UNIVERSITY OF TWENTE.

Faculty of Engineering Technology

The Comparison of the Accuracy of Camera-based Skeleton Trackers During the Stepping Response for Recovering Balance in a Home Based Environment.

Wessel W. Nieuwenhuys

M.Sc. Thesis

Document number BE-869
July 2022

Supervisors:
A. Ruiz Rodriguez MSc.
dr. E.H.F. van Asseldonk
dr. G.R.H Regterschot

University of Twente
Faculty of Engineering Technology,
Department of Biomechanical Engineering
P.O. Box 217
7500 AE Enschede
The Netherlands

Preface

In your hands, you are holding my MSc thesis named: ”The Comparison of the Accuracy of Camera-Based Skeleton Trackers during the Stepping Response for Recovering Balance in a Home Based Environment”. It has been written to finish my Biomechanical Engineering Master at the University of Twente. It has been in the work since October 2021 and is now finished in July 2022.

It was written for the HEROES Project at the University of Twente, which is currently developing an exergame for at-home stroke rehabilitation. I hope the work you are about to read will provide clear information to help this Project forward.

I would like to thank my supervisors for helping along every step of the way of performing and writing about this research. First of all, Aurora Ruiz Rodriguez MSc., who was always available for a quick question, a fun conversation or a brainstorming session about the project. Our (almost) weekly meetings were always useful and I really appreciated you not only looking on for my academic work, but also always checking in on my personal life. Second, Dr. Edwin van Asseldonk for providing critical looks at my written and experimental work and always having good feedback on those. Lastly, Dr. Ruben Regterschot for being willing to join the committee during the final phases of the project and still being able to provide my work with feedback, even though you had to be brought up to speed quickly.

For all people who were willing to help me perform my experiments, I would like to thank you as well for your willingness to do so. Your help has been invaluable.

I hope you enjoy your reading!

-Wessel Nieuwenhuys

Summary

Introduction: For the 350 000 stroke survivors in the Netherlands, frequent falling is one of the most common medical complications. The consequences of falling can be severe, with People with Stroke having four times greater odds of hip fractures in the event of a fall. The HEROES Project aims to reduce fall-related healthcare utilization and associated costs, prevent falls and help People with Stroke stay independent after being discharged from the hospital. This is done by developing an exergaming system for at-home rehabilitation, which is focused on the stepping response of patients. This exergame will use a Depth camera with a skeleton tracking algorithm to estimate the joint locations of users. Various skeleton tracking algorithms do exist without clear information on which performs the best in an at-home environment where various factors could influence the quality of skeleton tracking.
Materials and Methods: Two experiments have been performed to quantify the quality of skeleton tracking during a stepping exercise by three skeleton trackers: Cubemos, Nuitrack and Openpose. During one experiment, the three algorithms were compared to a golden standard, Qualisys, using a Root Mean Squared Error for both the entire movement and all steps separately. The second experiment compared the three algorithms during the same exercise in a home-based environment, analysing the effect of the amount of light, the visibility of a background and the type of clothes worn by the subject using mean step length error and percentage of frames with different types of wrong skeleton- and joint recognitions as outcomes.
Results: All skeleton trackers showed a large root mean squared error to Qualisys. Mean Step Length Error was lowest for Cubemos and Openpose and was largest during steps away from the camera. Ambient conditions did not have a significant effect on the mean step length error for any system. Cubemos and Nuitrack showed the lowest percentage of frames with wrong recognitions. In this regard, Cubemos was also not affected by any ambient condition, while Nuitrack showed a higher percentage of frames with wrong recognitions when the light is turned off and no background is visible, and Openpose showed a higher percentage when the lights were turned off versus turned on. For joint recognition, all three skeleton trackers perform better when the lights are turned on, and worn clothing is monochromatic.

Conclusion: Cubemos is the most robust skeleton tracker to changes in a home-based environment. For the best joint recognition, it is advised to turn on lights and wear monochromatic clothing. Even though steps in different directions could easily be identified, all skeleton trackers show a large error to the golden standard and are not deemed suitable for precise location estimations.

Contents

List of acronyms

CNN	Convolutional Neural Network
FoF	Fear of Falling
PAF	Part Affinity Fields
PwS	People with Stroke
SDK	Software Development Kit
TIA	Transient Ischemic Attack

Chapter 1

Introduction

1.1 Motivation

For the 350 000 stroke survivors in the Netherlands, frequent falling is one of the most common medical complications, with even a minor stroke, or Transient Ischemic Attack (TIA), elevating poor balance to be a high-risk factor [1]. Reported fall rates of stroke patients are up to 10 higher than those of their healthy peers, regardless of post-stroke stage. The consequences of falling can be serious, with People with Stroke (PwS) having four times greater odds of hip fractures in the event of a fall. In addition, actual falls or almost falling create a Fear of Falling (FoF), which could lead to avoidance of physical activities. This in turn leads to physical deconditioning, mobility limitation, and then loss of independence [2]. Costs related to rehabilitation and its specialized tools after a fall are high. Thus, prevention of falls and development of FoF for PwS will yield significant health gains for PwS in the form of more independence, and substantial savings in costs and utilization of health care resources.

1.2 Framework

The presented research was done as part of the HEROES Project at the University of Twente, ET Faculty, Department of Biomechanical Engineering (BE). The research was supervised by Aurora Ruiz Rodriguez MSc., dr. Edwin van Asseldonk and dr. Ruben Regterschot.

The HEROES Project aims to reduce fall-related healthcare utilization and associated costs, prevent falls and related injuries, and help PwS maintain independence once they are discharged from the primary care unit. The project aims to achieve this by developing an exergaming (see **Chapter 2.3**) system for home rehabilitation. The exergame will be focused on improving the stepping response of patients. The exergame will use a camera-based skeleton tracking system to identify the position of various joints of the patient. This way, only a specialized depth camera is needed.

Various algorithms exist for skeleton tracking without clear evidence about which is the most accurate in calculating the positions of the lower limbs for the stepping response in a home-based setting. Furthermore, the effect of ambience conditions, like low light, the busyness of the background, and user properties, like types of clothing or skin complexion, on this accuracy have never been quantified. This information is important for the application of camera-based skeleton trackers for rehabilitation in a home-based environment.

1.3 Goal of the Assignment

The goal of this assignment is to compare the accuracy of various camera-based skeleton trackers under various different ambience conditions and a user property. The skeleton trackers to be compared are Cubemos, Nuitrack, and OpenPose. The ambience conditions are brightness of light and visibility of a background. the user property to be tested for is the type of clothing. To this end, the research question is stated as follows: *How do certain ambience conditions and user properties affect the accuracy of different camera-based skeleton trackers?*. To answer this question, the next objectives are defined:

- Through literature research collect and summarize existing knowledge on how stroke affects patients, what the rehabilitation of PwS looks like and what the different types of camera-based skeleton tracking are.
- Set up a protocol, defining diverse ambience conditions and user properties in a home-based environment.
- Performing the experiment according to the protocol to collect a large data set that can then be analysed.
- Process the collected data set to find different types of accuracies for every skeleton tracker, ambience condition, and user property. Outcomes will be the quality of skeleton recognition, mean step length and step position.

1.4 Report Organisation

The remainder of the report is structured as followed. **Chapter 2** is a collection of summarised literature on the subject of stroke, rehabilitation thereof, exergaming, and the relevant skeleton trackers for this assignment. **Chapter 3** is Method and

Materials, describing the methodology of the experiments to measure the accuracy of the various skeleton trackers. **Chapter 4** will show the results of the experiment. Finally, in **Chapter 5** these results will be discussed. In **Chapter 6** conclusions will be presented and recommendations will be given.

Chapter 2

Literature Research

2.1 Clinical Image of Stroke

Stroke is a common, global healthcare problem. In high-income countries, stroke is the third most common cause of mortality. However, most patients survive the initial stroke, causing stroke to be the most common cause of adult-acquired disability. This manifests itself as long-term impairment, limitation in activities, and, as result, reduced participation [3]. In the Netherlands, 1,03% of the entire population suffered a stroke and its consequences [4].

There are multiple causes by which stroke can onset. *Ischemic stroke* occurs when oxygen-supplying blood vessels to the brain become occluded or become more narrow as the result of a buildup of plaque. This, as result, causes a critical decrease in oxygen supply to the brain cells. A *Hemorrhagic Stroke* can occur when a blood vessel in the brain becomes weak and then breaks or leaks. Blood then starts to collect in the brain and this can cause brain cells to die. A TIA is also called a minor stroke and is caused by a temporary clot in a blood vessel. The symptoms are those of a stroke, but the effects last at most a few minutes [5].

Stroke can have various types of long-term impairment. The most common types are lingual, cognitive and motoric decline. The motoric decline causes a limitation in mobility through a decrease in or even loss of muscle control. It is caused by damage to the (pre)motor cortex, motor tracks, or associated pathways in the brain. It usually affects movement of the face, lower and upper extremities of one side of the body, and is observed in about 80% of all stroke survivors. Examples of these impairments include but are not limited to, changes in muscle tone and sensation, muscle weakness and loss of inter-joint coordination, which all culminate in less body control. The decrease in body control also leads to less participation of PwS in life activities and higher dependency on others, causing an increased chance of developing depression [3]. FoF plays a great part in this decrease in participation. About 66 percent of PwS report suffering from FoF. FoF can start due to the PwS suffering from balance impairments, which approximately 83% of acute stroke patients suffer from. Furthermore, approximately 73% of acute stroke patients fall during the first 6 months after their initial stroke, which could raise rehabilitation costs [6]. Therefore there is a large focus on stroke rehabilitation to recover the impaired movement(s) and the associated functions [3], [7].

2.2 Rehabilitation practices for stroke

The goal of stroke rehabilitation is to help PwS relearn skills that have been lost due to the initial attack and limit the effects that this attack had on them. Effective rehabilitation is necessary to help PwS get used to their new level of physiological, physical and social functions. Furthermore, it has been proven that it also reduces the length of hospital stay, their re-admission rate and their use of primary care resources [8]. Reducing the length of hospital stay, re-admission rate and use of primary care resources is important due to the high cost of stroke rehabilitation, adding up to 1,5 billion euros in 2005. This was 2,2% of the total money spent on healthcare in that year. Societal costs per patient in that year were on average €29 484,-. Most of these costs are front-loaded in the first half-year after the initial attack [4].

Studies have been performed on the timing and duration of stroke rehabilitation. The earlier the rehabilitation starts, the better the patient recovers. [9] The duration of stroke rehabilitation depends on the severity. The rehabilitation duration varies between 8 weeks for a mild stroke and 17 weeks for a severe stroke to reach the best functional performance [10].

Stroke rehabilitation is usually referred to in two parts: Early and Later Rehabilitation. Early Rehabilitation is focused on using techniques to influence the potential for neuroplastic change. Later rehabilitation is focused on teaching adaptive responses and coping strategies. [10] There is evidence for improved outcomes if a patient was treated at a specialised stroke unit compared to alternative services. Under these conditions, PwS are more likely to be alive, living at home, and independent one year after a stroke. This is independent of sex, age, stroke severity and type. An increase in length of stay was noticed in specialised wards, but no strong evidence has been found to support this statement. [11]

Physiotherapy is one facet of rehabilitation focused on motoric improvement and is generally valued by patients and strong evidence has been found of its effectiveness, especially for balancing issues. Further evidence was found that task-oriented exercise training was favoured to restore balance and gait, as well as strengthening the lower affected limb. [12]

After some time, patients get discharged from the hospital. This does however not entail that the patient has reached full recovery. Transfer of care may therefore be a more apt description. The moment of transfer of care varies for each patient, mostly based on the amount of support they can receive at home. While rehabilitating from home, the patient stays in contact with rehabilitation services for regular checkups [10]. Earlier transfer of care for people who suffered a moderate stroke improved outcomes (less death and less dependency on others) and also increased patient satisfaction [13]. Early transfer of care also significantly reduces healthcare costs. [4] Creating new options to practice from home for PwS, who had an earlier transfer of care, could therefore be a positive change for patient well-being and societal cost.

2.3 Exergaming

In the past couple of years, the idea of using video games to help people with their rehabilitation process has started to surface. With the coming of Nintendo Wii and Microsoft Kinect, gaming systems that use movement as the main form of input became widely available, with the former using a controller with a motion sensor and the latter a camera-based system. Especially the Microsoft Kinect has not been ignored for research on the use of the system for rehabilitation purposes [8], [14]–[17]. An advantage that camera-based systems have for rehabilitation is the omission of the need for a controller, which could be difficult to hold for PwS and could also cause pain due to repetitive motion [18]. When used at home, a Kinect-based system can encourage the use of upper and lower extremities which may have been affected by the stroke, thus helping in the rehabilitation of the patient. Furthermore, playing non-sedentary games decreases the risks of all-cause mortality [8]. It has been proven that the use of exergames increases the patient's engagement and enjoyment of the rehabilitation while simultaneously improving their motor control [19]. This means that there are several benefits to letting patients practice using full-body, controller-less gaming. The Microsoft Kinect is but one example of a system that allows the tracking of a full human body without added hardware to the person. It is done by the use of camera-based skeleton tracking.

2.4 Skeleton Tracking

Tracking the positions of the joints of the human body is important to let the entire human body by the controller for a game. Skeleton tracking uses one or more cameras to film and analyse the human body for this purpose. As a core concept, the algorithms separate the detected human body from the background of the video and

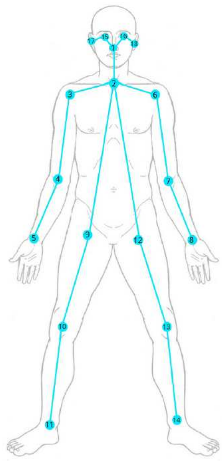


Figure 2.1: Example of the tracked joints by Cubemos.

then identify the position of several joints or features, such as hands, arms, elbows, knees, etcetera [20]. Different soft- and hardware that are able to track the human body are described below. Some of these software are freely available and open-source, while others are only commercially available and thus have less information freely available. Examples of software from both of these categories are explained in more detail in the following paragraphs.

2.4.1 Commercially available software

Cubemos

Cubemos first uses 2D RGB images to track the human body. After having recognised the x and y pixels of every joint on the 2D images, it deprojects these locations using the depth data, if available. To this end, a 3D camera is needed. Cubemos recognises thirteen key points on the human body and five key points on the human face. This can be seen in **Figure 2.1**. The Cubemos algorithm is able to detect up to five people at once. Seeing that Cubemos is not open source, not a lot of documentation on their skeleton tracking algorithm is available.

Nuitrack

Just like Cubemos, Nuitrack uses 2D RGB images to track the human body, after which it uses depth data to deproject the found joint locations to a 3D space. To this end, a 3D camera is needed. Nuitrack can track up to 6 different people at the same time. It uses face tracking to detect the front and the back of a person. Nuitrack allows for the tracking of 20 different joints as shown in **Figure 2.2**. Because Nuitrack

Figure 2.2: Example of the tracked joints by Nuitrack.

is not open source not a lot of documentation on their skeleton tracking algorithm is available.

2.4.2 Open-source software

Microsoft Kinect

To properly track bodies, the Microsoft Kinect makes use of an algorithm of which the focus was to be as robust and computationally efficient as possible. Using a depth image, it works frame by frame and is thus not dependent on temporal data of previous frames. The algorithm segments and labels the image per pixel in expected body parts. The labels have been chosen in such a way they lay close to skeletal joints. The inferred labels get projected back into the world space and from these coordinates, expected coordinates for various joints are estimated. This can also be seen in **Figure 2.3**. From these estimated joints, an estimation of the skeleton can now be made [21]. The newest version of Microsoft Kinect can recognise up to 32 Keypoints and is able to distinguish 6 active players for motion tracking at the same time. More people can be in frame and recognised, but their movements won't be tracked [22]. The Microsoft Kinect software is only able to be used with Microsoft Kinect hardware.

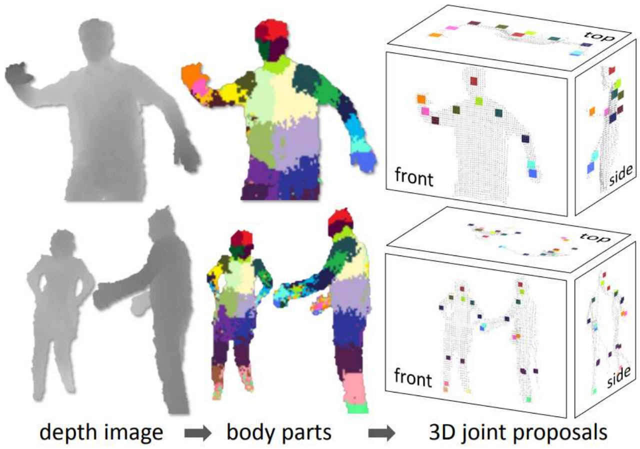


Figure 2.3: Example of Skeleton tracking algorithm that the Microsoft Kinect uses. Adapted from [21].

OpenPose

OpenPose uses 2D images to perform skeleton tracking and is thus able to use any type of webcam to collect video.

Skeleton trackers can either use a top-down or bottom-up approach for detecting multiple people in a frame. The top-down approach first employs a person detector and then uses single-person skeleton tracking on the detected person. This top-down method, however, suffers from early commitment. If a person does not get recognised by the person detector, their skeleton is also not tracked. This tends to happen more often when people are in closer proximity. Furthermore, this top-down approach also suffers from an increase in runtime the more people are visible in the frame. Therefore, OpenPose employs a bottom-up approach for multiple-person skeleton tracking. These types of approaches are attractive due to their robustness to early commitment and potential to decouple runtime from the number of people in the frame. However, since the final parse required a costly global inference, the efficiency of these methods was significantly lower. OpenPose presents an efficient bottom-up approach for multi-person skeleton tracking. Their skeleton tracking algorithm has no limit on the number of people tracked at the same time [23].

OpenPose uses Part Affinity Fields (PAF), a set of 2D Vector flow fields that encode the location and orientation of limbs. First, an image is loaded in, then a Convolutional Neural Network (CNN) is used to predict body parts and PAFs for body part association. The algorithm then uses bipartite matching to associate body parts. The collected data is then assembled into full-body poses and parsed onto the image. This is made visible in **Figure 2.4** [23].

Just like Cubemos and Nuitrack, adding depth data to deproject the found 2D joints allows OpenPose to perform 3D skeleton tracking.

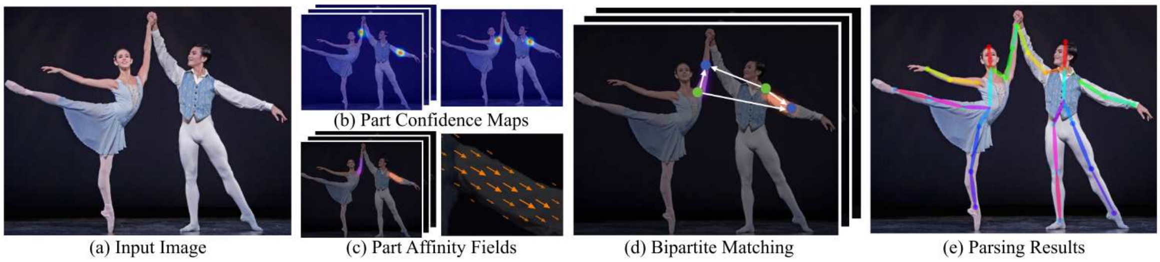


Figure 2.4: Example of Skeleton tracking algorithm that the OpenPose uses. Adapted from [23].

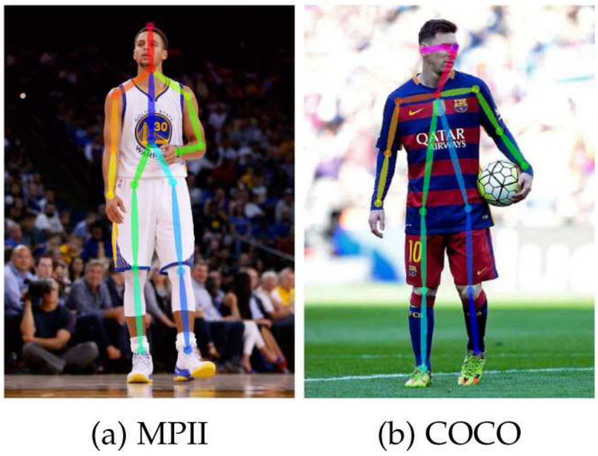


Figure 2.5: Keypoints for different datasets. Adapted from [23].

OpenPose consists of 3 different blocks: body+foot detection, hand detection and face detection. Depending on the goal of the research these can either be turned on or off. Its CNN is trained on the COCO and MPII datasets, both of which detect and display different key points on the human body. This is displayed in **Figure 2.5** The COCO dataset allows for the recognition of 17 Keypoints, while the MPII dataset only allows for 14 [23].

2.4.3 Accuracy of Skeleton Trackers

No information was available about the accuracy of the Cubemos and Nuitrack Skeleton Tracker, because they are only commercially available. The depth accuracy of the Microsoft Kinect varies depending on the distance of the to-be-measured object to the camera. The closer the object is to the camera, the more depth accurate the Kinect becomes. The Kinect has a depth accuracy of $<$2mm right in front of the camera up to 3m away. Starting at a distance of 4m, the accuracy depth accuracy becomes $>$4mm. The further a person moves towards the side of the frame, the worse the accuracy becomes as well [24]. In a multi-camera setup, OpenPose was able to track human movement with an accuracy of 30mm or less. The larger errors that were found in this research were due to the human body not being recognised or wrongly being recognised by the OpenPose software [25]. Both of these tests were performed in artificial, sterile environments. No papers were found discussing the accuracy of the Skeleton Trackers in a home-based environment, nor were any papers available on the accuracy of measuring the stepping response.

2.4.4 Skeleton Trackers and Stroke

Van den Pol, who previously worked for the HEROES Project, used Cubemos to check the effects of lighting, clothing and background objects on the accuracy of the system while the subject performed certain balance recovering stepping motions. He found an indication that the Cubemos skeleton tracker was less accurate in the dark, that background objects make the tracker more accurate, and that wearing multichromatic clothing makes the tracker more accurate. No significance was described, however. [26]

Exergaming applications for stroke rehabilitation using Nuitrack as the skeleton tracker are already in development. Ribé et al. designed an exergame in Unity and made use of the Nuitrack Software Development Kit (SDK), combined with an Intel®Depth Sensor Camera. This application includes 6 exercises for various types of movement. The player has to mimic an avatar displayed on the screen while seeing their own body mirrored on the screen as well. The application also includes ”serious games”, where the exercise is put into context on the screen. An example is a balancing game where the user needs to move their upper body to avoid obstacles that float towards them while grabbing objects that grant them additional points. The application has only been tested on healthy subjects. A clinical trial will be started in the near future [27].

The Microsoft Kinect is one of the most used hard- and software for research into body tracking and rehabilitation, because of its properties as described in **Chapter 2.3**. Various applications have been made and tested for stroke rehabilitation using the Microsoft Kinect SDK. In 2012, a study was performed by Llorens et al. who made stroke patients perform balance exercises using the Microsoft Kinect as skeleton tracker. Balance performance increased significantly for patients between the initial and final assessment [14]. Other research was performed by Proffitt et al. who looked into the feasibility of a home-based exergame stroke rehabilitation program. None of the four participants underwent further therapy at this point. The participants were instructed to play the rehabilitation game Mystic Island [28]. One of the four participants had difficulties understanding the gaming system. The participants enjoyed the fact there were multiple games available. Patients did find it difficult to reach the recommended four hours per week of playtime. Proffitt et al. concluded that Mystic Isle demonstrated to be a feasible option for at-home rehabilitation, but that more research should be performed with a larger sample size [29]. A preliminary study has been performed by Park et al. where they compared a control and intervention group. The control group got half an hour of physical therapy a day, while the intervention group got half an hour of additional exergaming using the Microsoft Kinect. There was a significant increase in performance for the intervention group, indicating the use of the exergaming for stroke rehabilitation could be beneficial to PwS [30].

Research into how well OpenPose is able to track the human gait, which is often affected by stroke, has been performed by A. Viswakumar et al. In this experiment they used the 2D skeleton tracking algorithm of OpenPose to track the gait of healthy subjects by calculating their knee flexion/extension during the movement. The subjects walked parallel to the camera. The measured knee angles were compared and concur well with the normative gait database. [31].

2.4.5 Conclusions

Four different skeleton trackers have been discussed in the previous paragraphs. They all have various characteristics and ways to track human skeletons. Microsoft Kinect is fully 3D image-based, while Cubemos, Nuitrack and OpenPose can work on 2D images, but supports 3D images. Keypoints recognised and their connections vary between all five described skeleton trackers. The number of people that can be recognised at once also varies between the skeleton trackers. This is summarised in **Table 2.1**.

Table 2.1: Summary of Discussed Skeleton Trackers

Skeleton Tracker	3D Camera Needed?	Opensource?	Amount of Keypoints	Max People at once
Microsoft Kinect	Yes (Microsoft Kinect)	Yes	32	6 for motion tracking, Unlimited for amount in frame.
Cubemos	Optional	No	18	5
Nuitrack	Optional	No	20	6
OpenPose	Optional	Yes	COCO Dataset = 17 MPII Dataset = 14	No limit

2.5 Possible complications in skeleton tracking

Because the HEROES Project will be focused on helping patients with at-home rehabilitation, the setting will not always be perfect for the skeleton tracker to reach peak accuracy. Various ambience conditions and user properties could influence the ability of skeleton trackers to properly track the user. Currently, there is a knowledge gap as no research has been performed on the influence of ambience conditions and user properties on accuracy. Various conditions that could influence the ability of the skeleton trackers to accurately track the human body are listed below.

- Lighting conditions
- Types and/or color of clothes worn
- Objects and additional people in the background

After noticing these complications, two experiments will be conducted to test how these complications affect the skeleton trackers and some recommendations for use will be proposed. In **Chapter 3** is described how there will be tested for these possible complications. In **Chapter4** the results of these experiments will be presented. The results will be discussed in **Chapter 5**. In **Chapter 6** the recommendations based on the results of the experiment will be proposed.

Chapter 3

Method and Materials

In this chapter, two experiments will be described to quantify the accuracy of the three skeleton trackers, Cubemos, Nuitrack and OpenPose. The earlier discussed Microsoft Kinect was not tested for in these experiments, since it requires the Microsoft Kinect as a camera, which was not used in the experiments. The first experiment to be described has the goal to compare the three skeleton trackers to a golden standard, Qualisys. Qualisys is a multi-camera marker-based motion capture system with sub-millimeter accuracy [32]. The second described experiment has the goal to quantify the effects of ambient conditions and user properties on the accuracy of the Cubemos, Nuitrack and OpenPose Skeleton tracker. Both experiments are observational cross-sectional studies and were held at the University of Twente. The reason two experiments have been performed is that Qualisys needs a lab setting, while the second experiment needs a (simulated) at-home setting. These experiments will be described in **Chapter 3.3** and **Chapter 3.4** respectively.

3.1 Study Population

For both experiments, healthy subjects were selected from the students and staff of the University of Twente and Saxion Hogeschool. The participants provided informed consent (See **Appendix B**) before starting their respective experiments. The experimental procedures used in these experiments were approved by the Ethics Committee of the University of Twente. The age of the subjects was between 18 and 40 years old. Their heights were between 150cm and 195 cm. For the comparison with Qualisys, one subject was included. For the comparison under different ambience conditions and user properties, thirteen subjects were included.

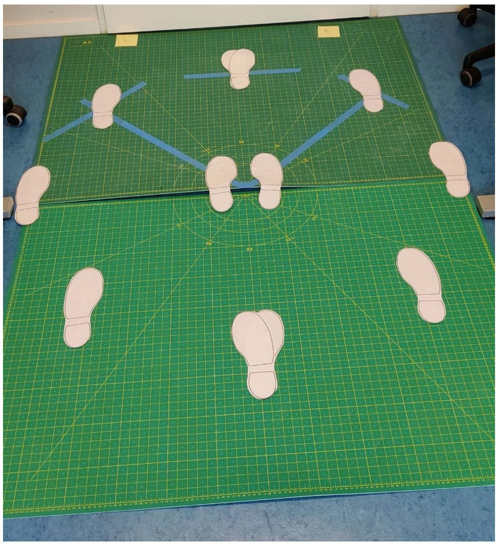


Figure 3.1: Mat with applied markers at the cardinal and intercardinal positions to help subjects perform proper steps. The distance between the middle and the outer markings is 60cm.

3.2 Materials

The following materials were used to perform these experiments:

- Intel Realsense Depth Camera D435i
- Computer with the following software installed:
  - Intel Realsense Viewer v2.38.1
  - MathWorks MATLAB r2020b
  - Cubemos SDK
  - Nuitrack SDK
- Mat with placed markers following **Figure 3.1**
- **Additional Materials for the Comparison with Qualisys experiment**
  - Computer with QTM Qualisys Track Manager installed.
  - Qualisys
- **Additional Materials for the comparison under different ambience conditions experiment**
  - A green screen

3.3 Skeleton Trackers compared to Qualisys

3.3.1 Summary

This experiment has the goal to compare the three skeleton trackers, Cubemos, Nuitrack and OpenPose to a golden standard, Qualisys. An Intel®Realsense™Depth Camera D435i and Qualisys were used to record the participant. Passive markers that the Qualisys cameras can recognise were attached to the two ankle joints of the subject. The participant stood in the center of a marked mat with 9 keypoints (one key point in the middle, the other eight on the cardinals and intercardinals) in front of the 3D Camera and in the center between the Qualisys cameras. Both a 3D video and a Qualisys joint location recording were made of the participant, while they performed the instructed movements. The 3D video recordings were saved and analysed by all three skeleton trackers. After the trail, the ankle joint location data given by the three skeleton trackers were compared to the ankle joint location data given by Qualisys. Root Mean Squared Errors were calculated for every step and the total movement to compare the three systems.

3.3.2 Goal of the Experiment

Collect a 3D video and Qualisys recording of the same stepping movement a subject performs to compare the three skeleton trackers, which use the 3D video to estimate joint locations, to the golden standard Qualisys, which uses 12 cameras to estimate marker locations, which have been added on joint locations. The joints that will be compared to Qualisys markers are the right and left ankle joint.

3.3.3 Hypothesis

We expect that the two commercially available skeleton trackers, Cubemos and Nuitrack have a lower Root Mean Squared Error than the open-source skeleton tracker, OpenPose.

3.3.4 Overview of the Data Collection

The experiment was performed in the Movement Laboratory at the University of Twente as this location is where Qualisys is set up and calibrated.

The subject was asked to perform several stepping motions while being recorded by a 3D camera, the Intel Realsense™Depth Camera D435i, and the Qualisys system. To guide these stepping motions, a mat with markers was made. This mat is visible in **Figure 3.1**. This mat was placed in front of the 3D camera and in the middle

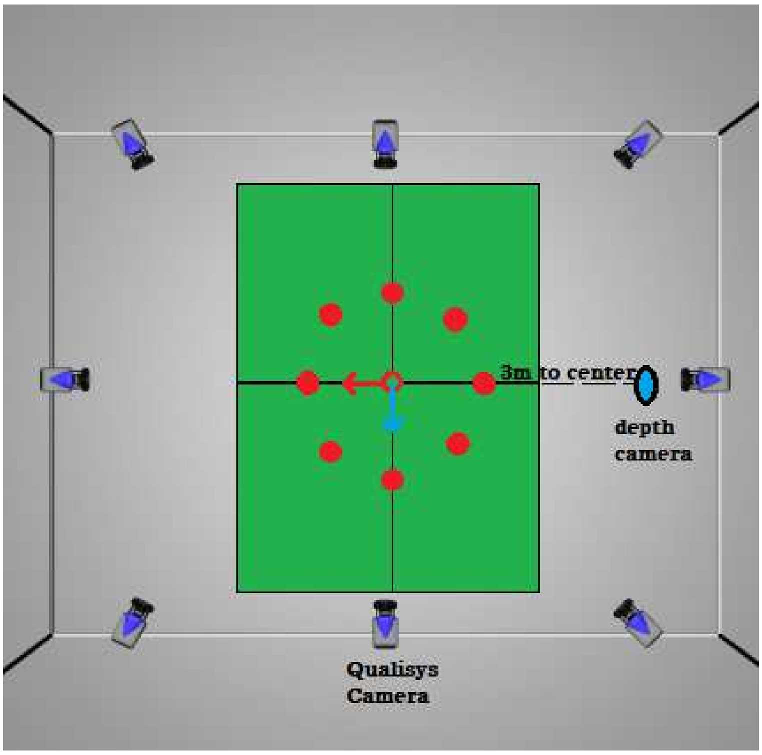


Figure 3.2: Schematic drawing of the mat with markers placed in the middle of the Qualisys. The X and Z axis of the Qualisys are shown in blue and red, respectively. The Y-axis is pointing toward the reader.

of the Qualisys cameras. Qualisys was first calibrated and a coordinate frame was defined using the L frame before the experiment was started. The setup, including the coordinate frame, is shown in **Figure 3.2**.

Reflective markers that can be recognised by Qualisys were added to the subject's body, as shown in **Figure 3.3**. The XYZ-location of the circled marker on both feet was used as a comparison to the XYZ-location of its respective recognised ankle joint by all three systems.

The subject was instructed to take place on the two middle markings on the mat. To synchronise the data between Qualisys and the 3D video recording, the subject performed a right leg raise to knee height, creating a sharp increase for its Y-coordinate. The subject was then instructed to perform ten steps. For every step, the subject was asked to step with their feet to the marker and then stand still in this position for approximately one to two seconds before moving that foot back to its respective middle marker position. This was repeated in every direction. The directions and their order are shortly summarised in **Table 3.1**. The distance between the center of the mat and an end position is 60cm. These motions were captured using both the Qualisys and the Intel Realsense™Depth Camera D435i and their data was saved for further analysis.

Figure 3.3: Image showing the markers added the subject's body for tracking with Qualisys. The circled marker is used as the reference for the skeleton trackers.

3.3.5 Data Management and Analysis

The Qualisys marker location data is saved in a .mat file, which was read into MATLAB for further analysis. The recording software for the 3D camera, Intel Realsense Viewer v2.38.1, saves the 3D video as a .bag-file. For each of the three skeleton trackers, the .bag file was run through their respective Python script, which collected the joint locations and saved these as a .mat file. These .mat files contain the 3D joint locations of every recognised joint on every recognised skeleton of every frame of the video by the chosen skeleton tracker. After all the .mat files were obtained, they were put through another MATLAB script, which analysed the collected data and compared them. The analysis is schematically shown in **Figure 3.4** and further explained in the following paragraphs.

As the Qualisys data has a higher frame rate and a different starting time than the 3D camera, the data first had to be synchronised. To find a corresponding Qualisys frame for every Intel Realsense frame, the following equation is used:

$$I_{Qualisys} = \left(\frac{I_{Intel}}{FPS_{Intel}}\right) \times FPS_{Qualisys} \tag{3.1}$$

Where I_{Intel} is the frame number of the Intel Realsense recording and $I_{Qualisys}$ is the

Table 3.1: Stepping order for all experiments

Step number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Foot	Right					Left				
Direction	Anterior	Right - Anterior	Right	Right- Posterior	Posterior	Posterior	Left- Posterior	Left	Left- Anterior	Anterior

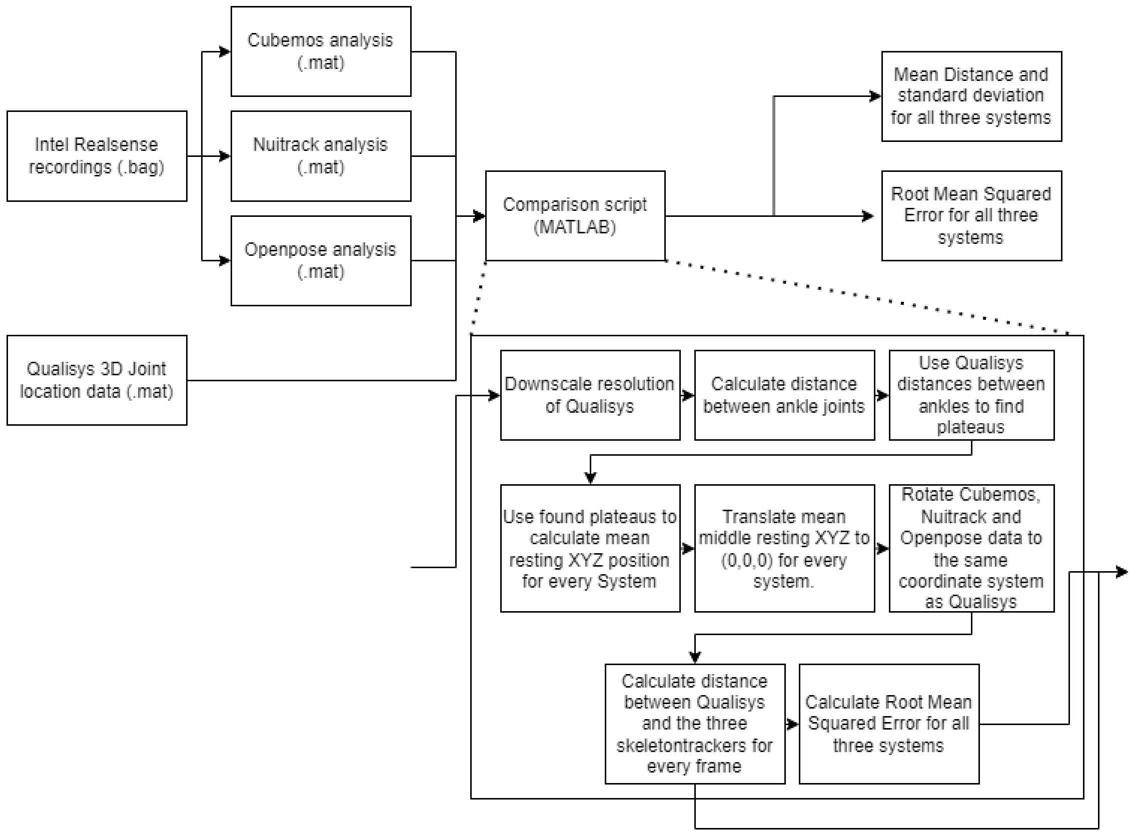


Figure 3.4: Schematic of the data flow after the experiment

corresponding Qualisys frame.

After the Qualisys data has been downscaled, the starting time needs to be synchronised as well. To this end, we did let the subject move their right foot up to knee height. This sharp increase in distance over Y-axis is used to synchronise the starting time. This is made visible in **Figure 3.5**.

After the synchronisation of the frames, the distance between the ankle joints is calculated for the Qualisys data. This data is then displayed in a graph, where the user is asked to select the breakpoints between standing still (platforms) and movement (increases or decreases in the distance). To give an idea of a proper selection, refer to **Figure 3.9**. These breakpoints are also used for the analysis of the data of the three skeleton trackers.

The selected breakpoints are used to find the plateaus where the feet are the closest together. In these frames, the feet are standing still in the middle of the mat. For all four systems, mean X-, Y-, and Z-coordinates were now calculated for the two ankle joints for every frame in this situation. Then an average X-, Y-, and Z- coordinate was calculated for the mean standing still middle position for every system.

Now, to properly compare the X-, Y-, and Z-coordinates of the ankle joints, all

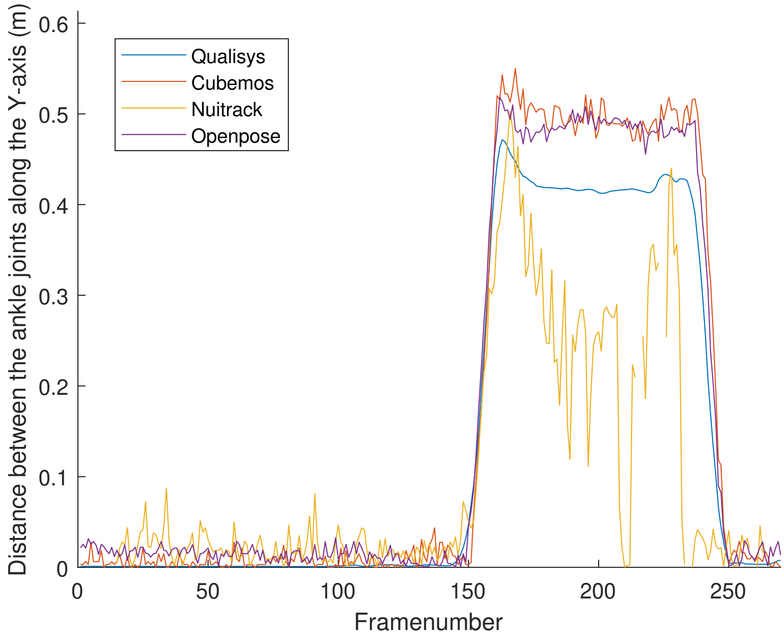


Figure 3.5: The distance between the two ankle joints over the Y-axis for all four systems. Displayed is the sharp increase that exists due to the subject moving their right foot to the knee.

systems first needed to be moved to the same origin. For this reason, the mean standing still middle position was calculated. All data for every system was translated to (0,0,0) by subtracting the mean standing still middle position. Then, a rotation was needed for Cubemos, Nuitrack and OpenPose due to them having a different coordinate system. Their data was first rotated by 180° over the Y-axis, and then rotated by 180° over the X-axis. To correct for possible rotation differences between Qualisys and the 3D camera, an optimisation has been performed, calculating a rotation around the Y-axis, minimising the total RMSE between all three systems and Qualisys.

3.3.6 Outcomes

At last, now all the systems are in the same coordinate system, the distance between the measured location of the ankle joints with Qualisys and the measured location of the ankle joints of the three skeleton trackers was calculated for every frame. For both ankle joints, Root Mean Square Errors were calculated, for both the entire movement and every step separately.

3.4 Accuracy of skeleton trackers under different ambient conditions

3.4.1 Experiment Summary

To measure the effects of ambient conditions and user properties on the accuracy of the Cubemos, Nuitrack and OpenPose Skeleton tracker, the following experiment has been performed. Participants were chosen from students and personnel from the University of Twente and Saxion Hogeschool. An Intel Realsense™Depth Camera D435i was to record the participants. Depending on what variables were tested for, the setting was changed. The ambient conditions and user property that were tested for are Light (ON/OFF), Background (Visible/no background) and Clothing (multichromatic/monochromatic). The participants stood in the center of a marked mat with 9 key points (one key point in the middle, the other eight on the cardinals and intercardinals) in front of the camera. 3D videos were made of the participant. While being recorded the participants performed the movements as instructed. These videos were saved and analysed by all three skeleton trackers. After the trial, the data from the skeleton tracker was used to analyse the accuracy of the measurements. Outcomes are mean error to expected step length, percentage of frames with wrong recognitions (Missing joints/skeleton, recognising additional skeletons) and mean step location for every step.

3.4.2 Goal of the Experiment

Collect a large dataset of 3D videos of subjects performing stepping motions under different ambient conditions (Light, Background) and a user property (Clothing). This data set was then used to analyse the accuracy of the three skeleton trackers, Cubemos, Nuitrack and OpenPose.

3.4.3 Hypotheses

We expect that the low light condition will negatively affect the accuracy of the skeleton trackers because skeleton trackers are image-based. Having objects in the background could affect the ability of the skeleton tracker to recognise the human pose. Previous work in the HEROES Project by Van den Pol [26] indicates background objects may have a positive effect on the accuracy of the skeleton tracker. This was tested on the Cubemos skeleton tracker. Van den Pol also found indication that wearing multichromatic (casual) clothing may positively affect the accuracy of

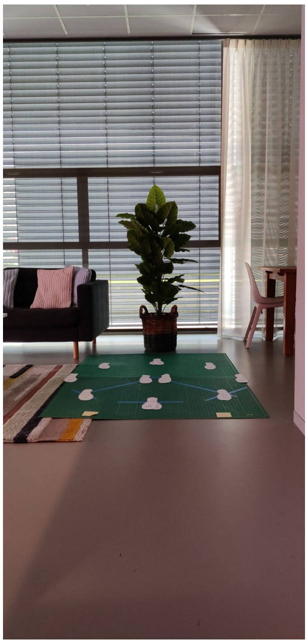


Figure 3.6: The background in the eHealth house when it is not blocked off by the green screen.

the skeleton tracker in comparison to monochromatic (black) clothing. However, no statistical significance was found due to the small sample size.

3.4.4 Overview of the Data Collection

The experiment was performed in the eHealth House at the University of Twente as this location simulates an at-home setting and allows better control of ambient conditions (e.g. light).

The subject was instructed to perform several stepping motions while being recorded by a 3D camera, the Intel Realsense™Depth Camera D435i. The same stepping motions and mat from the previous experiment were also used in this experiment. This mat was placed in front of the 3D camera. Behind the mat, either the background was made visible or blocked off by a green screen. The background without the greenscreen is visible in **Figure 3.6**. To better control the amount of light during the experiment, all blinds were closed in the eHealth house. This way, the light intensity could be controlled by using the light switches in the room. Depending if a particular recording required mono- or multichromatic clothing, the subject was asked to change. As every variable has a total of 2 settings, there are a total of eight combinations of settings. These are made visible in **Appendix A**. Every combination of settings has been recorded thrice per subject. The setup is visible in **Figure 3.7**.

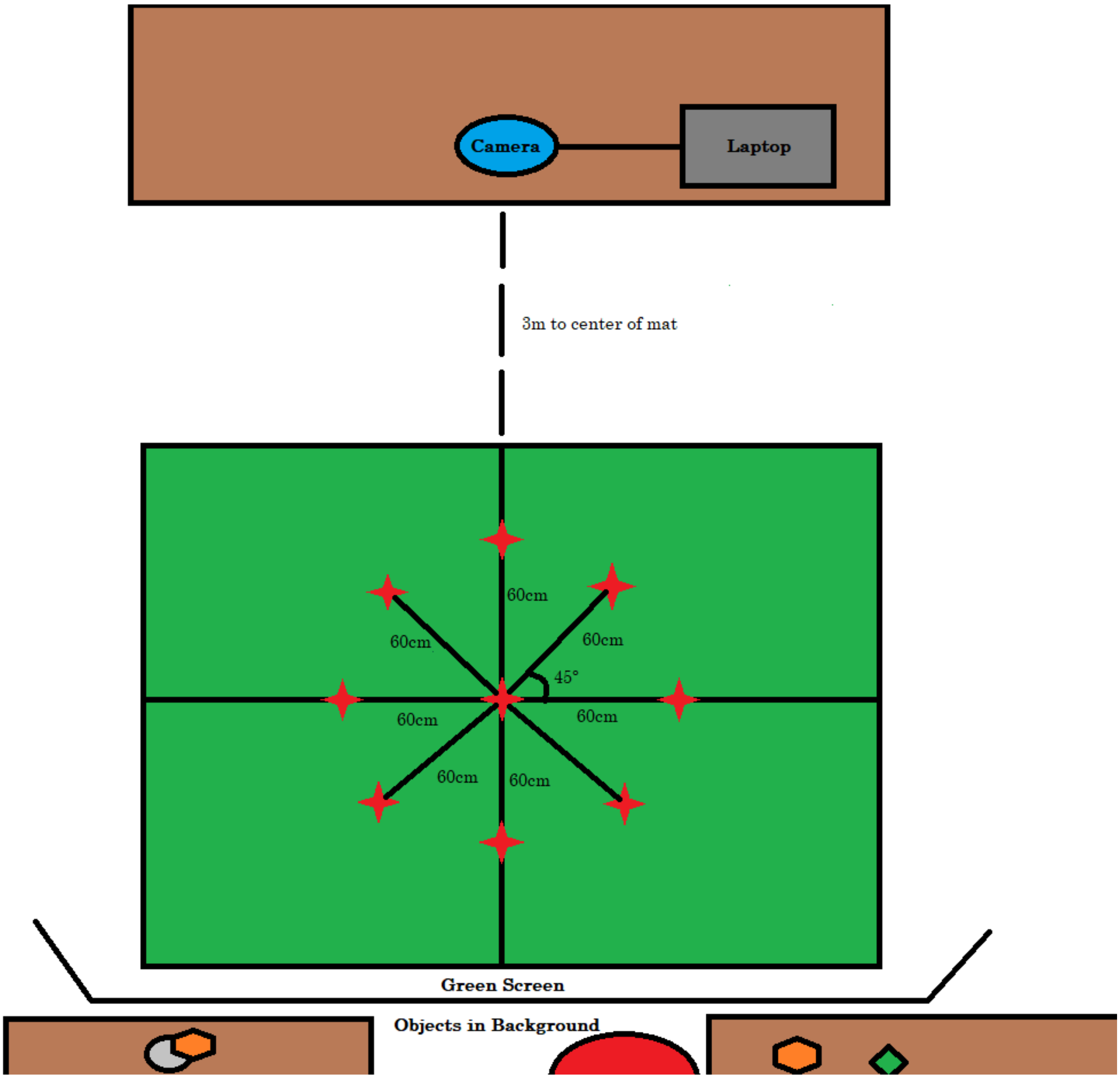


Figure 3.7: Schematic drawing of the objects and distances during the experiment

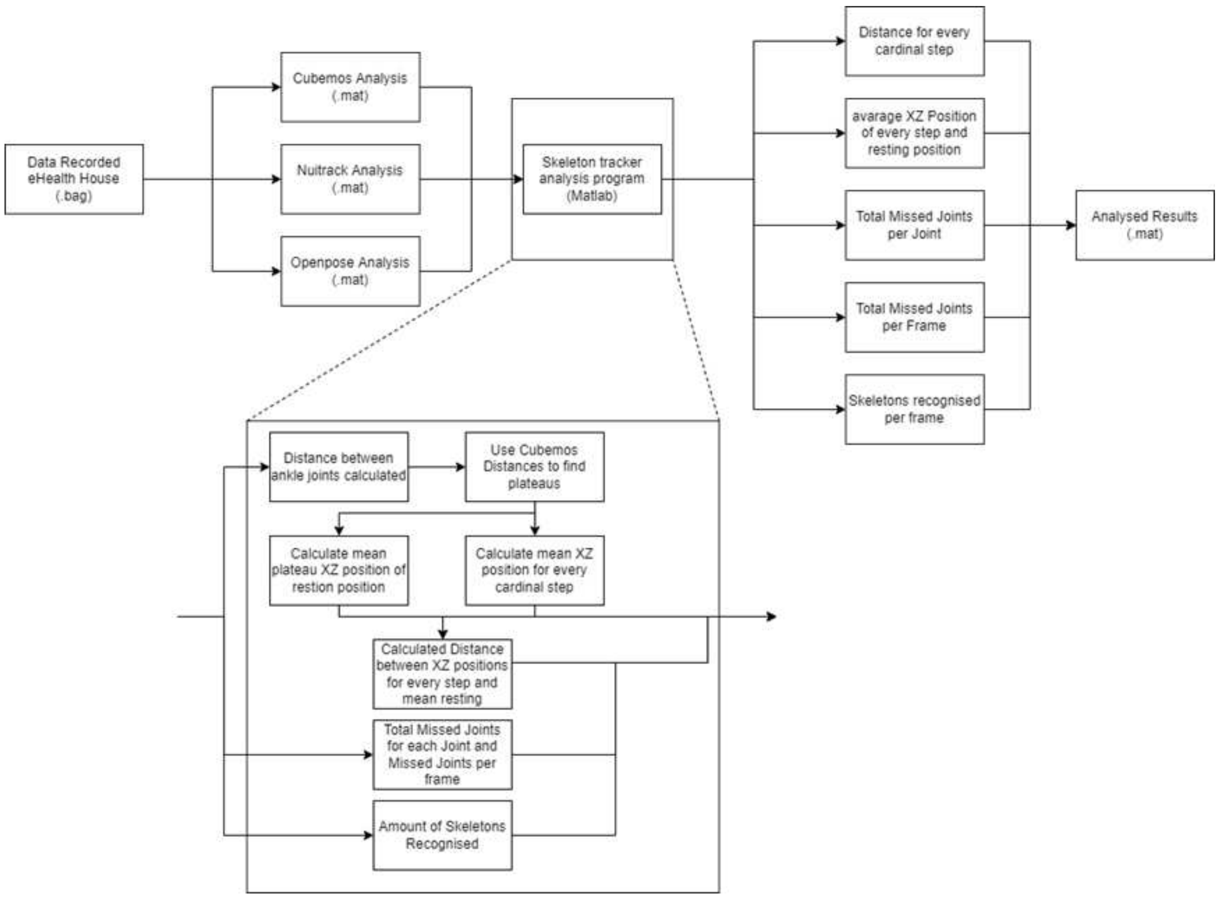


Figure 3.8: Schematic of the data flow after the experiment

3.4.5 Data Management and Analysis

The experiments were recorded using Intel Realsense Viewer v2.38.1, which saved the 3D video as a .bag-file. For each of the three skeleton trackers, the .bag file was then run through their respective Python script, which collected the joint locations and saved these as a .mat file. These .mat files contain the 3D joint locations of every recognised joint on every recognised skeleton of every frame of the video by the chosen skeleton tracker. Then, these .mat files were put through a MATLAB Program, which analysed the collected data in these files. The flow of data is shortly summarised in **Figure 3.8** and further explained and shown in the following paragraphs:

MATLAB Analysis:

To properly analyse and use the raw joint data that the three skeleton trackers have as output, a MATLAB application was written. Screenshots from this application can be found in **Appendix D**. After the three generated .mat files from the skeleton trackers have been loaded in on **Figure D.1**, the next screen, visible in **Figure D.2**,

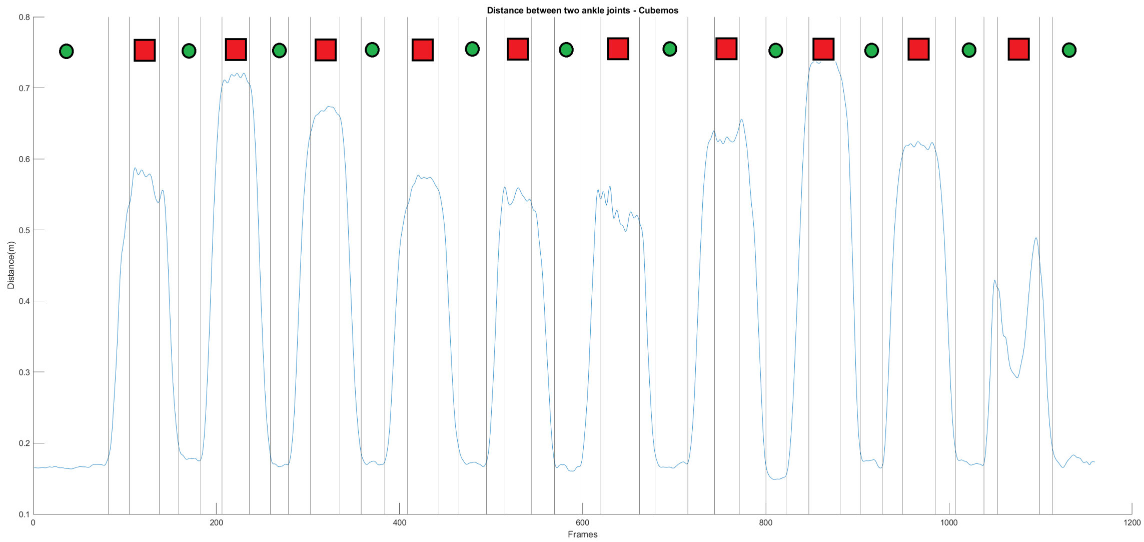


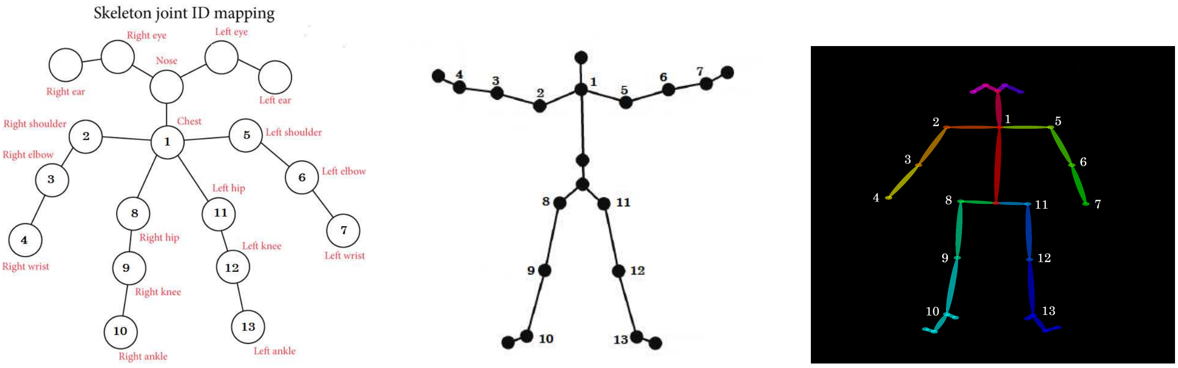
Figure 3.9: Automatic plateau detection with markers. The red markers indicate the segment where a foot is in an end position, the green markers indicate where both feet are close together on the middle markings.

can be used to get a general idea of the joint locations of both ankles through time. They are plotted in the XZ-plane, seen from above.

To be able to calculate the mean end positions of a step, the distance between the two ankle joints for every frame is calculated and plotted. An algorithm has been written to automatically recognise transitions from standing still, a plateau in the graph, and movement, an in- or decrease in the graph. This algorithm uses a maximum speed and a minimum length of a section as input variables. This is visible in **Figure D.3**. Additional transitions can be added by clicking on the graph if the algorithm did not function properly. The Cubemos dataset is used to recognise the plateaus, due to (usually) having the least amount of noise, which makes the recognition of plateaus easier for the algorithm and the user alike. After this segmentation, the data is divided into three different types of segments: One where the foot is on one of the (inter)cardinal markers, one where both feet are in the middle, and one where one of the feet is moving. This is displayed in **Figure 3.9**. The selected transitions are also made visible on Nuitrack's and OpenPose's data. (**FigureD.4**) Their calculated distance between ankle joints is plotted here together with the selected transitions. The choice was made to not recognise the transitions for every skeleton tracker separately for two reasons: First, the Nuitrack data usually contains so much noise that separate selection is difficult for both the algorithm and human eye alike and second, this way, the same frames for transitions are chosen for all skeleton trackers, which allows for a more fair comparison. Due to an observed characteristic of Nuitrack, where it takes a varying amount of frames to start recognising a skeleton, its transitions are shifted by the automatically recognised amount

Table 3.2: All joints, made visible in **Figure 3.10**, numbered and named.

Joint number	Joint name
1	Chest
2	Right Shoulder
3	Right Elbow
4	Right Wrist
5	Left Shoulder
6	Left Elbow
7	Left Wrist
8	Right Hip
9	Right Knee
10	Right Ankle
11	Left Hip
12	Left Knee
13	Left Ankle



(a) Cubemos, adapted from [33]
(b) Nuitrack, adapted from [34]
(c) OpenPose, adapted from [35]

Figure 3.10: Selected Joints numbered for their respective skeleton trackers. Not numbered joints are not used for calculating missing joints. The names of the numbered joints can be found in **Table 3.2**

of ”startup-frames” to properly align the data.

This segmented data can be used to calculate the mean end X- and Z-location of every step. The mean middle position can also be calculated. For every frame where both feet are on the middle markings on the mat (the segments with the green markings in **Figure 3.9**), the mean X- and Z-position of the ankle joints of both feet are calculated. A total average is then taken over all the frames, and an average X- and Z- location for the middle of the mat is calculated. This is displayed in **Figure D.5** with the green marking. The mean end positions of every step are calculated by taking the X- and Z-positions of the moving ankle of every frame in a segment, that is marked red in **Figure 3.9**, and averaging these values.

In **Figure D.6** and **D.7**, the screens are shown displaying the frames on which joints were not recognised, on which additional skeletons were recognised, or on which zero skeletons were recognised. Since the three skeleton tracker recognise joints in different locations on a person, a selection of joints has been made which are going to be used for comparison. Even though the nose joint gets recognised by every skeleton tracker, the choice has been made to omit it, because some subjects turned their heads away to see where they had to step, making the non-recognition of the nose joint proper. The selected joints for each skeleton tracker are made visible in **Figure 3.10**.

Finally, a bar graph is shown with the mean distance of every end position of a step compared to the earlier estimated middle of the mat. The standard deviation of this distance is also included. Instead of using the distance between the ankle joints as step distance, as is done in **Figure D.3 and D.4**, now the distance between the calculated middle position and the moving ankle joint is used to correct for possible misrecognition of the nonmoving ankle joint during the step. This is displayed for all three skeleton trackers, see **Figure D.8**.

Outcomes

The distance between the mean end-of-step position of every step to the calculated middle position is calculated for every video. The error between this value and the expected 0,60m is calculated. These results will be used to calculate mean errors for every step over all videos. A polar plot will be made to visualise the mean end position and variance of these positions for every single system.

Further calculations will be performed to see how proper the skeleton recognition is. For every video, the percentage of frames where a joint was not recognised, a skeleton was not recognised, or an additional skeleton was recognised is calculated. These are considered wrong recognitions. Also, the percentage of frames where either an additional skeleton is recognised, or a joint, or a skeleton is missed, is calculated.

Mean step length error and all skeleton/joint recognition outcomes will be compared against one another for which skeleton tracker (System) was used(1), if the light was on or off (2), if the subject was wearing mono- or multichromatic clothing (3) and if the background was or was not blocked off by the green screen(4). Factorial Repeated Measures ANOVAs are used to test for the significance of these results.

Chapter 4

Results

Two following chapter is divided into two subchapters, each showing the results of their respective experiment. **Chapter 4.1** shows the results of the experiment that compares the camera-based skeleton trackers to Qualisys. **Chapter 4.2** shows the results of the experiment that compares the accuracy of the camera-based skeleton trackers under different ambient conditions. All results are considered significant for p-values below 0.05. All bar graphs include the standard deviations of the collected data.

4.1 Skeleton Trackers compared to Qualisys

To be able to properly compare the three 3D camera-based skeleton trackers to the more accurate golden standard Qualisys, the coordinates of the skeleton trackers were rotated and translated as discussed in the **Chapter 3.3**. Furthermore, to minimise the total RMSE error of all systems compared to Qualisys, the coordinates for the three skeleton trackers were rotated an additional 0.3 degrees around the Y axis. Three XZ-plots, one for every skeleton tracker, have been made, displaying the final trajectories of both ankles over this plane, comparing them to the trajectory of Qualisys. This is visible in **Figure 4.1**.

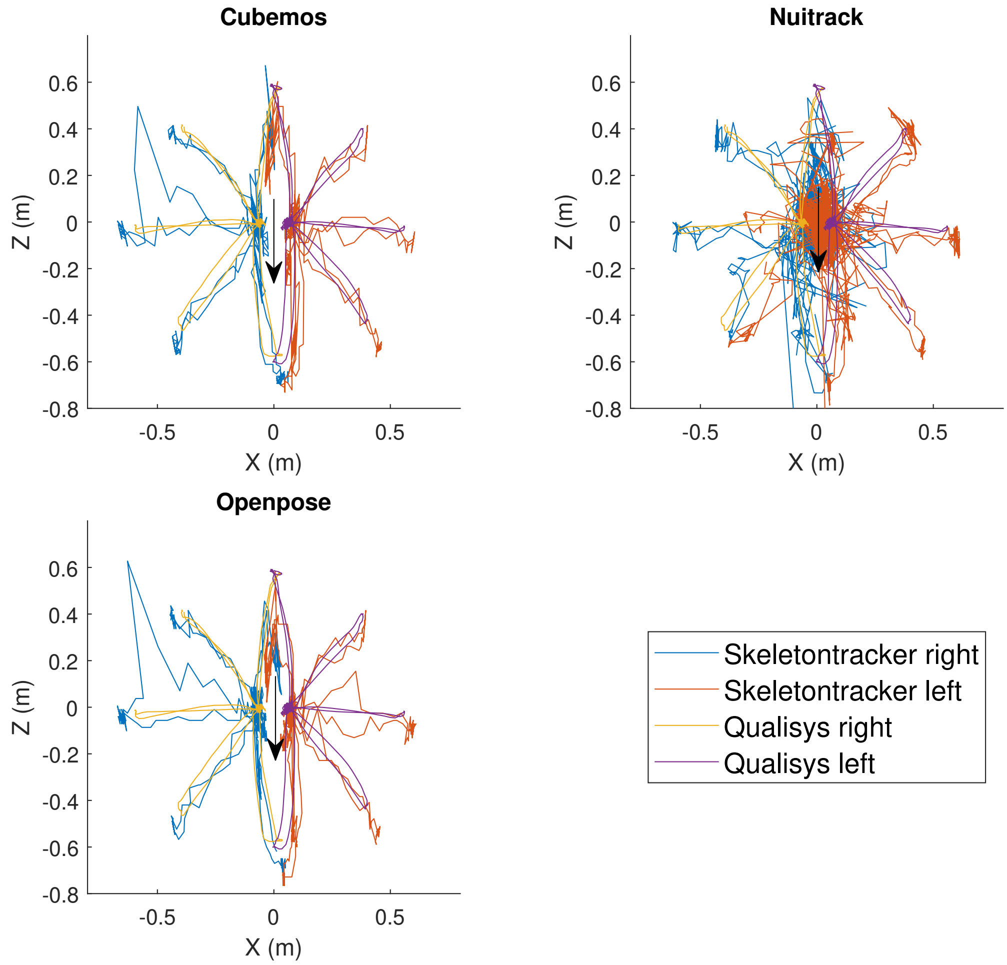


Figure 4.1: XZ plots of both ankles for all three skeleton trackers compared to Qualisys. Subject is looking in the direction of the arrow.

Visible is that all steps can be easily recognised by Cubemos and OpenPose, and these follow the trajectory of Qualisys well. Usually a overestimation of stepping distance is visible for these two systems. Nuitracks data contains way more noise in comparison to Cubemos and OpenPose. The steps towards the Left, Right, Left Anterior, Left Posterior, Right Anterior and Right Posterior marker can still be distinguished, but even these contain a lot of noise in its movement. To quantify the error between the recognised ankle coordinates of all skeleton trackers and the recognised marker coordinate of Qualisys RMSEs have been calculated. This was done both for every step separately and for the entire movement combined. These RMSEs are visible in **Table 4.2**.

Table 4.1: Root Mean Square Error for every system, ankle, and step The unit is meters.

Step		**Cubemos**		**Nuitrack**		**OpenPose**	
		Left	**Right**	**Left**	**Right**	**Left**	**Right**
Right leg moving	**Legraise**	0.033	0.162	0.210	0.712	0.032	0.132
	Anterior	0.167	0.145	0.249	0.216	0.181	0.155
	Anterior Right	0.043	0.164	0.097	0.247	0.045	0.172
	Right	0.046	0.134	0.081	0.223	0.044	0.144
	Right Posterior	0.039	0.119	0.091	0.238	0.045	0.122
	Posterior	0.053	0.260	0.151	0.334	0.044	0.349
Left leg moving	**Posterior**	0.201	0.057	0.370	0.085	0.338	0.049
	Left Posterior	0.095	0.042	0.289	0.104	0.110	0.046
	Left	0.165	0.050	0.368	0.173	0.199	0.052
	Left Anterior	0.123	0.051	0.521	0.534	0.141	0.049
	Anterior	0.155	0.405	0.360	0.302	0.157	0.426
RMSE over entire video		**0.086**	**0.134**	**0.262**	**0.296**	**0.109**	**0.156**

Table 4.2: Root Mean Square Error (m) for every system, ankle, and step

	Cubemos		**Nuitrack**		**Openpose**	
	Left	**Right**	**Left**	**Right**	**Left**	**Right**
RMSE over entire video	**0.086**	**0.134**	**0.262**	**0.296**	**0.086**	**0.124**

The average RMSE for the non moving leg is lower than that of the moving leg. For Cubemos and OpenPose, the RMSE of the moving leg towards the posterior marker for both legs is always the highest. A very large error can also be seen for these systems for the Right Leg, when the left leg is moving towards the Anterior marker. Taking a look at **Figure 4.1**, it is visible that somewhere during the movement the right leg is recognised approximately half way towards the posterior marker. In **Figure 4.2** is shown that this decrease over the Z-axis happens during this motion. Qualisys is recognised to be in the right location, the middle of the mat, while the skeleton trackers recognise the foot more forward. This also explains the higher RMSE for the right foot over the entire video for these two systems.

Nuitrack continuously has higher RMSEs than the other two skeleton trackers, which can be explained by its visibly more noisy data. Its largest RMSEs are during the leg raise for the Right ankle and the Left anterior movement for the Left foot.

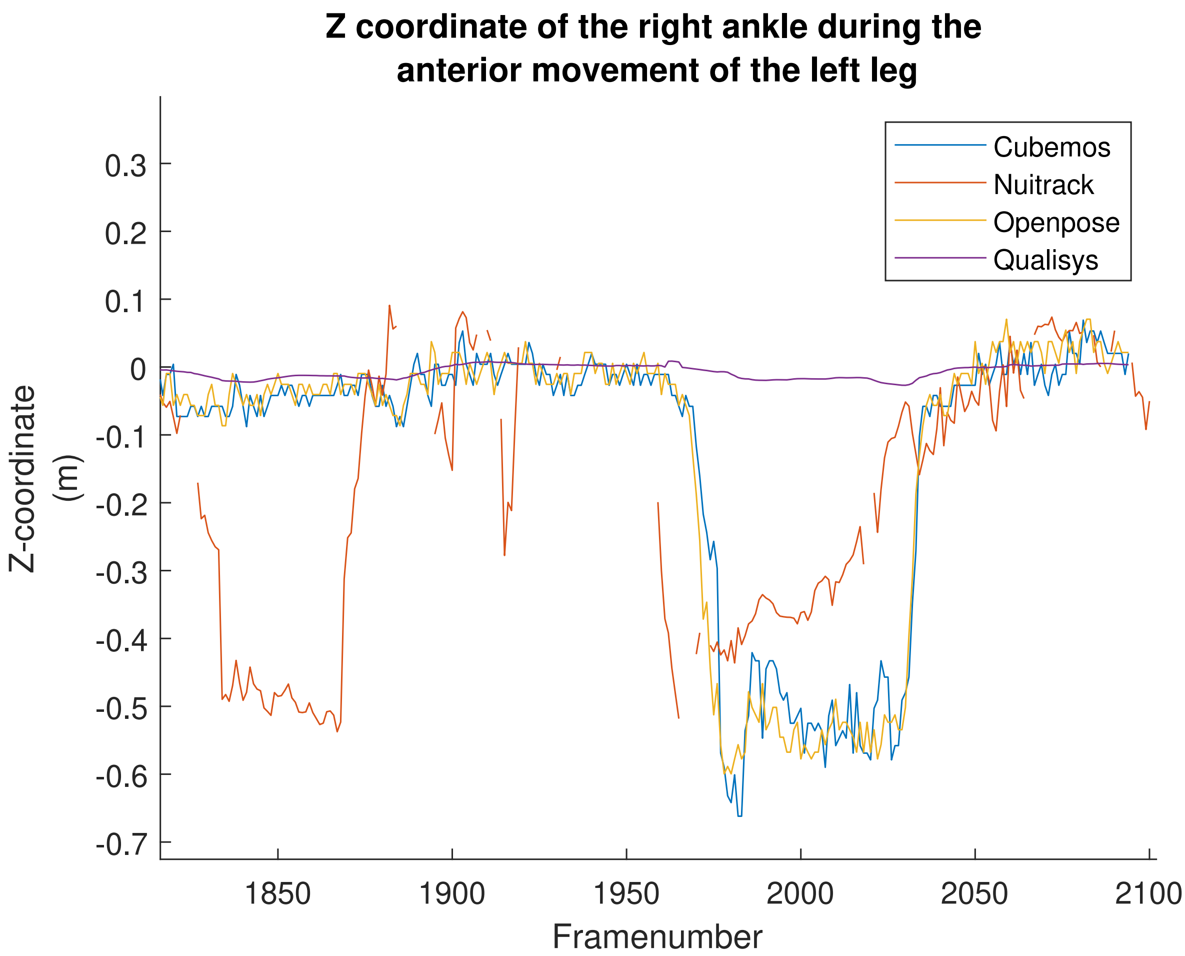


Figure 4.2: Z-coordinate of the right leg recognised by all three skeleton trackers and Qualisys during the anterior movement of left leg. Gaps in the data happen due to the respective skeleton tracker not recognising the joint during those frames.

4.2 Accuracy of skeleton trackers under different ambient conditions

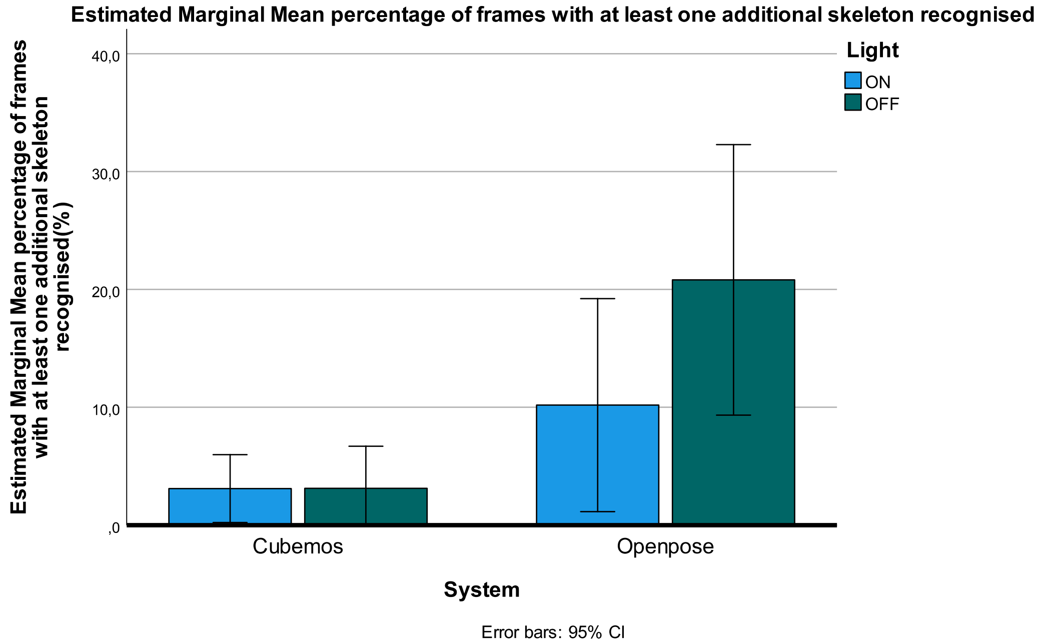


Figure 4.3: The effect of Lights on the mean percentage of frames where at least one additional skeleton was recognised by Cubemos and OpenPose

Due to a technical error in the analysis program, one of the thirteen subjects was not able to be analysed properly and has thus been excluded from this experiment.

During the analysis of the results, normality was checked for. For mean step length error, the data was normally distributed. For percentage of frames with missing skeletons, extra skeletons, missing joints and total wrong recognitions, the data was strongly left skewed. Several transformations have been tried on this data to make it more normally distributed, but none had the desired effect. As Repeated Measures ANOVA is quite robust to the violation of normality the choice has been made to still use it to quantify the significance of the results.

4.2.1 Wrong recognitions: Missing and Additional Skeletons

During the analysis of the percentage of frames where skeletons were either missed or the percentage of frames where additional skeletons were recognised, only Nuitrack sometimes did not recognise a skeleton. Both other skeleton trackers did recognise at least one skeleton on every frame. Nuitrack never recognises more than one skeleton, while the other two skeleton trackers do wrongly recognise addi-

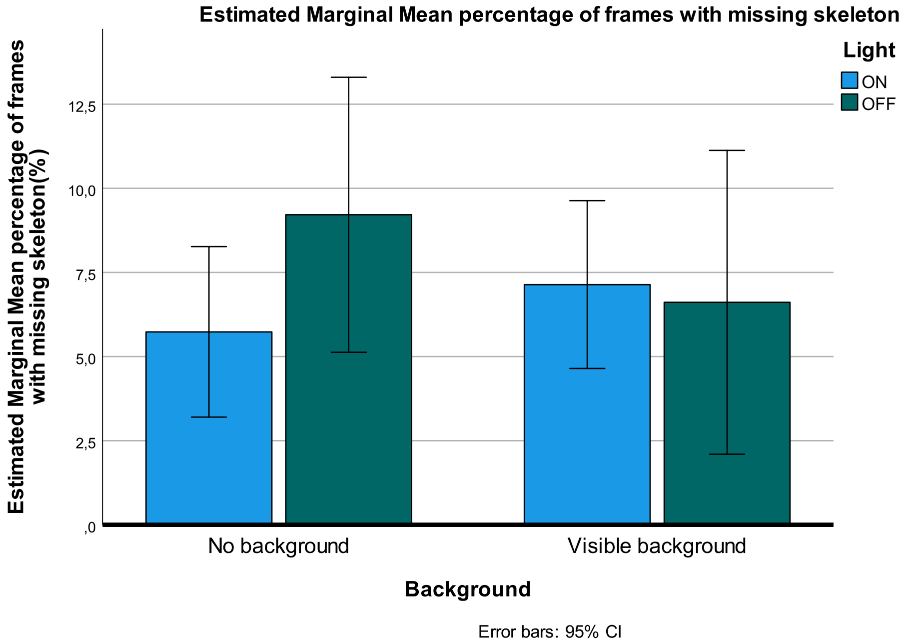


Figure 4.4: The effect of Light on the mean percentage of frames where skeletons were not recognised by Nuitrack when there is a visible or no background.

tional skeletons. For this reason, Cubemos and OpenPose will not be analysed for missing skeletons, and Nuitrack will not be analysed for extra recognised skeletons.

Extra Skeletons: A four-way repeated-measures ANOVA was conducted on the influence of four independent variables (System, Lights, Clothing, Background) on the percentage of frames where at least one more skeleton was recognised by the respective skeleton tracker. A significant interaction was found between Lights and System ($F(1, 11) = 10.780, p = .007$). A significant effect for System was found ($F(1, 11) = 10.018, p = .009$), indicating a significantly higher percentage of frames where an extra skeleton is wrongly recognised by OpenPose ($M = 15.498\%, SD = 4.471$) in comparison to Cubemos ($M = 3.113\%, SD = 1.216$).

Post hoc analyses were performed. No main effects were found for Cubemos that influenced its percentage of frames where additional skeletons were recognised.

For OpenPose, a significant effect was found for Lights ($F(1, 11) = 13.785, p = .003$), indicating a significant increase in the percentage of frames where at least one additional skeleton is wrongly recognised when the Lights are turned off ($M = 20.811\%, SD = 5.217$) versus when the Lights are turned on ($M = 10.185\%, SD = 4.106$). No further significant effects were found for OpenPose. The effect of Light on both Cubemos and OpenPose is made visible in **Figure 4.3**.

Missing Skeletons: A three-way Repeated Measures ANOVA was conducted on the influence of three independent variables (Lights, Clothing, Background) on the percentage of frames where no skeletons were recognised by Nuitrack. Only a significant interaction was found between Lights and Background, $F(1,11) = 6.920, p = .023$. During post hoc analysis, the main effect of Light showed an F ratio of $F(1,11) = 9.288, p = .011$ for when there is no background, indicating a significant increase in the percentage of frames with a missed skeleton between lights on $(M = 5.733\%, SD = 1.151)$ and lights off $(M = 9.213\%, SD = 1.856)$. No significance was found for Lights when there is a visible background. This is made visible in **Figure 4.4**

4.2.2 Wrong recognitions: Missing Joints

A four-way repeated-measures ANOVA was conducted on the influence of four independent variables (System, Light, Clothing, Background) on the percentage of frames where at least one joint was not recognised. No significant interactions were found between the four main variables. The main effect of System showed an F ratio of $F(2,22) = 5.308, p = .013$, indicating a significant difference between the three skeleton trackers. Nuitrack ($M = 0.388\%, SD = 0.088$) showed a significant lower percentage of frames with at least one missed joint than OpenPose($M = 2.789\%, SD = 0.671$) and Cubemos($M = 3.752\%, SD = 1.263$), which both don't differ significantly from each other. The main effect of Light showed an F ratio of $F(1,11) = 6.006, p = 0.032$, indicating a significantly higher percentage of frames with at least one missed joint when the lights are turned off ($M = 3.144\%, SD = 0.694$) compared to when the lights are turned on ($M = 1.476\%, SD = 0.603$). The main effect of Clothing showed an F ratio of $F(1,11) = 5.338, p = .041$, indicating a significantly higher percentage of frames with at least one missed joint when the subject is wearing multichromatic clothing ($M = 2.978\%, SD = 0.796$) in comparison to monochromatic clothing ($M = 1.641\%, SD = 0.38$). This is shown in **Figure 4.5** and **Figure 4.6**.

4.2.3 Wrong recognitions: Total amount of frames with wrong recognitions

A four-way repeated-measures ANOVA was conducted on the influence of four independent variables (System, Light, Clothing, Background) on the percentage of frames where there was a wrong recognition, i.e. either a joint was not recognised, a skeleton was not recognised or an additional skeleton was recognised. A significant interaction was found between System and Light ($F(2,22) = 9.843, p < .001$). This interaction is shown in **Figure 4.7**.

Post hoc tests were performed. For Cubemos, no significant interactions exist.

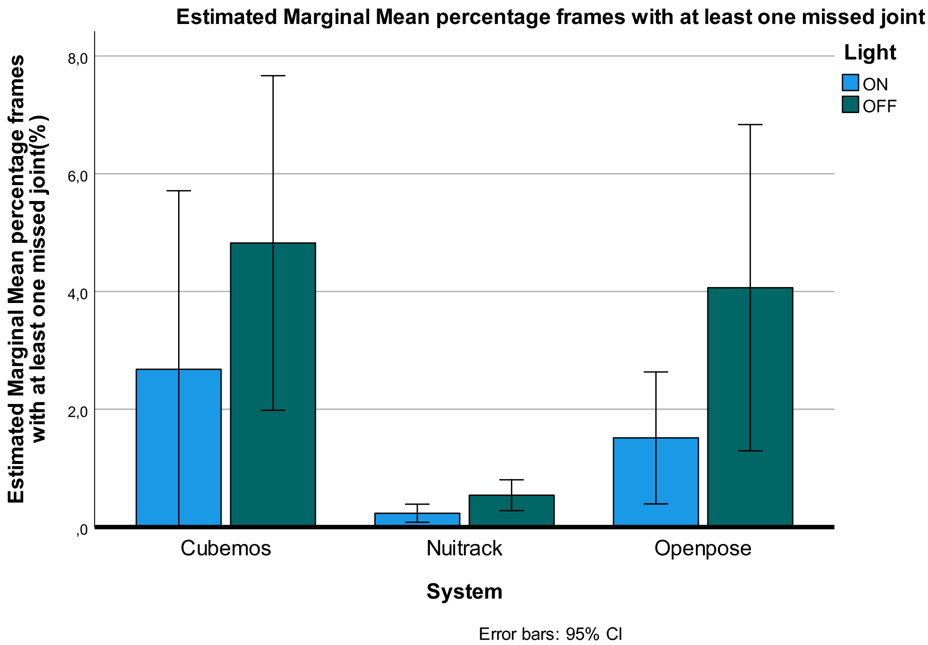


Figure 4.5: The effect of Light on the mean percentage of frames where at least one joint was not recognised for all skeleton trackers.

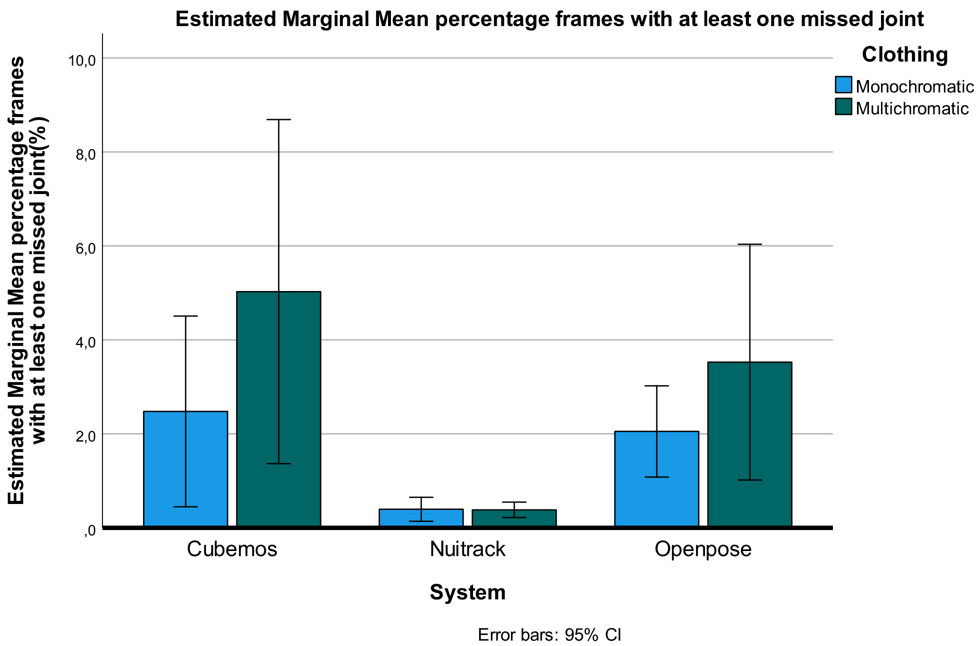


Figure 4.6: The effect of Clothing on the mean percentage of frames where at least one joint was not recognised for all skeleton trackers.

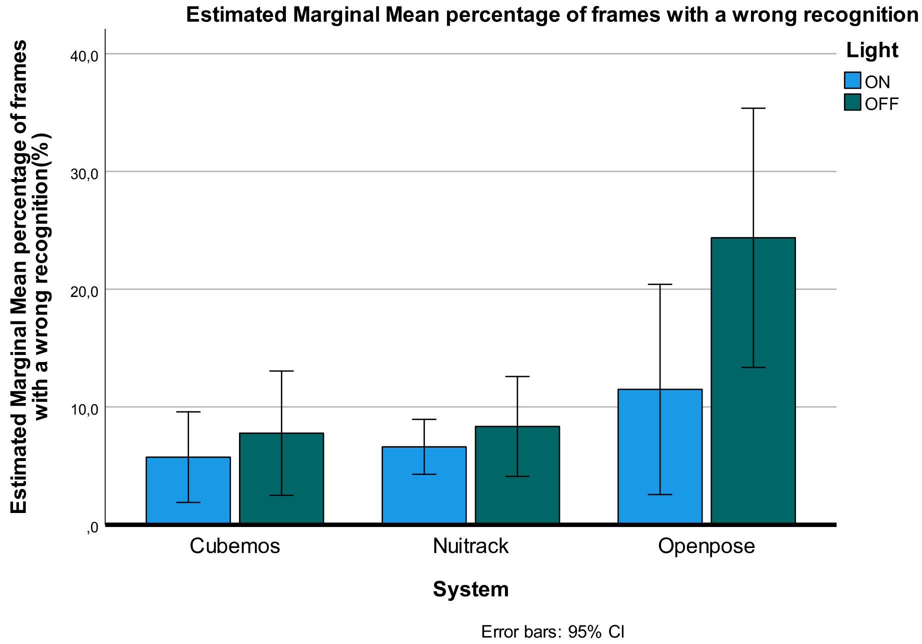


Figure 4.7: The effect of Light on the mean percentage of frames where a wrong recognition was made by all skeleton trackers.

Furthermore, none of the three main factors (Light, Clothing, Background) did indicate a significant effect on the percentage of frames with wrong recognitions.

For Nuitrack, a significant interaction was found between Lights and Background with an F ratio of $F(1, 11) = 8.243, p = .015$. This interaction is made visible in **Figure 4.8**. When the lights are turned on, no significant effects of Clothing or Background are found. When the lights are turned off, a significant effect for Background is discovered ($F(1, 11) = 5.290, p = .042$) indicating a significant increase in the percentage of frames with wrong recognitions when the background is not visible ($M = 9.723\%, SD = 1.902$) in comparison with when the background is not visible($M = 6.970\%, SD = 2.131$).

For OpenPose, the main effect of Light was significant ($F(1, 11) = 21.425, p < .001$) indicating a difference between lights on ($M = 11.490\%, SD = 4.055$) and lights off($M = 24.368, SD = 4.999$) for the percentage of frames where a wrong recognition was made, with Lights off having the higher percentage. This is made visible in **Figure 4.7**.

System showed to have a significant effect on the percentage of frames with wrong recognitions $F(1.193, 13.120) = 5.696, p = .028$. A post hoc test showed that OpenPose ($M = 17.929\%, SD = 4.334$) had a significant higher percentage of frames with wrong recognitions compared to Cubemos ($M = 6.756\%, SD = 1.840$) with and Nuitrack ($M = 7.483\%, SD = 1.408$) both. Cubemos and Nuitrack did not differ significantly.

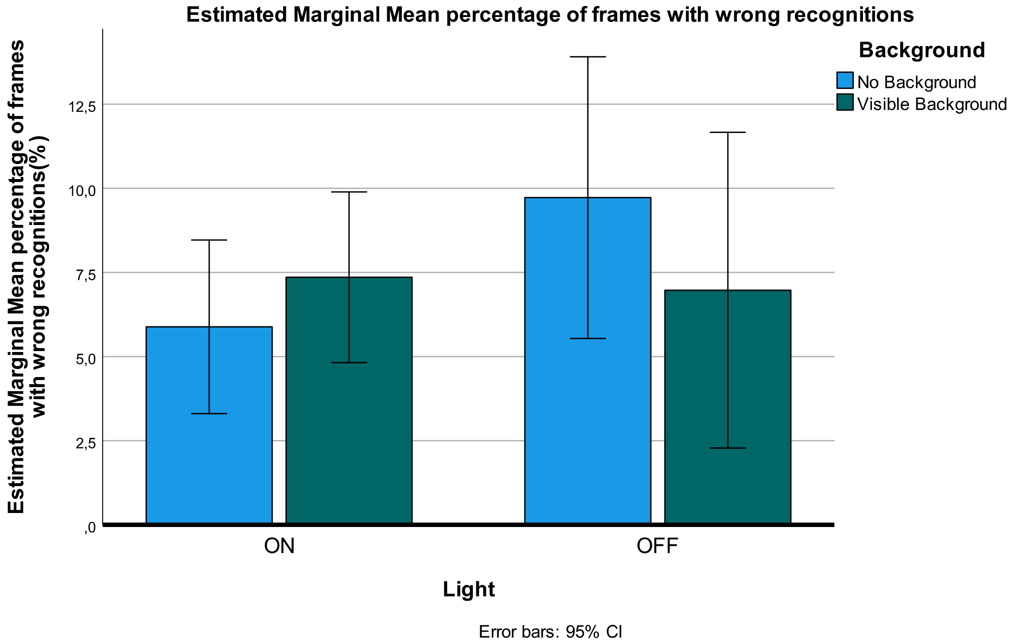


Figure 4.8: The effect of Light and Background on the mean percentage of frames where a wrong recognition was made by Nuitrack.

4.2.4 Mean step length error

To compare the mean error to the expected 60cm step length, a five-way repeated-measures ANOVA was conducted to check for the effect of five independent variables (Step, System, Light, Clothing and Background). No significant interactions were found between these variables. Light ($F(1, 10) = 3.288, p = .100$), Clothing ($F(1, 10) = 2.849, p = .122$) and Background ($F(1, 10) = 0.154, p = .703$) did not have a significant effect on the mean step length error. The main effect of System showed to be significant with $F(1, 659, 16, 591) = 7.991, p = .005$. A post hoc test showed significant differences between Cubemos ($M = -0.017m, SD = 0.022$) and Nuitrack ($M = -0.078m, SD = 0.017$) with $p = .019$. A significant difference was also found between OpenPose ($M = -0.012m, SD = 0.008$) and Nuitrack with $p = .001$. No significant difference was found for OpenPose and Cubemos ($p = .774$). This indicates a significantly higher mean step length error for Nuitrack over Cubemos and OpenPose.

The main effect of the Steps showed to be significant with an F ratio of $F(2.884, 28.843) = 44.361, p < .001$. The mean step length error is shown in **Figure 4.9** for every step. The results of the post hoc test are displayed in **Appendix C** in **Table C.1**. The most clear difference is between the two steps towards the Posterior marker (away from the camera) and all other steps. The means of all other steps, vary between $M = 0.050m, SD = 0.010$ (Right foot, Anterior marker) and $M = -0.030m, SD = 0.017$

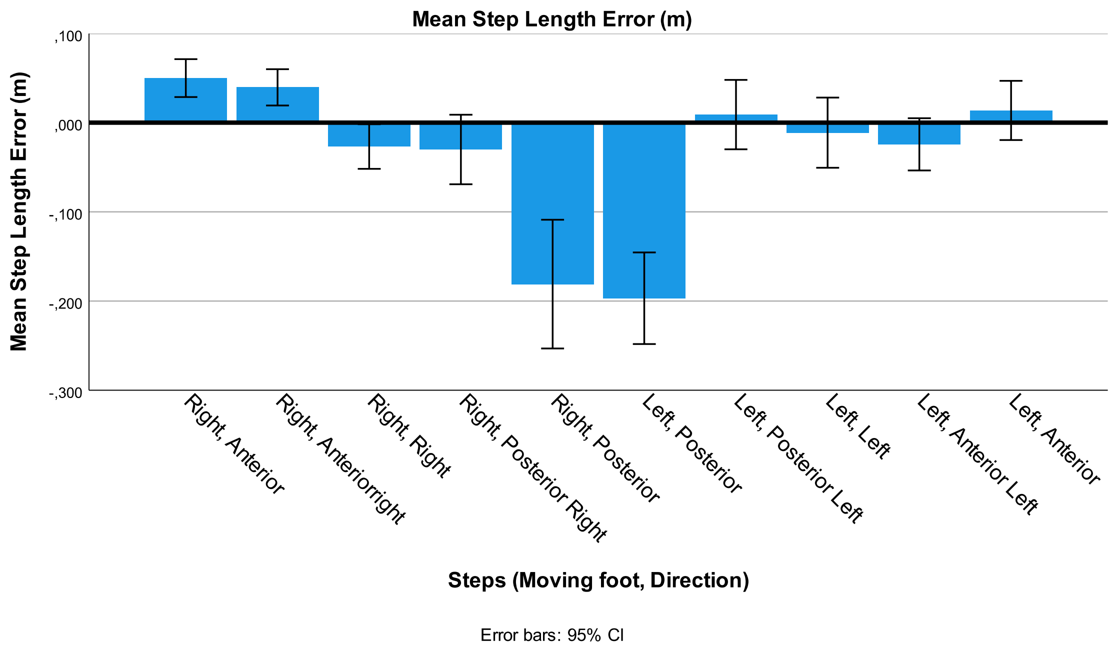


Figure 4.9: Mean Step Length Error for every step

(Right foot, Posterior Right marker), while the right and left foot steps to the Posterior marker show errors of $M = 0.181m, SD = 0.032$ and $M = 0.197m, SD = 0.023$ respectively.

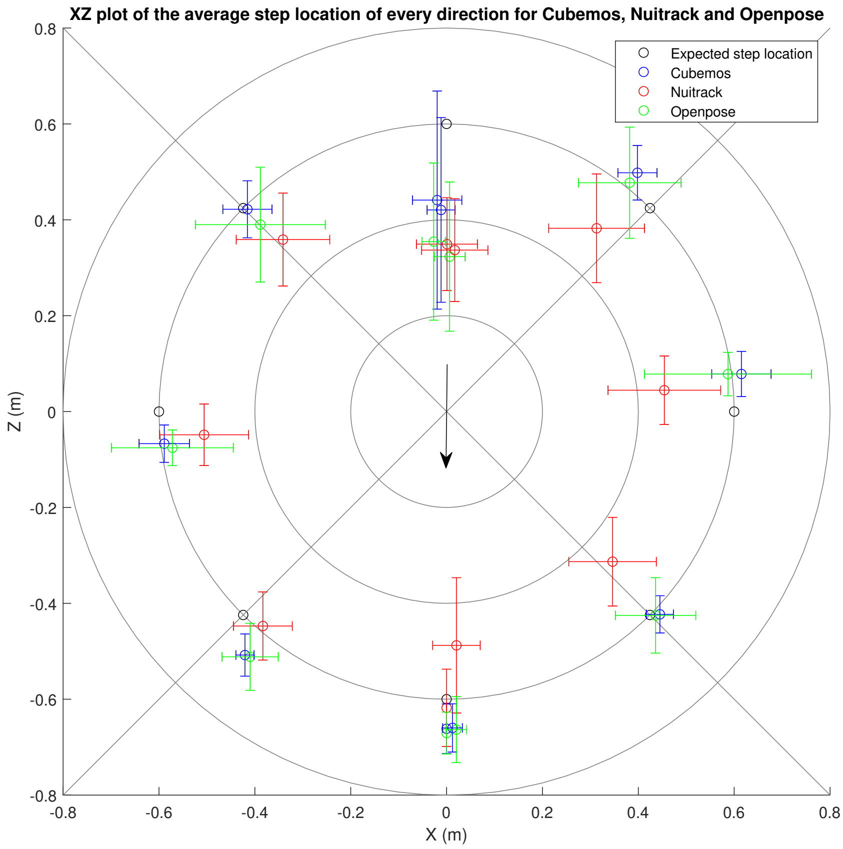


Figure 4.10: Mean Step Length Error for every step. The arrow points in the direction the subject is facing.

Figure 4.10 shows where the mean stepping end position is for every step after each X- and Z-coordinate has been translated to the middle position by subtracting the distance over the X- and Z- coordinate between the origin and middle mean resting position for each video. As described, for every video a small rotation around the Y-axis was also performed to correct for small changes in setup between videos.

Chapter 5

Discussion

The following chapter will interpret the results presented in the previous chapter and compare those to existing literature on the subject, if available.

5.1 Skeleton Trackers compared to Qualisys

The most important finding of this experiment is that the root mean squared error over the entire video was the lowest for Cubemos for both feet, followed by OpenPose and then Nuitrack, indicating that of the three skeleton trackers Cubemos performed the best when being compared to a golden standard.

The higher root mean squared error for the right foot can be explained by a larger error when the subject was stepping to the Anterior marker with the left foot, as was made visible in **Figure 4.1** and **4.2**. Large RMSEs were also found for steps backward towards the Posterior marker. Steps towards the Anterior and Posterior marker are the steps where occlusion can happen: One of the legs blocks the camera from seeing the other leg. This same effect was visible in **Figure 4.10**, where the final step locations towards the Posterior marker are underestimated in distance and also have a large standard deviation along the Z-axis. Occlusion causes the skeleton trackers to still recognise both ankles, even though one is hidden behind the other. The infrared camera that the 3D camera uses to deproject the 2D images back to 3D now encounters the anterior leg first and estimates the 3D position of the posterior leg to be on the position of the anterior leg, thus failing to recognise its actual depth position. When designing stepping exercises using this setup, it is recommended to avoid movement where this type of occlusion can occur.

All Root Mean Square Errors were large for all steps and the total movement both, indicating a large error between the single-camera skeleton trackers and Qualisys with its twelve cameras. However, for all skeleton trackers it is visible that steps can be distinguished from one another, as none of the mean step locations and their standard deviations overlap (**Figure 4.10**) With Root Mean Squared Errors ranging from 8.6cm to 29.6cm over the entire recording, all three skeleton trackers have difficulty accurately estimating the actual ankle location and are not suitable for precise location estimation of the ankle joints. However, if precise location estimation is not necessary, it can be stated that steps in different directions can be distinguished from one another. As is better made visible in **Figure 4.1**, this is more problematic for Nuitrack, as the noise surrounding the middle position is high, which means that smaller steps could fall in this noise area and not be recognised. OpenPose and Cubemos show less noise, making these two systems more suitable for step direction recognition.

Earlier research into the accuracy of OpenPose showed a similar result. Qualitatively recognised joint positions corresponded well to marker-based motion capture and followed its trajectory closely. Quantitatively, this research showed mean absolute errors ranging from 20mm to 40mm. This research however used five normal cameras surrounding the subject in comparison to this research's one 3D camera, which could explain the difference in found accuracies. [36]

5.2 Mean Step Length Error

The most important finding of this experiment, is that even though position accuracy compared to Qualisys has proven to be low, comparing measured step length errors to the expected 60cm has proven to be acceptable in most stepping directions, showing a max mean error of 5cm towards the anterior marker with the right foot and smaller errors for all other steps. The two steps towards the Posterior marker have a mean higher error to the expected 60cm distance ($M = -0.181m, SD = 0.032$ and $M = -0.197m, SD = 0.023$ for the right and left foot respectively). As the two steps towards the Left-Posterior and Right-Posterior marker do not show this large error while being at a similar depth level, it can be concluded this is not due to the skeleton tracker failing at larger distances, but rather due to the earlier explained effect of occlusion. Occlusion can not happen at the Right-Posterior and Left-Posterior marker, due to no overlap of the two legs occurring during these steps, while they do occur during the steps towards the Posterior marker. Occlusion was also noticed during steps towards the Anterior marker. The non-moving foot was estimated to be further forward than the actual middle position. This is not visible in the results of the second experiment due to the fact only the mean step length error of the moving foot has been analysed. It was however made visible in the results of the comparison to Qualisys with a large RMSE for the nonmoving foot for both steps towards the Anterior marker. (**Table 4.2**) With these results in mind, for a single-camera setup, the recommendation is to not include design stepping exercises that could cause occlusion as these will likely lead to wrong measurements.

A study performed by Dev et al. focused on tracking upper limb movement found a mean wrist endpoint error of $M = 2.6cm, SD = 1.8cm$ compared to a Motion Capture system, which is similar to the results we found for OpenPose's mean step length error ($M = -0.012m, SD = 0.008$), indicating that movement length errors for OpenPose are expected to be in this range [37].

5.3 Skeleton and Joint recognition

Skeleton recognition: During this experiment, Nuitrack never recognised additional skeletons, while Cubemos and OpenPose did. Of the two, OpenPose is the most prone to recognising additional skeletons, which becomes more pronounced when the lights are turned off. This was an expected result due to the bottom up-approach OpenPose uses. [23] Cubemos was not affected by any of the ambient conditions on the percentage of frames where it recognised an additional skeleton. During the analysis of the XZ plots that the MATLAB program did output, it was noted that during frames where multiple skeletons were recognised by Cubemos or OpenPose, the skeletons were always kept apart properly and did not influence the accuracy of the tracking of the actual skeleton of the user. This means that even though Cubemos and OpenPose sometimes make this type of wrong recognition, it does not influence the skeleton tracking of the original user, and is thus not something that needs to be kept in mind during the design of an exergame.

When there is no background available, Nuitrack has significantly more trouble recognising the skeleton of the user when the lights are off. When a background is visible, the state of light does not significantly influence the percentage of frames with a missing skeleton. A possible explanation for this effect could be the contrast a background provides. When a visible background is providing contrast to the user, the light is not needed to provide this contrast, while when no background is visible to provide contrast, only the amount of light is available to do so.

Joint recognition: Nuitrack is significantly better at recognising joints on frames where it recognises a skeleton in comparison to Cubemos and OpenPose, which are more likely to recognise a skeleton but miss certain joints. As expected, the amount of light significantly influences the number of frames where at least one joint is not recognised. When the lights are turned off, the percentage of frames where at least one joint is not recognised is higher for all three skeleton trackers. It was furthermore shown that wearing monochromatic clothing works positively for the recognition of joints. In comparison to speculation by van de Pol, Background does not seem to significantly influence the ability of skeleton trackers to recognise joints on a person [26].

Total wrong recognitions: When looking at the total percentage of frames with wrong recognitions, it can be observed that Nuitrack and Cubemos have similar percentages of frames with wrong recognitions in comparison to OpenPose, which has a significantly higher percentage of wrong recognitions. It has to be kept in mind that the wrong recognitions of Nuitrack are mostly due to missing entire skeletons, while the wrong recognitions of Cubemos and OpenPose are due to only missing single joints or seeing additional skeletons where there aren't any. It could be argued that not recognising the user's skeleton is a worse wrong recognition, than either missing only single joints or recognising additional skeletons (which does not impact stepping accuracy, as mentioned earlier). Cubemos is not significantly affected by either of the three main variables (Light, Background, Clothing) for proper recognition, which is a positive for an at-home user, as they can use this system anywhere, under all tested conditions. Nuitrack was only significantly affected by the visibility of the background when the lights were turned off, where having a background lowered the number of wrong recognitions. OpenPose is significantly affected by the amount of light in the room, strongly showing that having the lights turned on lowers the percentage of frames with wrong recognitions.

Considering all results, for a robust system for single-depth camera skeleton tracking with a low position- and mean step length error, Cubemos is a good option. It is not affected by the type of clothing, the amount of light, and the type of background with regards to mean step length error and percentage of frames with wrong recognitions. OpenPose shows generally the same results as Cubemos, but has a higher percentage of wrong recognitions and is more heavily influenced by the amount of light available. Nuitrack is not a good option, as it is the only skeleton tracker that sometimes fails to recognise the entire user, and has high errors compared to Qualisys and the expected 60cm step length. Its joint position data is also visibly the noisiest of the three skeleton trackers.

5.4 Limitations

Data obtained from Cubemos was used to select plateaus in the distance between the two ankle joints. This data was then used to calculate various mathematical values. Due to the fact these plateaus were selected on Cubemos' data, it could have skewed the results more favorably for this system. This choice was however made due to the high amount of noise in the Nuitrack data, which would make a proper selection of plateaus (nearly) impossible. Due to the automatic plateau selection algorithm not being completely proper, user input was allowed to fix the automatic detection. A different user may therefore find different results for the same video. Even though finding plateaus in the data could be difficult for the algorithm, for the human eye this was always easy. No big differences are to be expected between users. No testing was done to find inter-and intra-observer variability, however.

Since the recordings were made over several different days at several different times during the day, two things may have influenced the results. First of all, the different times of day and weather outside could have altered the amount of light in the room. Closing all the blinds in the room was performed to homogenise the setting as much as possible. Because the measurements were performed on different days and the setup had to be done again on every one of those days, small differences in the setup may have happened as well. (Angle mat, distance greenscreen to subject) Corrections after the analysis of the videos by the three skeleton trackers were performed to better compare the videos on stepping location and stepping distance. However, if these (small) differences affected the skeleton/joint recognition by the skeleton trackers, this may have altered those results. This is however not expected.

Because there is no proper comparison with a gold standard during the second experiment, it is hard to say if the subjects indeed made steps of sixty centimeters. This makes the outcome measure of the mean distance between the recognised final stepping distance and the expected sixty centimeters less valuable. However, because a clear, large difference can be seen between the two steps where the person steps backward and all other remaining steps, it can still be concluded that all systems have more trouble properly measuring the steps away from the camera.

Something that has to be kept in mind while interpreting the statistical analysis of the skeleton and joint recognition, is that the data was not normally distributed. Performing a factorial repeated measures ANOVA on non-normally distributed data increases the chances of finding false-positive results. As most found p-values found for these results were $p < 0.032$ it can be assumed that these found differences are still significant, as these p-values are still far below the 0.050 threshold. The highest p-values found were those of the significance of the effect of Clothing on the percentage of frames with missing joints with a p-value of 0.041 and the effect of Background on the percentage of frames with wrong recognitions for Nuitrack if the lights are turned off with a p-value of 0.042. These results could therefore be false positive, and in reality not have an influence on joint- and skeleton recognition.

5.5 Recommendations for future studies

The results of this study lay a good groundwork for comparing various skeleton trackers and the effect of ambient conditions and user properties on their performance. During this research only one type of 3D camera was used, the Intel Realsense Depth Camera d435i. It would be worth investigating if other Depth Cameras perform differently for these skeleton trackers. More ambient conditions and user properties could also be tested for. Possible ambient conditions could be movement behind the user, light shining in the camera or small objects blocking the view of the user, like a small table. Possible user properties could be gender, height, tightness of clothing, and/or ethnicity.

Chapter 6

Conclusions and recommendations

To conclude, all three skeleton trackers show a large error to Qualisys. However, for every skeleton tracker, all steps are still clearly recognised and distinguishable from one another. For every analysed skeleton tracker, steps away from the camera are strongly underestimated in distance, probably due to occlusion of the non-moving leg. When designing exercises for at-home rehabilitation, this should be kept in mind. Recognised step length for all three systems showed not to be significantly affected by any environmental factors or user property. Nuitrack did have the biggest error to the expected 60-centimeter step length and also the highest RMSE compared to Qualisys, while Cubemos and OpenPose showed lower errors. Thus, if the exergame needs proper measurement of step distance, Nuitrack is not a good option.

For the percentage of frames with at least one missing joint, all Systems showed to be significantly affected by the amount of light and type of clothing. For the percentage of frames with wrong recognitions, Cubemos showed to be resistant to the influence of Light, Background and Clothing. OpenPose showed to be affected to be influenced by Light. For both Cubemos and OpenPose, for the user the recommendation is to wear monochromatic clothing and make sure the room is properly lit for the best skeleton recognition, as this lowers the percentage of frames where joints are not recognised. The background does not significantly influence the accuracy of the skeleton trackers, which means the user does not have the burden of preparing a proper background for exercising. As Cubemos showed a significantly lower percentage of wrong recognitions than OpenPose, thus becoming the recommended choice for single 3D camera at home skeleton tracking.

Bibliography

[1] J. Roelofs, I. Schut, A. Huisinga, L. Leeuw, J. Bussmann, A. Geurts, and V. Weerdesteyn, "Minor stroke, serious problems: the impact on balance and gait capacity, fall rate, and physical activity," *ASSESSING BALANCE CONTROL AFTER MINOR STROKE*, p. 89, 2020.

[2] V. Weerdesteijn, M. d. Niet, H. Van Duijnhoven, and A. C. Geurts, "Falls in individuals with stroke." 2008.

[3] P. Langhorne, F. Coupar, and A. Pollock, "Motor recovery after stroke: a systematic review," *The Lancet Neurology*, vol. 8, no. 8, pp. 741–754, 2009. [Online]. Available: http://dx.doi.org/10.1016/S1474-4422(09)70150-4

[4] M. Van Eeden, C. Van Heugten, G. van Mastrigt, M. van Mierlo, J. Visser-Meily, and S. Evers, "The burden of stroke in the netherlands: estimating quality of life and costs for 1 year poststroke," *BMJ open*, vol. 5, no. 11, p. e008220, 2015.

[5] D. Spencer, "Long-term risk of stroke in children treated with growth hormone for short stature: Looking at the big picture," *Neurology*, vol. 83, no. 9, pp. e103–e104, 2014.

[6] A. A. Schmid, S. E. Arnold, V. A. Jones, M. Jane Ritter, S. A. Sapp, and M. Van Puymbroeck, "Fear of falling in people with chronic stroke," *The American Journal of Occupational Therapy*, vol. 69, no. 3, pp. 6 903 350 020p1–6 903 350 020p5, 2015.

[7] C. Torkia, K. L. Best, W. C. Miller, and J. J. Eng, "Balance Confidence: A Predictor of Perceived Physical Function, Perceived Mobility, and Perceived Recovery 1 Year after Inpatient Stroke Rehabilitation," *Archives of Physical Medicine and Rehabilitation*, vol. 97, no. 7, pp. 1064–1071, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.apmr.2016.03.004

[8] C. L. Lai, Y. L. Huang, T. K. Liao, C. M. Tseng, Y. F. Chen, and D. Erdenetsogt, "A Microsoft Kinect-Based Virtual Rehabilitation System to Train Balance Ability for Stroke Patients," *Proceedings - 2015 International Conference on Cyberworlds, CW 2015*, pp. 54–60, 2016.

[9] M. Musicco, L. Emberti, G. Nappi, C. Caltagirone, I. M. S. on Outcomes of Rehabilitation of Neurological Patients *et al.*, "Early and long-term outcome of rehabilitation in stroke patients: the role of patient characteristics, time of initiation, and duration of interventions," *Archives of physical medicine and rehabilitation*, vol. 84, no. 4, pp. 551–558, 2003.

[10] J. Young and A. Forster, "Rehabilitation after stroke," *British Medical Journal*, vol. 334, no. 7584, pp. 86–90, 2007.

[11] P. Langhorne and S. Ramachandra, "Organised inpatient (stroke unit) care for stroke: network meta-analysis," *Cochrane Database of Systematic Reviews*, vol. 2020, no. 4, 2020.

[12] R. P. Van Peppen, G. Kwakkel, S. Wood-Dauphinee, H. J. Hendriks, P. J. Van der Wees, and J. Dekker, "The impact of physical therapy on functional outcomes after stroke: What's the evidence?" *Clinical Rehabilitation*, vol. 18, no. 8, pp. 833–862, 2004.

[13] P. Fearon, P. Langhorne, and E. S. D. Trialists, "Services for reducing duration of hospital care for acute stroke patients," *Cochrane database of systematic reviews*, no. 9, 2012.

[14] R. Lloréns, M. Alcañiz, C. Colomer, and M. D. Navarro, "Balance recovery through virtual stepping exercises using kinect skeleton tracking: A follow-up study with chronic stroke patients," *Annual Review of CyberTherapy and Telemedicine*, vol. 10, pp. 108–112, 2012.

[15] P. Ghanouni, T. Jarus, D. Collette, and R. Pringle, "Using virtual reality gaming platforms to improve balance in rehabilitation of stroke survivors," *International Conference on Virtual Rehabilitation, ICVR*, vol. 2017-June, 2017, cited By 3. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85034221510&doi=10.1109%2fICVR.2017.8007465&partnerID=40&md5=cd1abc582c70038f0e56033dbf1a8e11

[16] C.-L. LAI, C.-M. TSENG, D. ERDENETSOGT, T.-K. LIAO, Y.-L. HUANG, and Y.-F. CHEN, "A kinect-based system for balance rehabilitation of stroke patients," *IEICE Transactions on Information and Systems*, vol. E99.D, no. 4, pp. 1032–1037, 2016.

[17] S. Swee, L. You, B. Hang, and D. Kiang, "Development of rehabilitation system using virtual reality," *Proceeding of 2017 International Conference on Robotics, Automation and Sciences, ICORAS 2017*, vol. 2018-March, pp. 1–6, 2018, cited By 4. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=

2-s2.0-85050658024&doi=10.1109%2fICORAS.2017.8308045&partnerID=40&md5=84f90adcebec7512be398ff0abbb1d60

[18] M. P. Nett, M. S. Collins, and J. W. Sperling, “Magnetic resonance imaging of acute “wiiitis” of the upper extremity,” *Skeletal radiology*, vol. 37, no. 5, pp. 481–483, 2008.

[19] D. Putrino, H. Zanders, T. Hamilton, A. Rykman, P. Lee, and D. J. Edwards, “Patient Engagement Is Related to Impairment Reduction During Digital Game-Based Therapy in Stroke,” *Games for Health Journal*, vol. 6, no. 5, pp. 295–302, 2017.

[20] D. Osokin, “Real-time 2d multi-person pose estimation on cpu: Lightweight openpose,” *arXiv preprint arXiv:1811.12004*, 2018.

[21] J. Shotton, T. Sharp, A. Fitzgibbon, A. Blake, M. Cook, A. Kipman, M. Finocchio, and R. Moore, “Real-Time human pose recognition in parts from single depth images,” *Communications of the ACM*, vol. 56, no. 1, pp. 116–124, 2013.

[22] Oct 2021. [Online]. Available: https://docs.microsoft.com/en-us/azure/kinect-dk/body-joints

[23] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2021.

[24] L. Yang, L. Zhang, H. Dong, A. Alelaiwi, and A. E. Saddik, “Evaluating and improving the depth accuracy of kinect for windows v2,” *IEEE Sensors Journal*, vol. 15, no. 8, pp. 4275–4285, 2015.

[25] N. Nakano, T. Sakura, K. Ueda, L. Omura, A. Kimura, Y. Iino, S. Fukashiro, and S. Yoshioka, “Evaluation of 3d markerless motion capture accuracy using openpose with multiple video cameras,” *Frontiers in Sports and Active Living*, vol. 2, p. 50, 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fspor.2020.00050

[26] E. van den Pol, “What will be the effect of different ambience conditions on the accuracy of balance recovery measurements using a camera based system,” B.Sc thesis, University of Twente, 2021.

[27] J. Torner Ribé, G. Serrancolí, I. Ayestaran Arriaga, F. Alpiste Penalba, A. Iriarte, C. Margelí, B. Mayans, A. García, S. Morancho, and C. Molas, “Design

and development of biomedical applications for post-stroke rehabilitation subjects," in *Recerca i Tecnologia en Enginyeria Gràfica i Disseny a la Universitat Politècnica de Catalunya (Volum 2)*. OmniaScience, 2021, pp. 159–171.

[28] R. Proffitt, C.-Y. Chang, B. Lange, K. Feeley, and S. Koenig, "Mystic isle," 2013. [Online]. Available: https://ict.usc.edu/prototypes/mystic-isle/

[29] R. Proffitt and B. Lange, "Feasibility of a customized, in-home, game-based stroke exercise program using the microsoft kinect® sensor," *International Journal of Telerehabilitation*, vol. 7, no. 2, p. 23, 2015.

[30] D.-S. Park, D.-G. Lee, K. Lee, and G. Lee, "Effects of virtual reality training using xbox kinect on motor function in stroke survivors: a preliminary study," *Journal of Stroke and Cerebrovascular Diseases*, vol. 26, no. 10, pp. 2313–2319, 2017.

[31] A. Viswakumar, V. Rajagopalan, T. Ray, and C. Parimi, "Human gait analysis using openpose," in *2019 Fifth International Conference on Image Information Processing (ICIIP)*. IEEE, 2019, pp. 310–314.

[32] "Motion capture for engineering," Nov 2021. [Online]. Available: https://www.qualisys.com/applications/engineering/#:~:text=Qualisys%20motion%20capture%20cameras%20are,position%20tracking%20in%20engineering%20applications.

[33] S. Payandeh and J. Wael, "Experimental Study of a Deep-Learning RGB-D Tracker for Virtual Remote Human Model Reconstruction," *International Journal of Telemedicine and Applications*, vol. 2021, 2021.

[34] "Joints names and locations scheme," Sep 2018. [Online]. Available: https://community.nuitrack.com/t/joints-names-and-locations-scheme/685/1

[35] G. Hidalgo, Z. Cao, T. Simon, S.-E. Wei, Y. Raaj, H. Joo, and Y. Sheikh, "Openpose documentation," Jul 2017. [Online]. Available: https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/md_doc_02_output.html

[36] N. Nakano, T. Sakura, K. Ueda, L. Omura, A. Kimura, Y. Iino, S. Fukashiro, and S. Yoshioka, "Evaluation of 3d markerless motion capture accuracy using openpose with multiple video cameras," *Frontiers in sports and active living*, vol. 2, p. 50, 2020.

[37] T. Dev, S. Selvaraj, H. P. Magimairaj, S. Balasubramanian *et al.*, "Accuracy of single rgbd camera-based upper-limb movement tracking using openpose," in

International Conference on NeuroRehabilitation. Springer, 2020, pp. 251–255.

Appendix A

Table with all of the performed tests

Table A.1: All performed tests summarised.

Skeleton Tracker	Lighting Condition	Color of Clothes worn	Busy Background
Cubemos	Low	Monochromatic	Yes
			No
		Multichromatic	Yes
			No
	Normal	Monochromatic	Yes
			No
		Multichromatic	Yes
			No
Nuitrack	Low	Monochromatic	Yes
			No
		Multichromatic	Yes
			No
	Normal	Monochromatic	Yes
			No
		Multichromatic	Yes
			No
OpenPose	Low	Monochromatic	Yes
			No
		Multichromatic	Yes
			No
	Normal	Monochromatic	Yes
			No
		Multichromatic	Yes
			No

Appendix B

Consent Form Experiment

Consent Form for The Comparison of Accuracy of Camera-based Skeleton Trackers During the Stepping Response for Recovering Balance in a Home Based Environment.

YOU WILL BE GIVEN A COPY OF THIS INFORMED CONSENT FORM

Purpose of this research

The presented research was done as part of the HEROES Project at the University of Twente, ET Faculty, Department of Biomechanical Engineering (BE). The HEROES Project aims to reduce fall-related healthcare utilization and associated costs, preventing falls and related injuries, and help People with Stroke (PwS) maintain independence once they are discharged from primary care unit. The project aims to achieve this by developing an exergaming system for home rehabilitation. The exergame will be focused on improving the stepping response of patients. The exergame will use a camera based skeleton tracking system to identify the position of various different joints of the patient. Various different skeleton trackers exist and this research aims to quantify the accuracy of these skeleton trackers under different environmental conditions, like light, colors of clothes worn and crowdedness of the background.

Benefits and risks of participating

There are little to no risks involved in participating in this study. A possible risk is the participant falling over while performing the instructed motions. This risk is considered negligible. Benefits include a complementary snack after the completion of the experiment. This research project has been reviewed and approved by the ET Ethics Committee.

Procedures for withdrawal from the study

You are able to withdraw yourself from the study at any moment during the experiment or even thereafter. During the experiment you can declare you wish to stop to the supervisor. If you wish to withdraw afterward, you can send an email to the researcher listed on this sheet. In both scenarios, all collected data of you will immediately be deleted.

Collection of personal data

This study includes being recorded using an 3D video camera. This data will be stored and later used for the analysis of the skeleton trackers. Due to the nature of a video, it is not possible for this information to be anonymised. The video will be saved under participant[number] to ensure names are not needed. Personal information, like your age, gender, height, and skin complexion will also be collected. This information will be saved in the same manner (participant[number]). The processed data will also be saved anonymously in the same manner.

Retention period of data

The collected data will be used during the HEROES Project. This project will run for at least 3 more years. The data will be retained for 5 years.

Contact details

In case of complaints, questions or the wish to withdraw from the study, you can reach the researchers, Wessel Nieuwenhuys, Aurora Ruiz Rodriguez or Edwin van Asseldonk on the following email addresses: w.w.nieuwenhuys@student.utwente.nl, a.ruizrodriguez@utwente.nl or e.h.f.vanasseldonk@utwente.nl.

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-bms@utwente.nl.

Please tick the appropriate boxes	Yes	No
Taking part in the study		
I have read and understood the study information dated _________, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	○	○
I hereby declare that (currently) I do not suffer from balance impairment, or have suffered a stroke in the past	○	○
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	○	○
I understand that taking part in the study involves being recorded using a 3D Camera and that this video will be saved while this project is ongoing. These videos will be used for the analysis of camera based skeleton trackers for the HEROES Project. I also understand that personal characteristics such as age, height, skin complexion and gender will be notated. This information will be anonymised in the research paper.	○	○
Risks associated with participating in the study		
I understand that taking part in the study involves the following negligible risks: Falling over while performing the experiment.	○	○
Use of the information in the study		
I understand that information I provide will be used for a research paper comparing various skeleton trackers to better help the HEROES Project (and further movement analysis projects) select the most optimal skeleton tracker.	○	○
I understand that personal information collected about me that can identify me, such as the recorded video and personal details will not be shared beyond the study team.	○	○
Future use and reuse of the information by others		
I give permission for the 3D video recording that I provide to be archived in .bag-, .csv- and .mat-files so it can be used for future research and learning. The .bag file is a saved 3D video file, which will have your person visible. The .mat files only contain the (un)processed joint location data and are thus anonymised. This data will only be used by (future) people of the HEROES Project.	○	○

Signatures

______________________	______________________	________
Name of participant	Signature	Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

Wessel Nieuwenhuys	___________________	________
Researcher name	Signature	Date

UNIVERSITY OF TWENTE.

UNIVERSITY OF TWENTE.

Appendix C

Mean step length Error results

Table C.1: Pairwise comparisons of every step. Bold results mean significant differences. Continuation on the next page

Pairwise Comparisons

(I) Steps	(J) Steps	Mean Difference (I-J)	Std. Error	Sig.
South-Left	North-Right	**-,247***	**,017**	**<,001**
	Northeast-Right	**-,236***	**,018**	**<,001**
	East-Right	**-,170***	**,018**	**<,001**
	Souteast-Right	**-,167***	**,025**	**<,001**
	South-Right	-,016	,026	,556
	Southwest-Left	**-,206***	**,024**	**<,001**
	West-Left	**-,186***	**,019**	**<,001**
	Northwest-Left	**-,173***	**,014**	**<,001**
	North-Left	**-,211***	**,019**	**<,001**
Southwest-Left	North-Right	**-,041***	**,018**	**,047**
	Northeast-Right	-,031	,017	,106
	East-Right	**,036***	**,012**	**,017**
	Souteast-Right	**,039***	**,008**	**<,001**
	South-Right	**,190***	**,028**	**<,001**
	South-Left	**,206***	**,024**	**<,001**
	West-Left	,020	,010	,072
	Northwest-Left	,033	,017	,075
	North-Left	-,005	,020	,822
West-Left	North-Right	**-,061***	**,014**	**,001**
	Northeast-Right	**-,051***	**,014**	**,004**
	East-Right	,016	,008	,091
	Souteast-Right	,019	,012	,152
	South-Right	**,170***	**,022**	**<,001**
	South-Left	**,186***	**,019**	**<,001**
	Southwest-Left	-,020	,010	,072
	Northwest-Left	,013	,011	,271
	North-Left	-,025	,016	,152
Northwest-Left	North-Right	**-,074***	**,006**	**<,001**
	Northeast-Right	**-,064***	**,006**	**<,001**
	East-Right	,002	,007	,738
	Souteast-Right	,006	,019	,763
	South-Right	**,157***	**,027**	**<,001**
	South-Left	**,173***	**,014**	**<,001**
	Southwest-Left	-,033	,017	,075
	West-Left	-,013	,011	,271
	North-Left	**-,038***	**,007**	**<,001**
North-Left	North-Right	**-,036***	**,009**	**,003**
	Northeast-Right	**-,026***	**,008**	**,008**
	East-Right	**,040***	**,012**	**,007**
	Souteast-Right	,044	,023	,084
	South-Right	**,195***	**,030**	**<,001**
	South-Left	**,211***	**,019**	**<,001**
	Southwest-Left	,005	,020	,822
	West-Left	,025	,016	,152
	Northwest-Left	**,038***	**,007**	**<,001**

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

Pairwise Comparisons

(I) Steps	(J) Steps	Mean Difference (I-J)	Std. Error	Sig.
North-Right	Northeast-Right	**,010***	**,003**	**,010**
	East-Right	**,077***	**,007**	**<,001**
	Souteast-Right	**,080***	**,019**	**,002**
	South-Right	**,231***	**,029**	**<,001**
	South-Left	**,247***	**,017**	**<,001**
	Southwest-Left	**,041***	**,018**	**,047**
	West-Left	**,061***	**,014**	**,001**
	Northwest-Left	**,074***	**,006**	**<,001**
	North-Left	**,036***	**,009**	**,003**
Northeast-Right	North-Right	**-,010***	**,003**	**,010**
	East-Right	**,066***	**,007**	**<,001**
	Souteast-Right	**,070***	**,018**	**,003**
	South-Right	**,221***	**,029**	**<,001**
	South-Left	**,236***	**,018**	**<,001**
	Southwest-Left	,031	,017	,106
	West-Left	**,051***	**,014**	**,004**
	Northwest-Left	**,064***	**,006**	**<,001**
	North-Left	**,026***	**,008**	**,008**
East-Right	North-Right	**-,077***	**,007**	**<,001**
	Northeast-Right	**-,066***	**,007**	**<,001**
	Souteast-Right	,003	,012	,795
	South-Right	**,154***	**,026**	**<,001**
	South-Left	**,170***	**,018**	**<,001**
	Southwest-Left	**-,036***	**,012**	**,017**
	West-Left	-,016	,008	,091
	Northwest-Left	-,002	,007	,738
	North-Left	**-,040***	**,012**	**,007**
Southeast-Right	North-Right	**-,080***	**,019**	**,002**
	Northeast-Right	**-,070***	**,018**	**,003**
	East-Right	-,003	,012	,795
	South-Right	**,151***	**,029**	**<,001**
	South-Left	**,167***	**,025**	**<,001**
	Southwest-Left	**-,039***	**,008**	**<,001**
	West-Left	-,019	,012	,152
	Northwest-Left	-,006	,019	,763
	North-Left	-,044	,023	,084
South-Right	North-Right	**-,231***	**,029**	**<,001**
	Northeast-Right	**-,221***	**,029**	**<,001**
	East-Right	**-,154***	**,026**	**<,001**
	Souteast-Right	**-,151***	**,029**	**<,001**
	South-Left	,016	,026	,556
	Southwest-Left	**-,190***	**,028**	**<,001**
	West-Left	**-,170***	**,022**	**<,001**
	Northwest-Left	**-,157***	**,027**	**<,001**
	North-Left	**-,195***	**,030**	**<,001**

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

Appendix D

Screenshots Skeleton Analysis Program

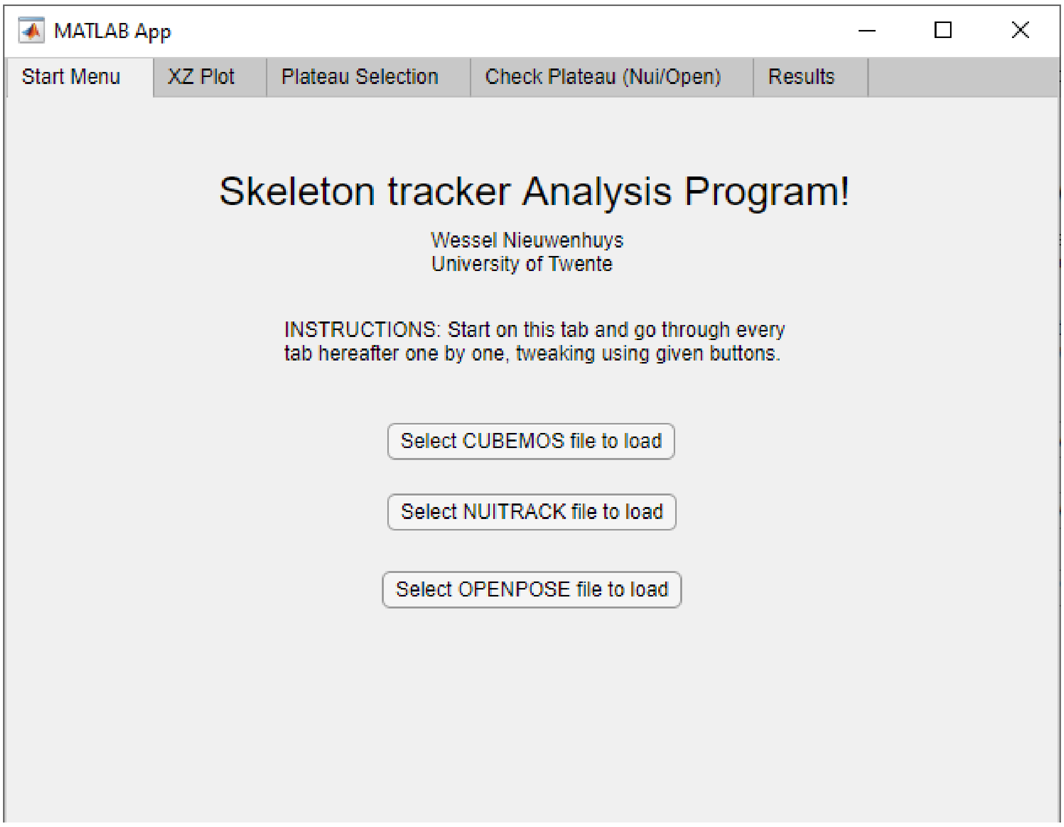


Figure D.1: Opening screen of the application where data can be loaded in

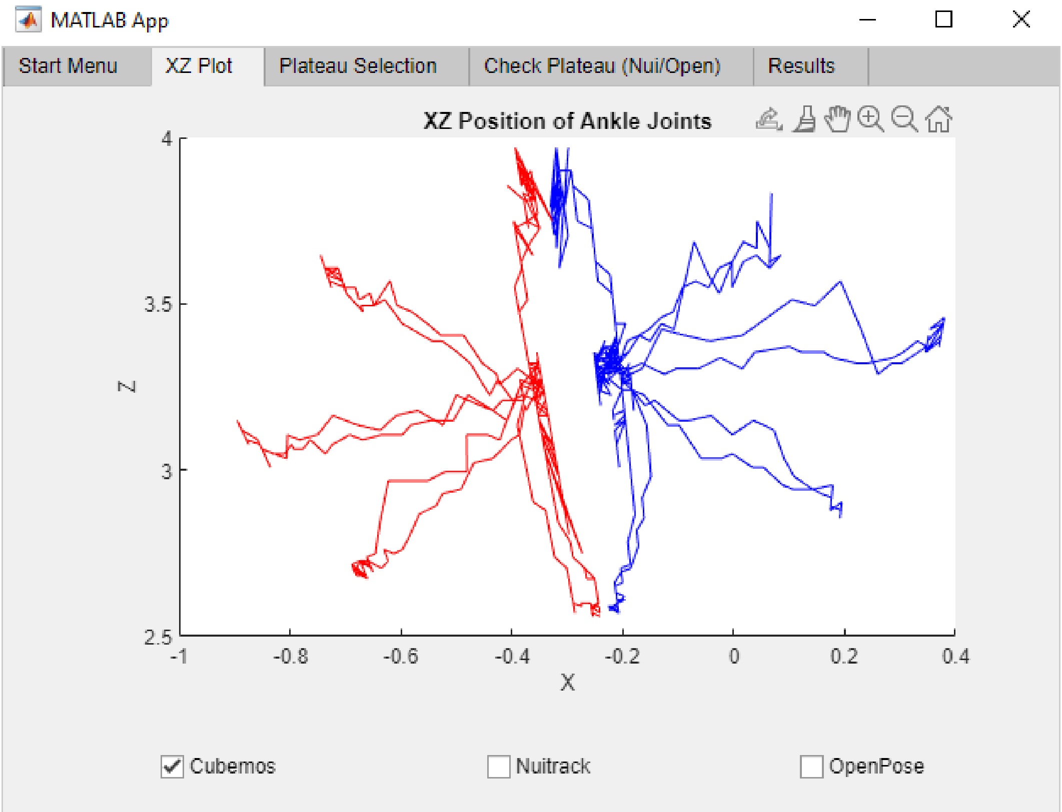


Figure D.2: XZ-plot screen

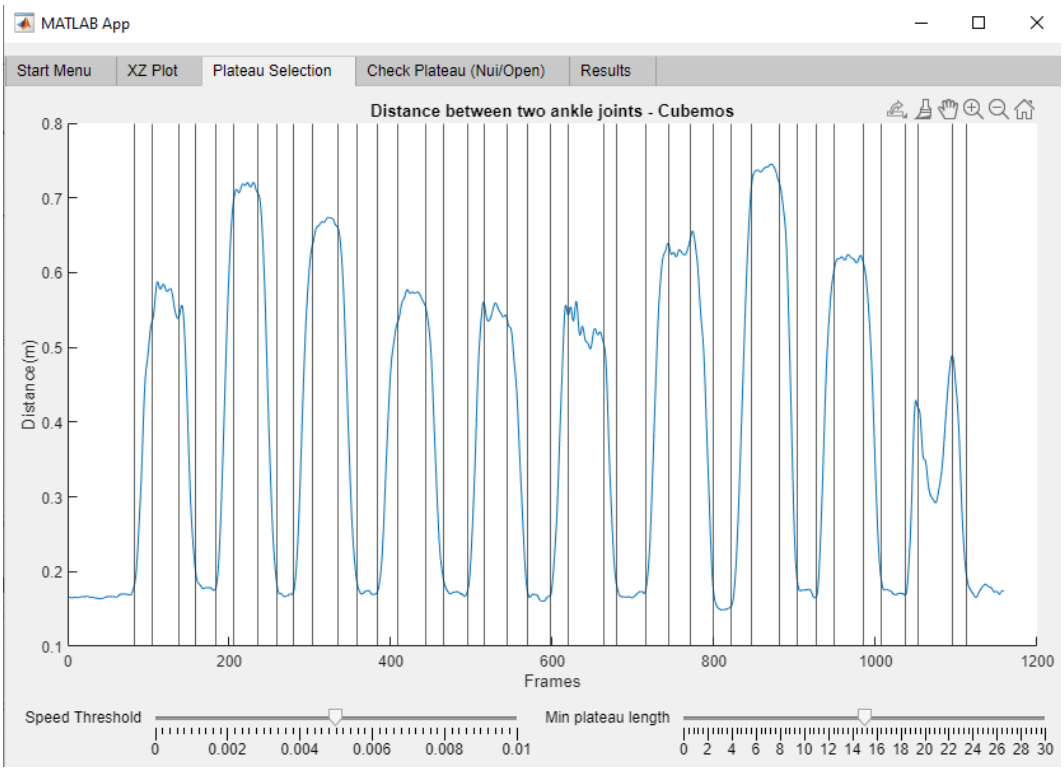


Figure D.3: Result of automatic plateau detection on the Cubemos data

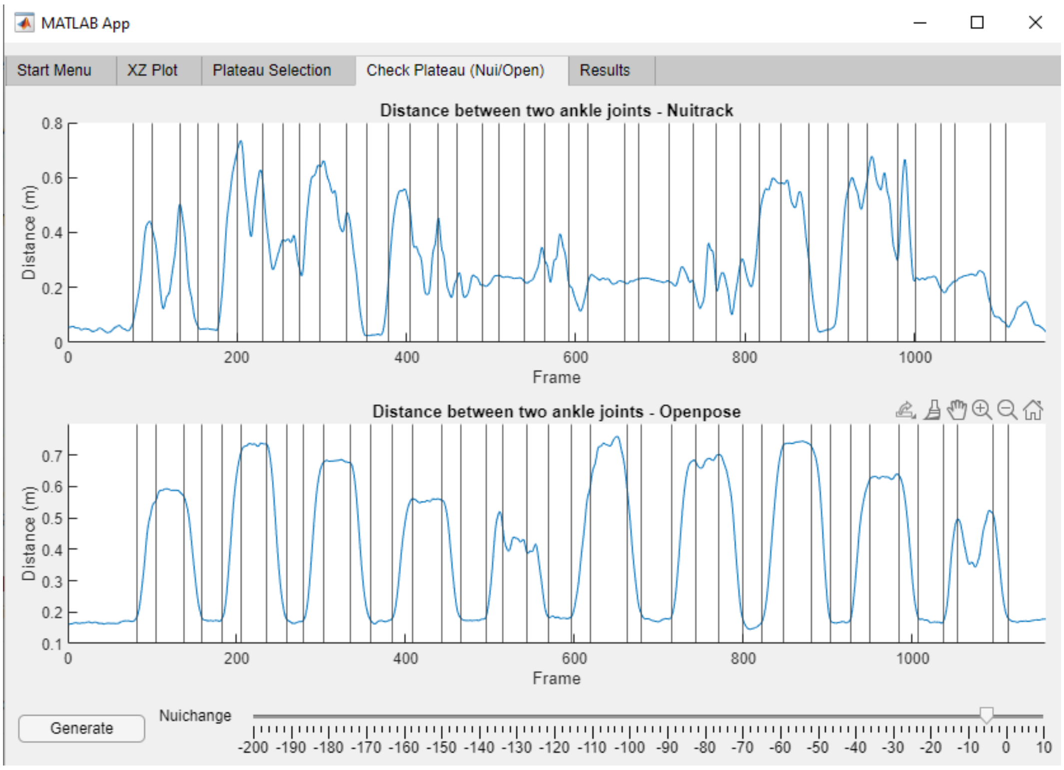


Figure D.4: The same plateau detection pushed through to Nuitrack and Openpose

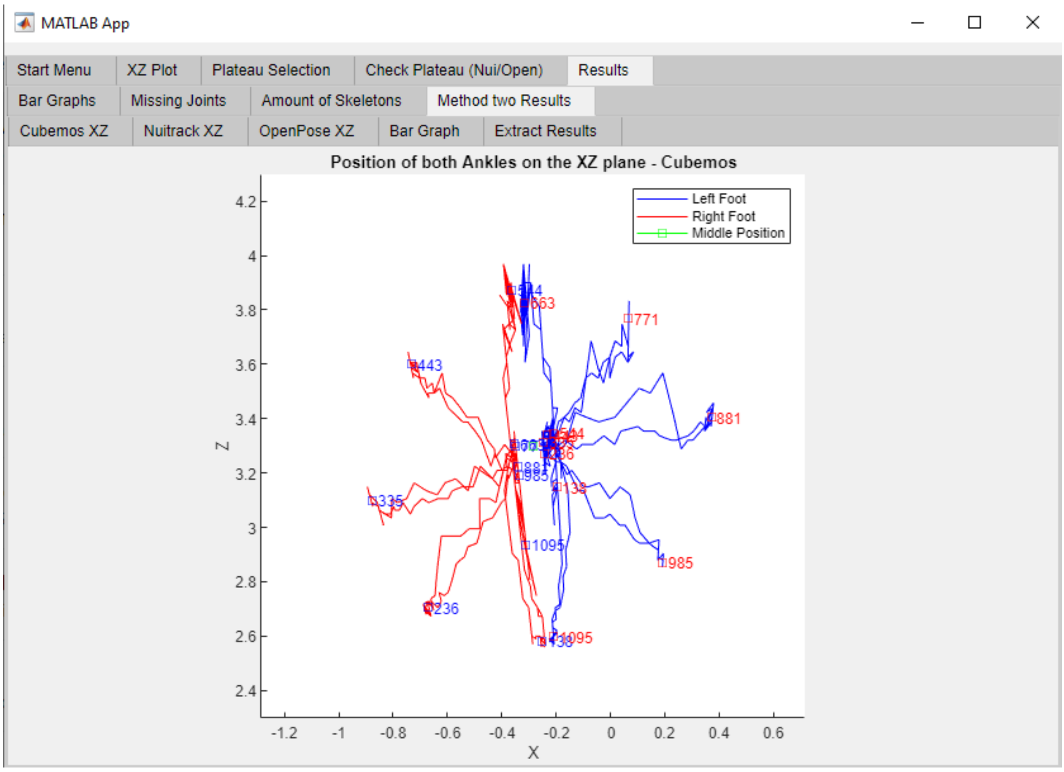


Figure D.5: Detection of the mean end position of every step

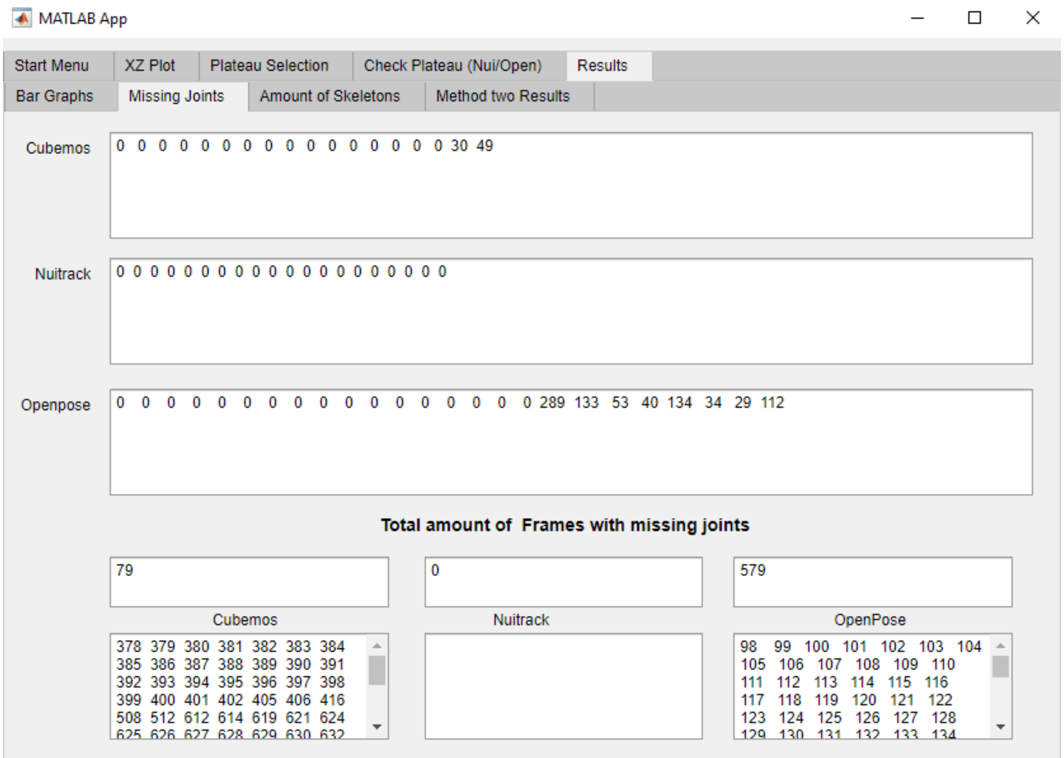


Figure D.6: Screen displaying all the data for missing joint calculations. The top half shows the amount of frames a certain joint was not recognised. The bottom half shows first the total amount of frames with at least one missing joint, and then the frame numbers of these frames.

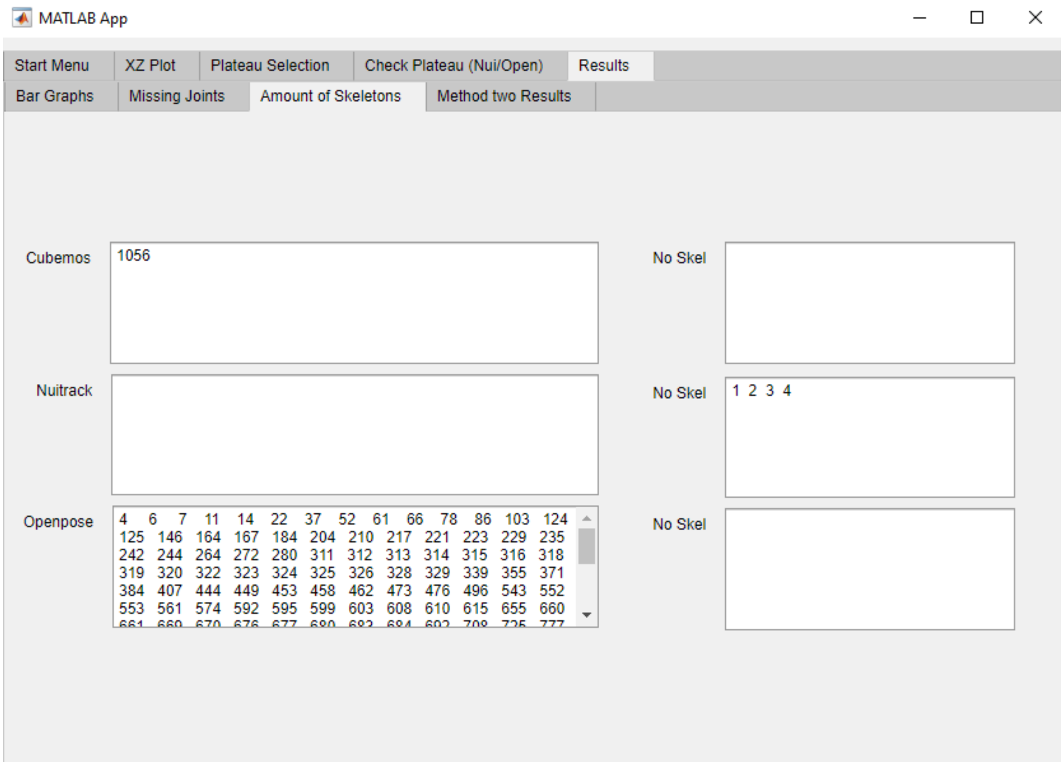


Figure D.7: Screen displaying all the data for missing skeleton calculations. On the left side, all frames with at least one additional recognised skeleton are shown. On the right side, all frames with zero recognised skeletons are shown.

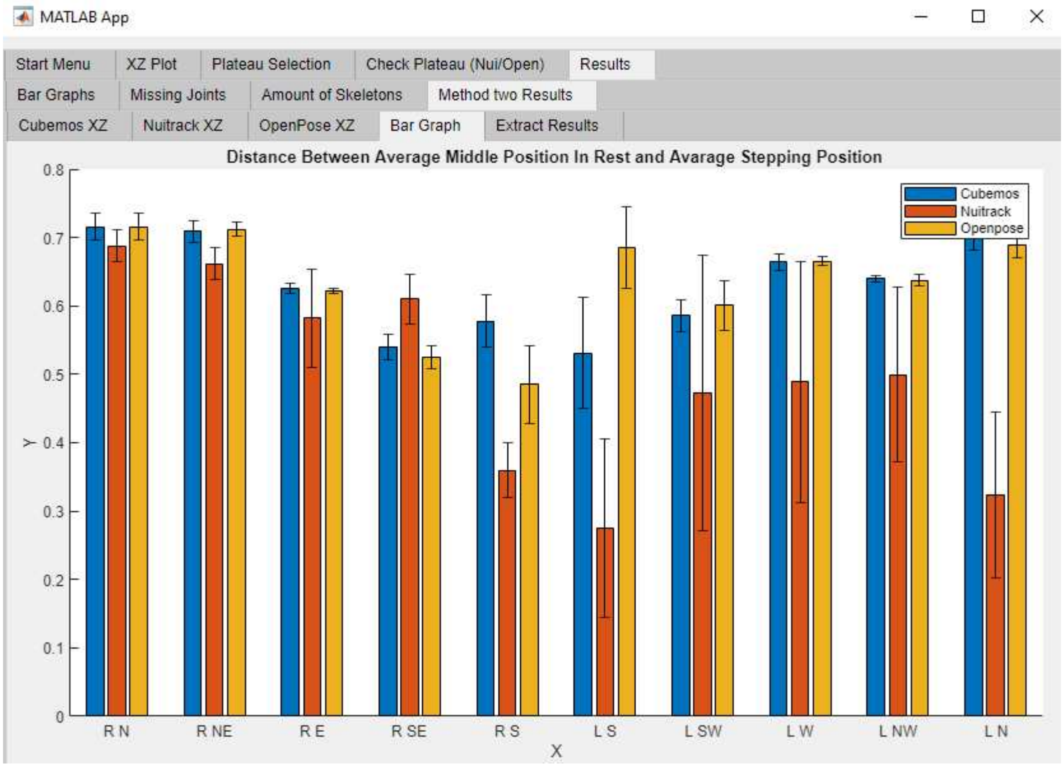


Figure D.8: Bar graph displaying all step distances for every skeleton tracker