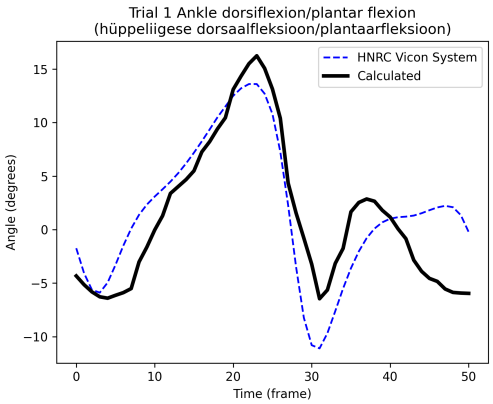
Appendix 7 – Gait Kinematics Measured In This Thesis

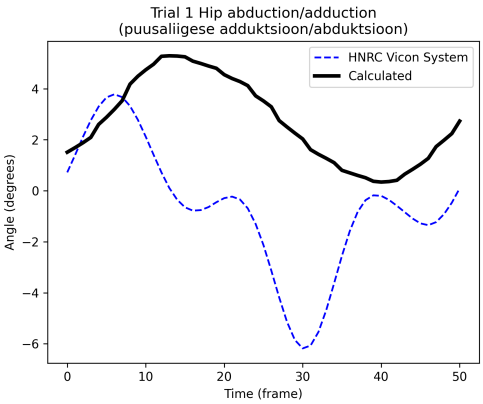
- Sagittal plane
  - Hip flexion/extension
  - Knee flexion/extension
  - Ankle dorsiflexion/plantar flexion
- Frontal plane
  - Hip abduction/adduction
- Transverse plane
  - Foot progression angle

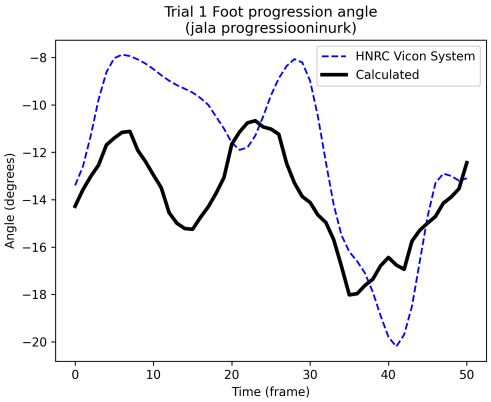
Appendix 8 – Visual Comparisons for Trial 1



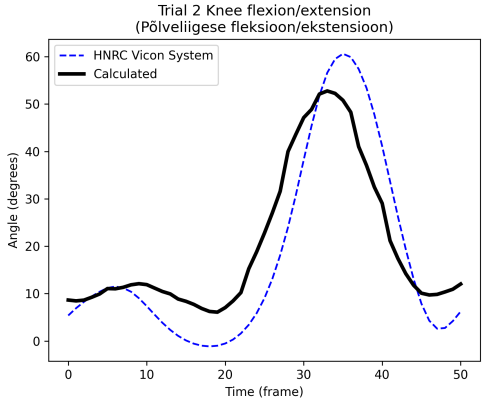


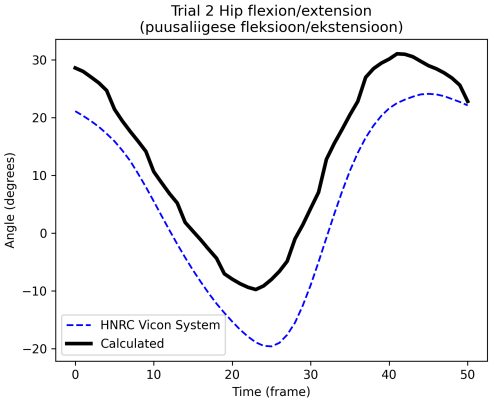


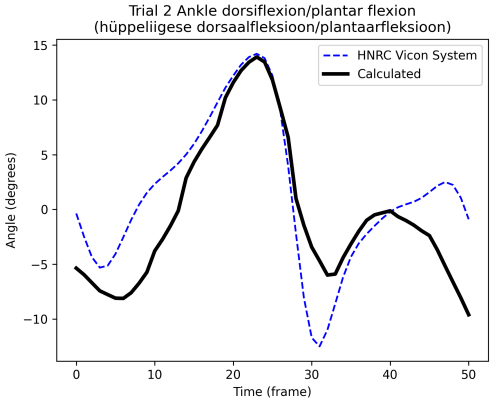


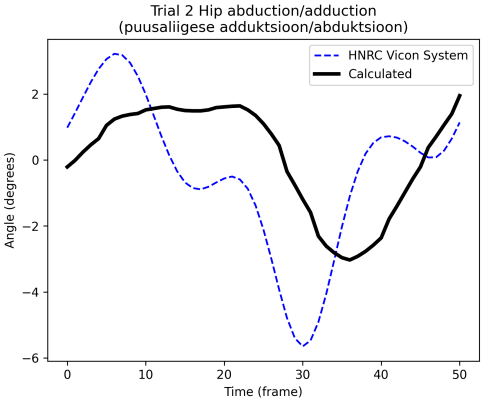


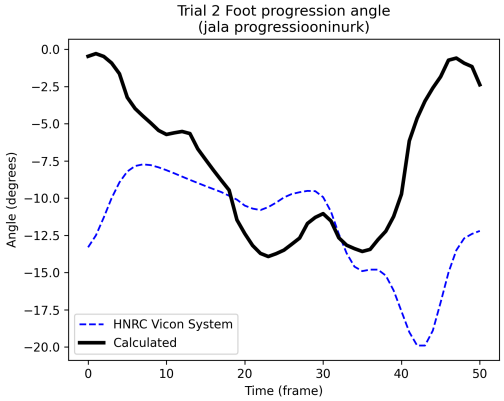
Appendix 9 – Visual Comparisons for Trial 2











Appendix 10 – Visual Comparisons for Trial 3