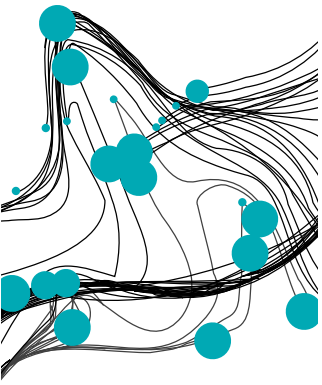
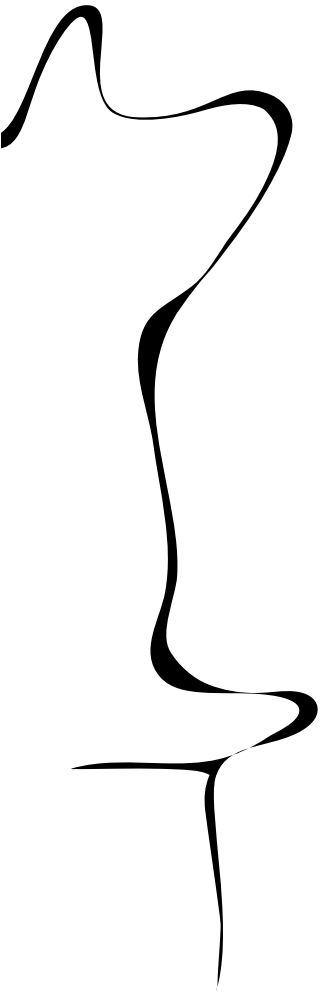


RAM ROBOTICS AND MECHATRONICS



INVESTIGATING THE FEASIBILITY OF USING A REALSENSE DEPTH CAMERA D435I BY CREATING A FRAMEWORK FOR 3D POSE ANALYSIS

R.S. (Reihaneh) Shahmoradi Zavareh

MSC ASSIGNMENT

Committee:
dr. F.J. Siepel
dr. E.H.F. van Asseldonk
A. Ruiz Rodriguez

July, 2022

037RaM2022
Robotics and Mechatronics
EEMCS
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

UNIVERSITY OF TWENTE. | TECHMED CENTRE

UNIVERSITY OF TWENTE. | DIGITAL SOCIETY INSTITUTE

Acknowledgements

I am thankful to Dr Edwin van Asseldonk who guided and advised me during the whole study, without his guidance the research could not be done. I should also present my thanks to Aurora Ruiz Rodriguez and Dr Francois Siepel who in various stages I used their help and advice.

I am also thankful to Jelle Bijlsma and to all my other friends who helped me and volunteered to be subjects of the experiments.

Finally, last but not least, I thank my family for their moral and financial support.

Abstract

The main task in the present study is to investigate the feasibility of using an optimum depth camera to monitor step locations. The Real Sense D435i depth camera is selected for its user friendliness, low cost and ease of availability.

In the general framework used, RGB results from RealSense camera are fed into a software, which is based on machine learning algorithm. This software, called OpenPose, produces the required joint data on human anatomy. Back projection for 3D pose estimation is then done by using the depth values and the output of OpenPose. Filtering of the data is the last stage in the applied framework.

The accuracy of step locations is found by comparing the results with golden standard (Qualisys) measurements. The highest accuracies are obtained for forward steps (steps towards the camera). The results show errors of 1.1 cm and 1.12 cm for right and left forward steps at a distance of 293 cm. The signal to noise ratio (SNR) values are derived and again the mentioned forward steps show better results i.e. 19.73 and 18.9 respectively. The higher distance to the camera is assessed to be the main cause of lower accuracy in the results of backward steps.

The occlusion occurs in some images. The occlusion effect in these images appears as a fake peak in the time series graph for forward steps. This effect appears as a missing the peak for backward steps. The effected results have been left out in the calculations of accuracy assessment. Further works could be performed by using more depth cameras of this type to gain more efficient measurements in all directions.

Keywords: exergames, RealSense D435i, Qualisys, stepping accuracy, SNR, occlusion

Contents

1 Introduction

Frequent falling has been an important subject in patients who experienced brain stroke or other neurological disorders. Among the people with a neurological condition worldwide, patients who have had a stroke (89%), Parkinson's disease (77%), dementia (60%) or epilepsy (57%) suffer from falling [29]. The risk of post-stroke patients falling is ten times higher than people without any neurological condition. This not only leads to having four times greater odds of hip fractures [23] but also a fear of falling which often leads to avoidance of activities, physical deconditioning, mobility limitations or loss of independence [32]. Therefore, prevention of falling is one of the key factors in research in this area.

One of the ways to reduce falling is improving balance quality. This improvement can be achieved by doing exergames at home, which is especially recommended for people with neurological disorders. The most important reasons for doing exergames at home are decreasing the health care costs and reduction of medical centers loads [25]. Further, doing exergames at home gives motivation to the patient to recover without needing a specialist present [28]. This in turn helps feasibility of the long-term treatment since it is at home and always available [18].

Camera-based sensors can be used to provide indoor posing data, and used in quantifying the balance quality for exergames. The accuracy, however, depends on the magnitude of the sensor error [19]. Accuracy improvement of the posing measurements is important for both, quantifying the balance quality and for making exergames more effective. The advantages of camera-based sensors compared to other contact-based sensors are that they reduce system complexity, improve the portability and increase flexibility in use. Depth sensing cameras are mostly used for pose analysis and have wide application in exergames.

Generally, there are two main techniques used in depth cameras: Time-Of-Flight (TOF) and active Infrared Radiation (IR) stereo measurement. In TOF technology the projector radiates a single frequency laser beam. By some means of diffraction many points on the target are torched by the projected beam, i.e. cloud of points. The reflected beam from each of these points on the scene hit the camera sensor and the distance is worked out by measuring the TOF. Sunlight is a disturber in this mechanism of distance measurement and acts as noise. In stereo measurement technology, there are two IR sensors which are situated at different distances on both sides of the IR projector. The image formed on these sensors are compared and the overlap pattern is kept for analysis. Also, the shifts of the scene on the two sensors are used for their depth measurement.

Historically, one of the most-used depth camera in research is Microsoft Kinect V2, which uses TOF. Some studies have been conducted on the accuracy evaluation of Kinect cameras. For instance, David Webster et al [31] investigated the suitability of using the Kinect depth camera for stroke rehabilitation. The specific joints of interest were the shoulder, elbow and wrist. Their results are compared with OptiTrack motion capture system. They concluded that the normalized root mean squared error in position varied between 0.53 cm to 1.74 cm per data point among all axes and joints on average for their suggested distance range of 0.5 m to 2 m. These numbers demonstrate why Microsoft Kinect V2 have been mostly used.

Also Karen Otte et al [16], worked on the accuracy of Kinect V2 sensor. To assess the accuracy of Kinect, he compared the Kinect results with 16-camera Vicon system as a reference system i.e. golden standard. In this study nineteen healthy subjects were used, and six different movement tasks were recorded. 3D full-body kinematics from both systems were employed. In Vicon system 36 attached IR reflecting markers were attached and their positions recorded at 100 hertz frequency (with 2 mm accuracy within an area of 3 m by 6 m). For Kinect the same was performed with 25 landmark positions at 30 hertz frequency. Calculations were performed on

45 clinical parameters. These parameters cover movement velocities, ranges of motions, torso sway and cadence. The information for landmark stability of Kinect V2 was provided by signal to noise ratio and larger noise behavior was seen in feet and ankles. They reported that most of the clinical parameters result in good to excellent accuracy, i.e. Standard Error of Measurement (SEM) < 10%, for both systems.

In similar work, Mahvash Jebeli et al [11] performed the above comparison for application on workers in an industrial cell. Their results indicated that the Kinect sensor exhibited a correlation of the position of the center mass of the body and the location of the RGB-D data: 0.9996 on the depth axis, 0.9849 on the horizontal axis, and 0.2767 on the vertical axis. The imaging was performed in a distance range of 1.5 m to 2.5 m. This shows Kinect has its highest accuracy on the depth axis when compared with the other axes. The study indicates that the results accuracy of the Kinect sensor is good enough to be employed in the industrial areas.

Oliver Wasenmuller et al [30] also study the accuracy and the precision of the images obtained from the Kinect V1 and V2. They study the effect of various parameters such as temperature, the distance of the camera and the scene color on the depth images. The suggestion here is to pre-heat the Kinect V2 for 25 min to obtain stable and repeatable accuracy. This is due to the TOF characteristic of Kinect V2, but the Kinect V1 captures reliable pictures after a few shots. The study shows that the accuracy decreases exponentially for both the Kinect V1 and V2 when the distance increases. Furthermore, the scene color affects Kinect V2 in depth estimation in such a way that the lighter the color, the better the accuracy and precision. However, the study conclude that Kinect V1 is uninfluenced by the object color. The conclusion is that the distance affects both Kinect V1 and Kinect V2 results but the pre-heating and scene color do not influence Kinect V1.

Daphane J. Geerse et al [6] employed a Multi-Kinect V2 for quantitative gait assessments and considered 25 body points. A 10-Meter Walkway was instrumented. Measured positions are compared with the golden standard, OptiTrack. The study concludes all gait measurements, except step width, are in good agreement with the golden standard, i.e. ICC- intraclass correlation coefficient for consistency =< 0.88. In another work performed by Daphne Geerse et al [5], foot placement locations from ankle data is validated using the Kinect V2 depth camera. The effects of distance from the sensor, side and step length on the results were evaluated and then again compared with OptiTrack as golden standard. The between-systems agreement was investigated using intraclass correlation for absolute agreement, ICC(A,1), and consistency, ICC(C,1). Daphne and her colleagues considered values above 0.60 and 0.75 as good and excellent agreement, respectively. Stepping accuracy was seen to be in good agreement with golden standard. However, Kinect V2 underestimates the stepping accuracy at 2 m distance, while overestimates its accuracy at 3 m distance.

When Intel RealSense depth cameras were introduced to the market, they attracted some researchers involved in pose analysis. For instance, J.D.Mejia-Trujillo et al [12] compared measurements obtained from two devices of Kinect V2 and RealSense D435. They established an obstacle free corridor, 1.5 m wide and 6 m long, and measured distance, time and speed of walking volunteers by fixing the cameras at the end of the corridor. They concluded Kinect produces less noisy results. However, RealSense can be used to measure some spatial-temporal variables, especially by using the waist joint. The wrist and ankle joints present noise that may be difficult for the gait analysis. Therefore, signal processing is needed to be done on RealSense results.

Further, similar comparison of two devices, Kinect V2 and RealSense D435, had also been performed by Andres Navarro et al [12]. The obtained results suggest that the RealSense can be better used to measure some spatio-temporal variables for different scenarios, like eHealth and analysis for high-performance athletes, than the Kinect V2.

In another work, Vladimir Tadić et al [24] investigated the effect of additional post-processing of depth images for their robot in the RealSense D415 and the D435. Factory settings or random settings were used to filter during post-processing to improve the edge detection obstacle. It was concluded that the RealSense D415 provides more accurate depth information than the D435 camera. The main difference between the two cameras is the wider field of view of D435. Thus, the D435 has a lower pixel density when compared with the D415. This was presented by the authors as the main reason for the lower accuracy and the lower resolution of the D435 depth camera compared to the D415.

In the present study, the Intel RealSense D435i is selected as a depth camera for pose analysis. The Intel RealSense D435i has been used by some researchers, mostly in dynamic modes for large distances, but in our study this camera is used for static functioning in less than 3 meters distance. Within the 400 series of Intel RealSense, the D435i enjoys both wider view and the RGB sensor. Although the D455 features larger ideal range of 0.6 to 6 meters, the ideal range of the D435i, 0.3 to 3 meters, is sufficient enough for our study. Also, the D435i has a greater field of view (FOV) compared to most of the other D series Intel depth cameras and when compared to the Kinect V1 and the V2. This is an advantage in the home environment because the wider FoV ensures that the object can be closer to the camera. The above reasons give us strong grounds to choose the RealSense D435i as the depth camera for a pose survey.

In this analysis, step positions are measured by the RealSense D435i depth camera. In the present experiment, Mocap system (Qualisys), a passive marker based system, is used as the reference system, i.e. the golden standard. The RealSense measurements accuracy is derived by comparing with the golden standard. The aim in our study is to contribute to the falling prevention. To this end, the step analysis in different directions, using the left and the right foot, is considered to be an important task.

In this study, a general framework is designed (see Section 2.1) for assessing the usage of the RealSense D435i in 3D pose analyses. There is no evidence that other research before have developed this workflow to measure the stepping location of a subject by this camera. This will allow the HERoS project to see if the RealSense depth camera is accurate enough to be used in the exeragame.

2 Methods and Material

The main methods and material used in this study for pose estimation are described in this chapter. In the general framework the calibration method used is explained. The description of employed alignment, OpenPose software and back projection are then given. Finally, the protocol is explained and a solution for processing the data and deriving the position measurements are presented.

2.1 The general Framework for 3D pose estimation using Intel RealSense

In order to find the accuracy of position measurements of the ankle's points, we have to derive the 3D pose estimation of D435i.

Figure 2.1 presents the framework. As shown in this figure, the workflow includes the following phases: the depth camera gives the RGB and the depth data as outputs. The RGB data is used by OpenPose to obtain the 2D coordinates of the keypoints. The 2D positions of the keypoints, alongside with the depth data are then fed to the back projection module to find the 3D pose estimates. The main structure of the workflow and connections between various components is engineered by a *python* software program, developed by the author of the present study.

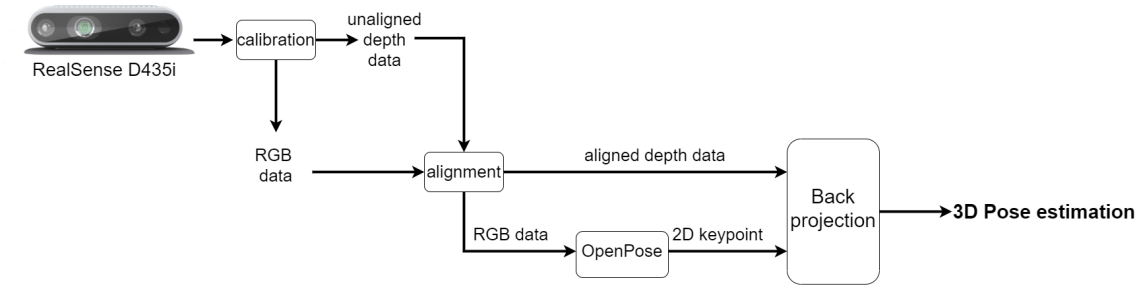


Figure 2.1: general framework for 3D pose estimation

2.1.1 Calibration

According to the manufacturer, an accuracy check must be done prior to using the camera. For a flat surface at a distance of 2 m, the obtained distance should have an accuracy of 2% or better, if not then the camera needs the calibration. An application program, the Intel RealSense D400 series Dynamic Calibration tool, is provided by the manufacturer for this purpose and a printed pattern or a mobile phone screen can be used as a target to calibrate the camera. The phone screen is used as the target in this work. The calibration is explained more in Appendix A.2.

2.1.2 Aligning depth image to RGB image

The Intel RealSense D435i contains a stereo-camera and an RGB camera. These cameras are not aligned spatially. Thus, when extracting the depth data using the same pixel-coordinates, the images have to be aligned first, otherwise the images would be mismatched.

An aligning rout is provided in *python* program for this purpose [22]. The route uses the intrinsic camera parameters to align the depth frame spatially to the RGB frame.

2.1.3 OpenPose software

The software, introduce previously, is capable of handling the total of 135 keypoints, but the present study only employs two keypoints, namely right and left ankle points. OpenPose gets the required joints data from the RGB outputs of the RealSense camera.

The OpenPose running time is constant and not dependent on the number of input keypoints data. The computer used an operating system of Ubuntu 20.04. Thanks to the dedicated GPU (Graphics Processing Unit) which made the computer fast enough for performing our calculations.

2.1.4 Back projection

As it is described in the Appendix A.4, back projection means converting 2D data of an image into 3D coordinate data by using their depth. Using 2D coordinates of keypoints obtained from OpenPose and depth data resulted from RealSense, the spatial 3D coordinates are derived.

The back projection algorithm used in this thesis is *pyrealsense2* which is a library produced in *python* by the Intel RealSense for this purpose. One of the functions in this library takes intrinsic parameters of the camera including focal lengths, principal point and skew factor, alongside with 2D coordinates and depth data, to output the spatial 3D coordinates [21].

2.2 Protocol

First, the subjects criteria is explained, showing the requirements and the demographics of the participants. Second, the materials needed for the protocol are explained. The used camera alongside with the configuration setup are shown. Later on, the data is collected with both the Mocap system and the RealSense camera and prepared for analysis in the last section.

2.2.1 Subjects

For this protocol, we considered the following criteria when selecting the seven subjects:

- Young adult: aged between 18 and 35 to reduce the likelihood that the participants have mobility problems due to age.
- Healthy subjects: participants must not have any Neuromuscular disorder or imbalance problem before or at the time of conducting the protocol.

Due to occurring some error in the golden standard system at the beginning of experiment, the data for the first subject is left out and only 6 other subjects have been dealt with in the continuation of the study.

2.2.2 Materials and Setup

For this protocol, the camera Intel RealSense D435i is used as a sensor for getting data images and depth measurements. The camera is mounted on a tripod which has a height of 0.9 meter. As seen in Figure 2.2 which shows the experimental setup, 12 Qualisys cameras are fixed surrounding the subject. For Qualisys, ten markers are placed on each subject, eight of which are used in the step position calculations. The markers positions on each foot (toe, two ankles on both sides of each foot and the point behind the foot) are shown in Figure 2.3. The setup also contains the computer for running the *Python program*, developed by the author and *QTM (Qualisys Track Manager)* golden standard data acquisition system.

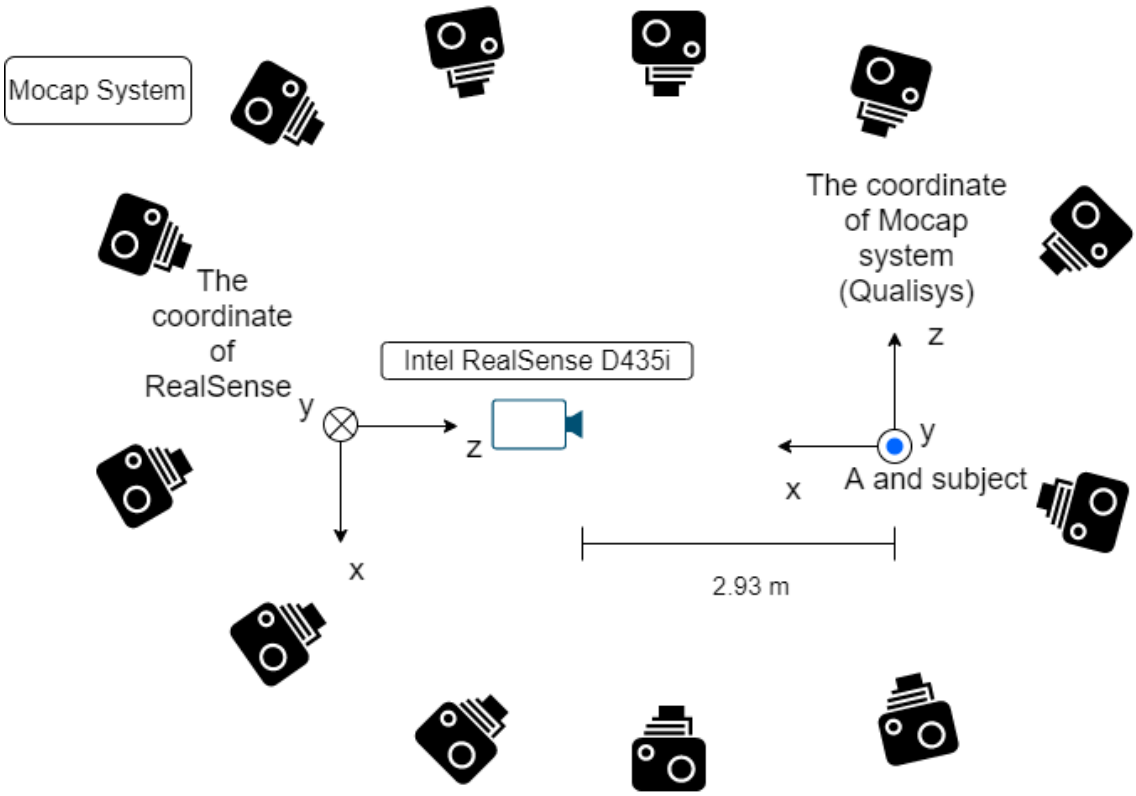


Figure 2.2: top view of the experimental setup, showing the cameras and the coordinates

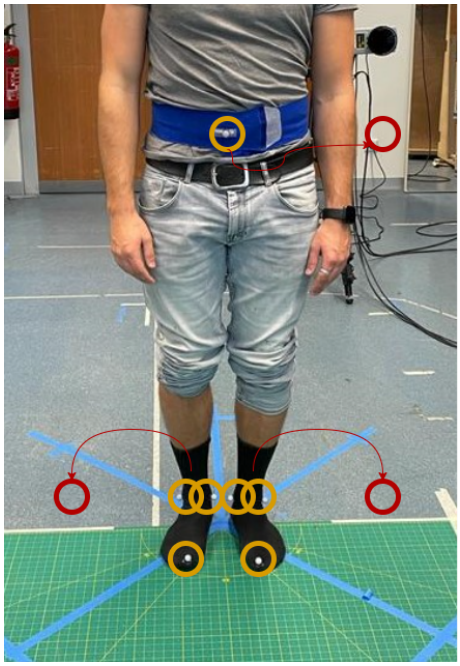


Figure 2.3: for golden standard (Qualisys); the circles show 10 locations of the markers. The rear ones, which can not be seen, are shown in red.

2.2.3 Procedure

After performing the Intel RealSense calibration and preparing the setup, the study procedure is followed as below: subjects are placed with bare feet at the distance of 2.93 meter from the camera (see Figure 2.2). The recording starts by both *python program* and *QTM*. Movement

of each subject starts with raising of right foot from his stand position at point A (see Figure 2.4). Then, the subject moves by placing the right foot at point B. The subject moves back to point A after staying on B for at least one second. This is called the basic step movement. The participant performs the basic step movement of right foot towards point C and back again to A. The procedure is repeated for point D and then in the same manner for the left foot, covering points E, F, G, H and I. Finishing with the left foot, it is the right foot turn again to complete the steps by moving to points J and K and end at point A. This whole procedure, from beginning to the end, is called a cycle.

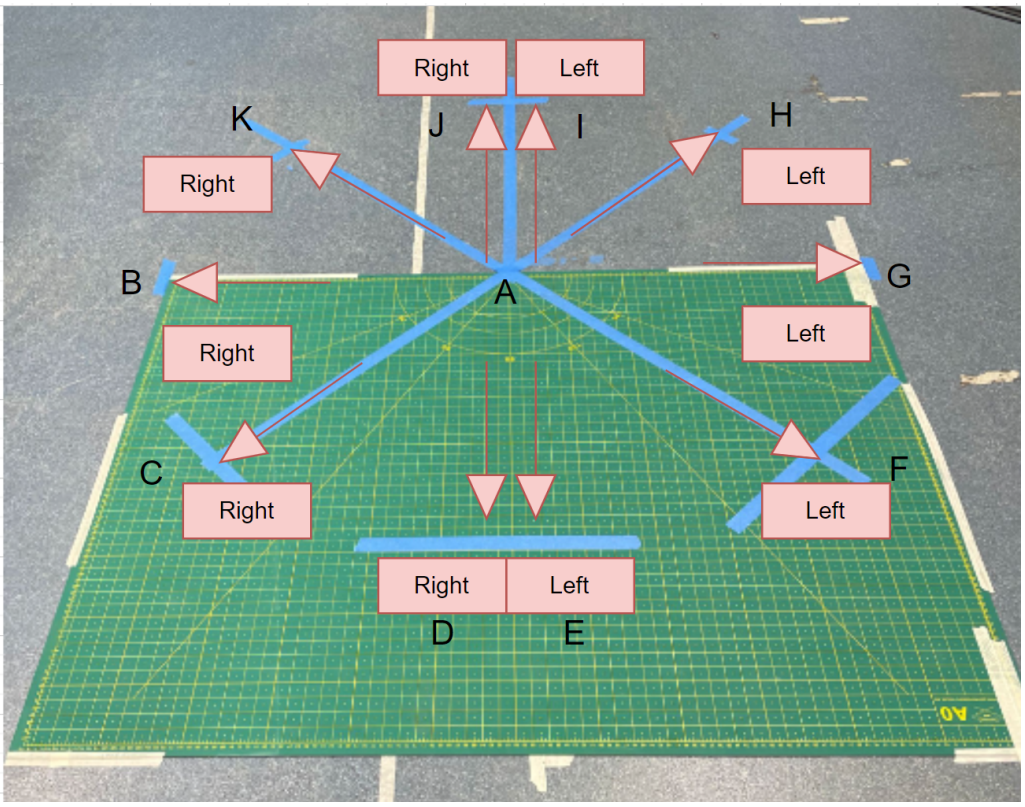


Figure 2.4: the stationary position A and points B to K to which the subjects place their right and left feet. The diagonal line have 45 degrees to the base line.

2.2.4 Data collection

As mentioned earlier, four markers are placed on each foot of all subjects for Qualisys measurements. When comparison with RealSense results are carried out, the point situated on the average value of these four markers is considered to be the step position for each foot. In RealSense RGB images, the two points for both feet, are points 11 and 14, shown in Figure 2.5. As It was practically not possible to stick markers on these 2 points (11 and 14), the above average method is applied.

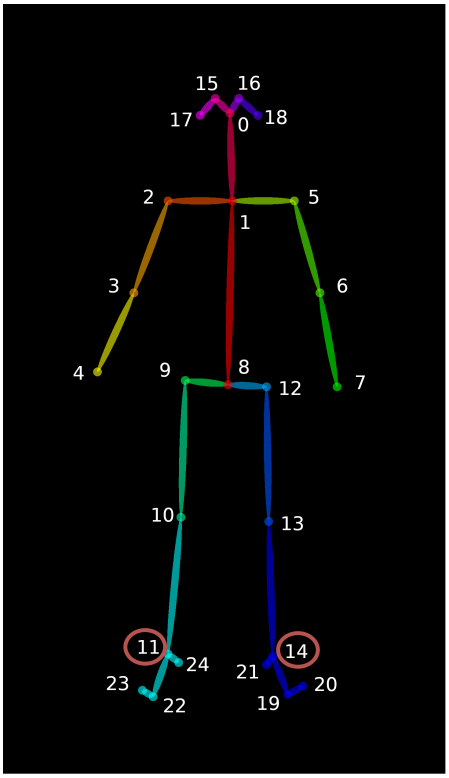


Figure 2.5: the keypoints considered by OpenPose as human body anatomy [15]. For Intel RealSense, points 11 and 14 are of our interest as right and left step positions respectively.

To collect data of the step movements, for Intel RealSense, the developed *Python program* is manually started at the beginning of the movements and terminated at the end of the movements. The Qualisys data is also collected by *QTM*. Following section explains how the data is treated.

2.2.5 Data Analysis

According to the manufacturer, the RealSense coordinate system (RCS) is defined [20], and as shown in Figure 2.2.

The RealSense camera is situated on the Z axis with the negative direction towards the camera. The Y and X axis, both perpendicular to Z, are vertical (negative up) and parallel to the floor, respectively.

The obtained data from RealSense camera are all transferred to the Qualisys coordinate system, in other words Qualisys system is chosen as the world coordinate system (WCS). As seen in Figure 2.2 the origin of the WCS is on point A, which is the start point of the step movements. The positive direction of X is toward the camera, and the Y and Z are perpendicular to X in vertical (positive up) and parallel to the floor, respectively.

In order to transfer the data outputs of RCS to the WCS, first the origin of the WCS is looked at by the RealSense camera and the obtained coordinates i.e. x0, y0 and z0 are considered as the offset of the origins of 2 systems.

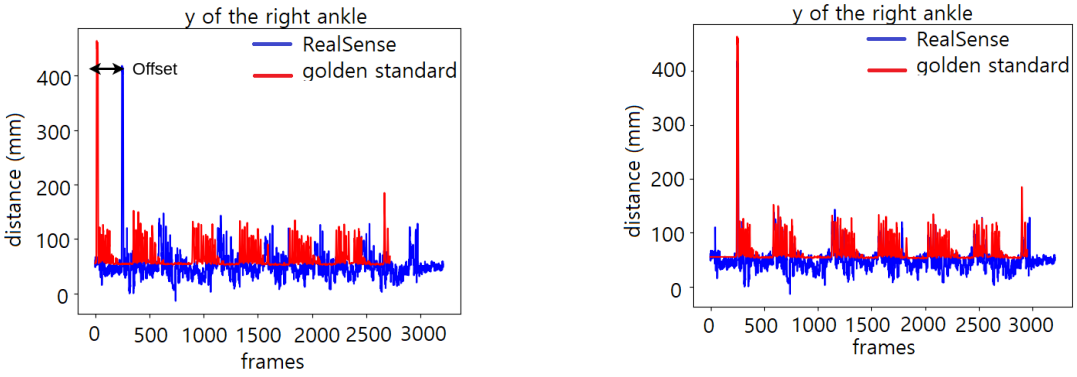
$$x_{WCS} = -z_{RCS} + x0 \tag{2.1}$$

$$y_{WCS} = -y_{RCS} + y0 \tag{2.2}$$

$$z_{WCS} = -x_{RCS} + z0 \tag{2.3}$$

The values for the x0, y0 and z0 are found to be 3830 mm, 700 mm and 69 mm, respectively.

Time aligning of the two systems is done by imaging a specific movement i.e. raising of the right foot for both systems as shown in Figure 2.6a. As seen in the figure, RealSense has a time-lag that we call it time offset. The value of time offset is measured up to millisecond by using the time stamping function of RealSense. The Qualisys data are shifted in time axis with offset size as shown in Figure 2.6b.



(a) time Offset between golden standard and RealSense **(b)** offset shift corrected

Figure 2.6: time-lag of RealSense

The recording rate of Qualisys is at a steady value of 128 hertz, however our general framework recording rate is 14 frames per second. This means that data alignment is to be done when comparing the images of two systems. Although, the RealSense camera is capable of capturing up to 90 frames in one second, but due to limitation of the employed OpenPose software in handling the data, 14 hertz imaging frequency is used.

Knowing the difference in framing frequencies in two systems, time synchronization is also needed when selecting the compatible frame of Qualisys. To find the compatible frame number in Qualisys, the time stamped for RealSense frame is multiplied by 128.

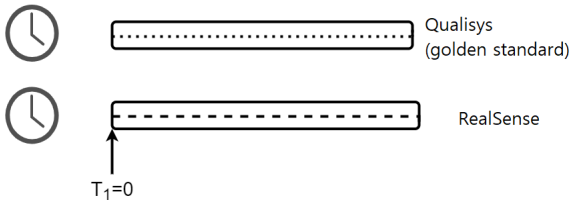


Figure 2.7: frequencies of the two systems

After we have dealt with time offset and space offset, the main data analysis is worked on. Step movements of each subject that is of our interest take place in X-Z plane as shown in Figure 2.8. The data for step locations of each subject is drawn on series of graphs for X and Z directions for both Qualisys and RealSense systems. The graphs are presented for right and left ankles separately. For instance, in Figure 3.1a, for subject number 1, step locations of the right foot in X direction is shown for two consecutive cycles. The horizontal axis is the frame number and the vertical is the related data for each frame in millimetre. Each peak value demonstrates the destination of one step i.e. step location. For example, in this graph steps locations 2 and 3 show positive peaks and step locations 9 and 10 show negative peaks.

Similar results are drawn in Figure 3.1b, for Z direction.

For some steps (3, 4, 8 and 9) in both right and left movements, the Z values are 0 and X enjoys its maximum amount. Vise versa, this occurs for some other steps (1 and 6) where their X values are 0 and have maximum Z amount. As it is seen in the Figure 2.8, diagonal steps (2, 5, 7 and 10) have components in both X and Z directions.

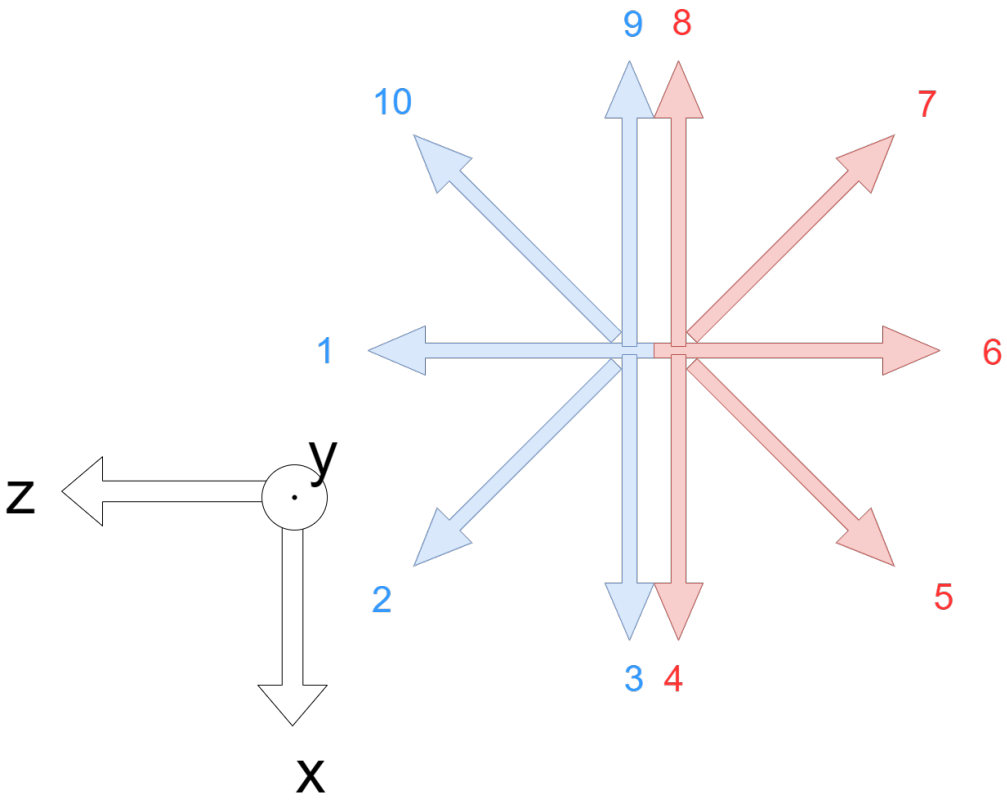


Figure 2.8: step directions 1 to 10 shown in XZ plane. Blue arrows show the right foot and the red ones show the left foot.

To evaluate the accuracy of RealSense D435i by comparing the step position measurements with golden standard results, the peak values in the obtained graphs (which represent the step locations) are of the most important data. The data obtained while the subjects are stationery, do not contain any significance. To this end, an algorithm is designed by the author to get rid of the worthless data and securely save the most valuable results. In the designed filter, main idea is to specify the important points by checking the difference in the values obtained from two consecutive frames. In this way, marking the points on the slopes towards the peaks and away from the peaks are guaranteed by the procedure used. Two significant aims are achieved by employing this filter. Firstly, noise fluctuations can not be mistaken as the time series peak value. Secondly, a proper average value is selected to represent step location in each movement.

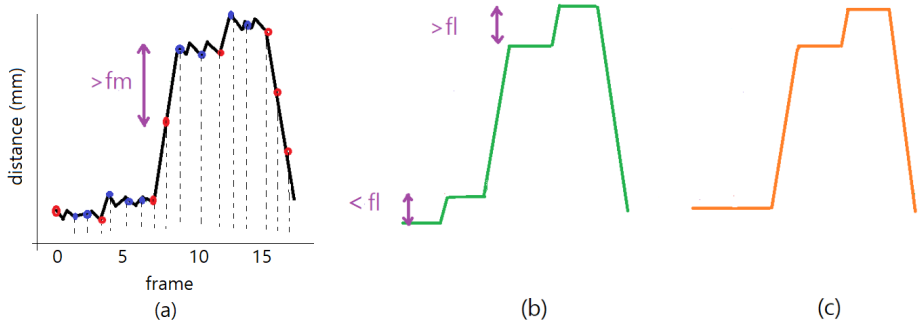


Figure 2.9: schematic presentation of data filtering, a) showing important (red) and non-important (purple) points b) after implying the first phase c) final results, the second phase applied.

Two parameters of *fm* and *fl* are introduced in the developed filter. They stand as the parameters for keeping the significant frame data and devaluation of non-significant frame data respectively. The value of the parameter *fm* is estimated by an initial survey of the results. The chosen value for *fm* is one third of the lowest step. In other words, the lowest step is 330 millimeters and the selected number for *fm* is 110 milliliters. As it is seen in Figure 2.9, if the absolute difference in values of two consecutive frames is greater than *fm* first point is considered as an important point. The first and the last points for each subject are also considered as important points. After all important points are marked, the values between them (including important points) are averaged and they are all replaced by this average value. Apparently, except for the first and last important points, all others have to take two of these average values, one from previous and one from subsequent frames. Therefore, at these important points a jump appears in the new plot. Now, *fl* comes into play. If the difference of two consecutive average values (the jump height) is less than the *fl*, the second average is replaced by the first value. The value for *fl* is estimated by surveying the graphs after applying *fm*. Surveying means choosing a value for *fl*, to flatten stationary part. The selected value is 40 millimeters for this parameter. These selected values for *fm* and *fl* are applied to all cycles of all subjects. Subsequently, the final stepping locations are determined from the time series. Figure 3.3 illustrates an example of the derived stepping locations for different steps.

After finding the step positions, the comparison of the results with Qualisys golden standard are carried out and the accuracy of the RealSense results are worked out. The absolute value of error is found from the following formula:

$$Error\ percent = |(\frac{GoldenStandard\ value - RealSense\ value}{GoldenStandard\ value})| \times 100 \tag{2.4}$$

It must be noted that the accuracy percent is 100 percent minus the error percent.

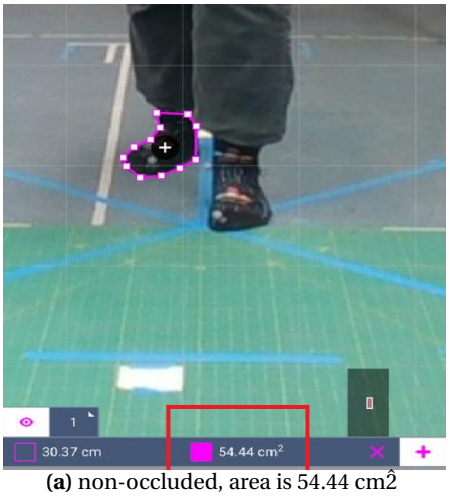
Apart from investigating the accuracy, in order to have a better evaluation of the measurements obtained from RealSense camera, Signal to Noise Ratio (SNR) is derived for all step positions of all subjects. SNR values are calculated and shown in the graph of Figure 3.2c, RealSense after filtering, for the significance values (peak values) which represent the step positions. The filtered value on the peak is taken as signals and the average absolute diversity from this value is considered as noise like the Equation 2.5.

$$SNR = \frac{signal}{noise} = \frac{signal\ without\ noise}{signal\ with\ noise - signal\ without\ noise} \tag{2.5}$$

The presented results in the next Section are based on the averaged values between all cycles for each subject. In further processing, the final values are derived also by averaging between subjects.

The method of calculating the error data for showing in the box plots, which is shown is Figure 3.5, is based on the median, first quartile, third quartile and outliers (read more on the A.8).

To assess the effect of occlusion on the results accuracy, the occlusion is quantized. Occlusion intensity is calculated at the end of each step were the maximum occlusion for that step occurrs. After all, we are seeking the occlusion effect on the step measurement i.e. ankle position at the end of each cycle. The quantization is performed employing [1] software, by calculating the area which is occluded in the image and finding its percentage relative to the area if it was not occluded at all, as it is seen in Figure 2.10. In other word, if it was not occluded at all it has 0 percent of occlusion, and vice versa it has 100 percent of occlusion.



(a) non-occluded, area is 54.44 cm$\hat{2}$

(b) occluded, area is 34.71 cm$\hat{2}$

Figure 2.10: not occluded and occluded image

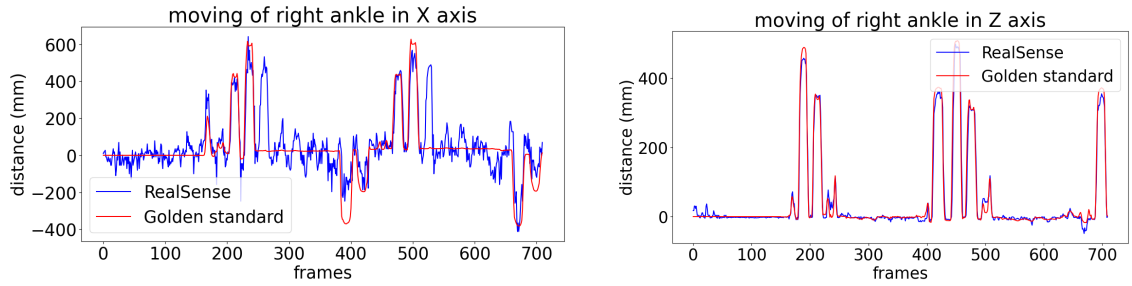
Therefore, the occlusion area comes from this formula:

$$Occlusionpercent = (\frac{area\ of\ image\ without\ occlusion - area\ of\ image\ with\ occlusion}{area\ of\ image\ without\ occlusion}) \times 100 \tag{2.6}$$

3 Results

The results are successfully attained by implementation of the general framework, described in Section 2.1. Stepping locations are determined and presented in both RealSense and golden standard systems. Further, accuracy and SNRs are derived for the RealSense data, followed by the results related to the occlusion issue.

The time series of the right ankle, measured with the RealSense and golden standard, are shown in Figure 3.1 for X and Z components (X_r and Z_r). Using the step numbering shown in Figure 2.8, the peaks in X_r represent steps 2, 3, 9, and 10. The peaks in Z_r stand for steps 1, 2, and 10. Graphs for the left ankle (X_l and Z_l) are not shown. The peaks in X_l represent steps 4, 5, 7, and 8. The peaks in Z_l stand for steps 5, 6, and 7.

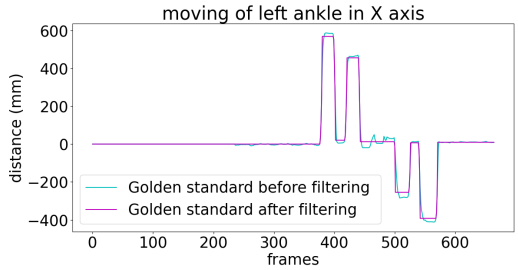


(a) X_r **(b)** Z_r

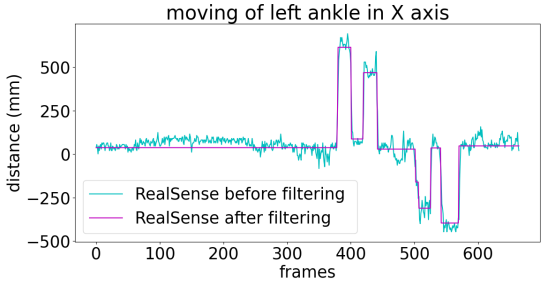
Figure 3.1: X and Z components for right ankle, two cycles of subject 1.

The acquired data has been filtered by a method (as described in Section 2.2.5) developed by the author of this project. The two aims, expected by using this filter, are fulfilled, i.e. no noise fluctuation is mistaken as peak value and a proper average value is determined as step location measurement.

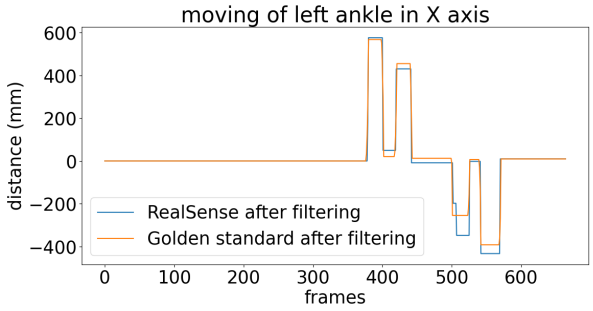
The golden standard and the RealSense signals before and after filtering are shown in Figure 3.3. These filtered results have been used for accuracy calculations. SNRs are derived by using the signal and noise values of the step positions. Figure 3.2b is used for finding SNR values on the significant locations, i.e. peaks.



(a) non-filtered vs filtered as acquired with golden standard



(b) non-filtered vs filtered as acquired with RealSense



(c) RealSense and golden standard signals after filtering

Figure 3.2: time series of left ankle positions in X axis before and after filtering for subject 2

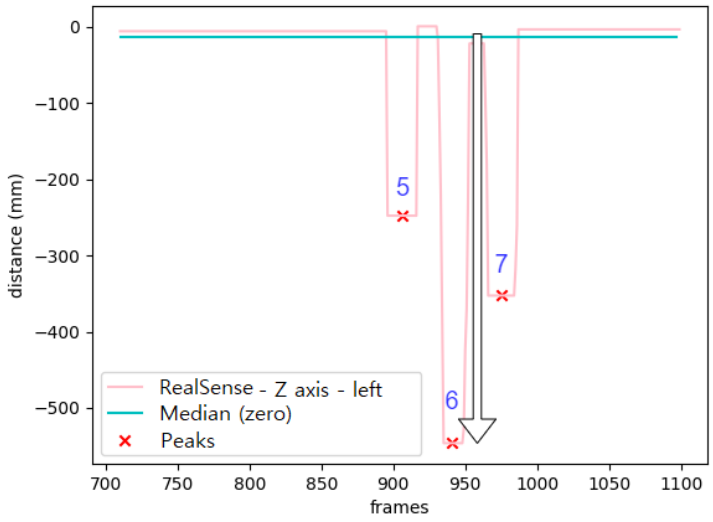


Figure 3.3: Z components of step positions 5,6 and 7, frames 710 to 1100 (cycle 2 of subject 5), shown in red crosses on the pink peaks. The blue line considered as zero is median of stationery data.

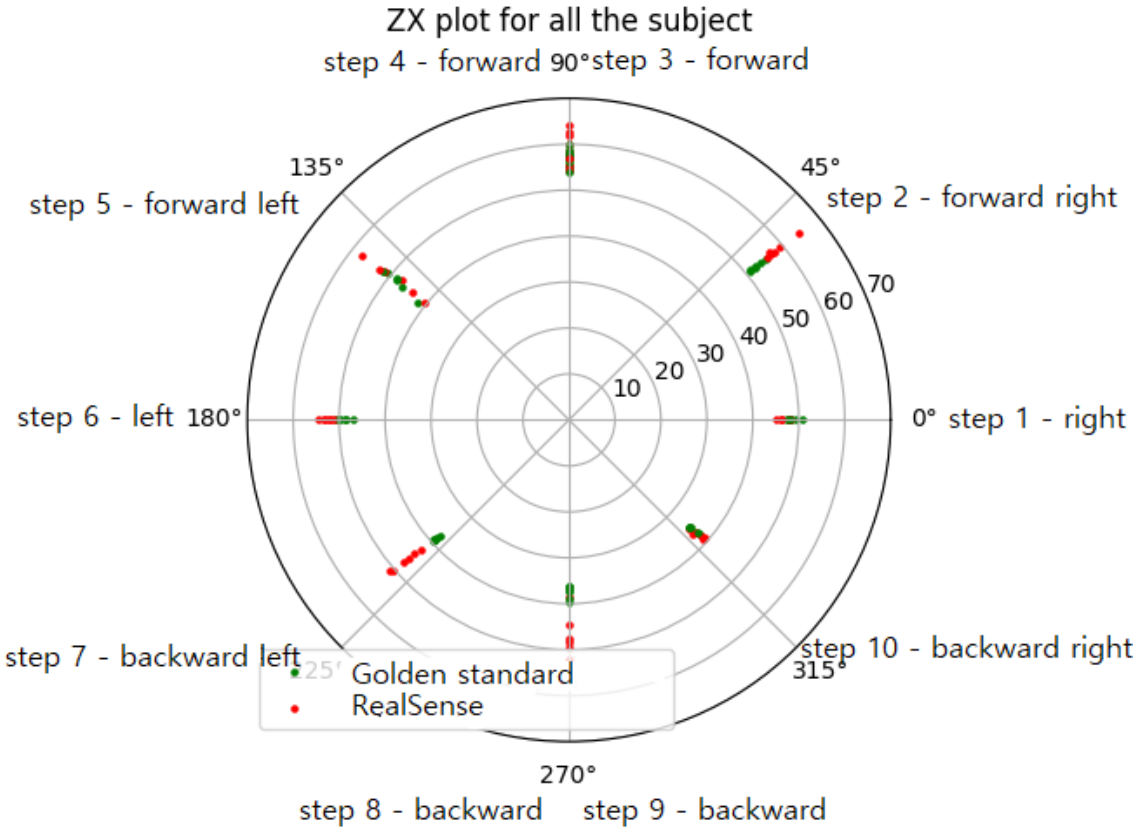


Figure 3.4: steps locations shown in zx plane for all subjects. The beginning point for all is zero and the ending points shown in dots.

The obtained results are shown in ZX plane in Figure 3.4. The center represent the stationary position of the subjects and the redial values stand for step measurements of both systems.

In Figure 3.5, the box plot for distribution of errors are shown for all subjects.

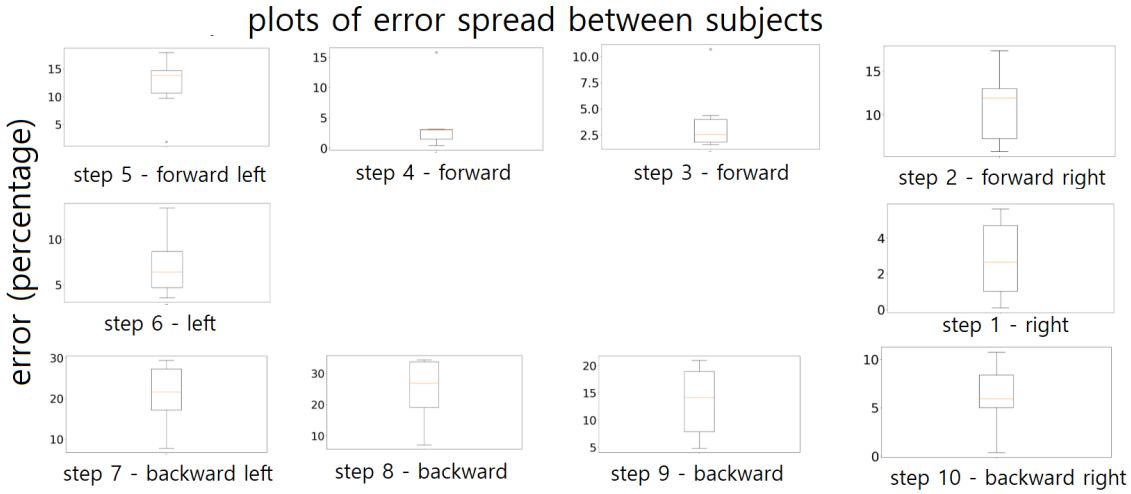


Figure 3.5: box plot shown for each step separately, all subjects are considered.

The final results are presented in Table 3.1. As shown, the step position values for both systems, RealSense outputs accuracy and its SNR are presented. All the values are derived according to the methods explained in the previous chapter.

Steps	Golden standard (cm)	RealSense (cm)	Error (percentage)	SNR (-)
Step 1 - right	48.59	46.92	3.43	20.85
Step 2 - forward right	52.67	58.49	11.03	17.02
Step 3 - forward	57.08	58.18	1.93	18.9
Step 4 - forward	56.94	58.06	1.97	19.73
Step 5 - forward left	47.93	48.99	2.91	15.08
Step 6 - left	49.18	52.71	7.17	16.49
Step 7 - backward left	38.89	47	20.86	15.05
Step 8 - backward	37.54	46.79	24.64	8.97
Step 9 - backward	37.92	41.57	9.63	6.94
Step 10 - backward right	36.14	37.5	3.79	14.39

Table 3.1: final results of the step locations and evaluation of RealSense outputs i.e. accuracy and SNR

In Table 3.1, all the errors are less than 12% except the step 7, step 8 with almost 22%. Regarding SNRs, the best result is obtained for steps 1, 2, 3 and 4 which stands on the values between 17 to 21. The least SNR values are seen to be for steps 8 and 9.

Since there is no occurrence of the occlusion for steps 1, 2, 5, 6, 7 and 10, only the effect of occlusion intensity on steps 3, 4, 8 and 9 are investigated. For forward steps (3 and 4), occlusion causes a fake peak and for backward steps (8 and 9) the effect is seen as a missing peak. These are shown in Figures 3.6 and 3.7.

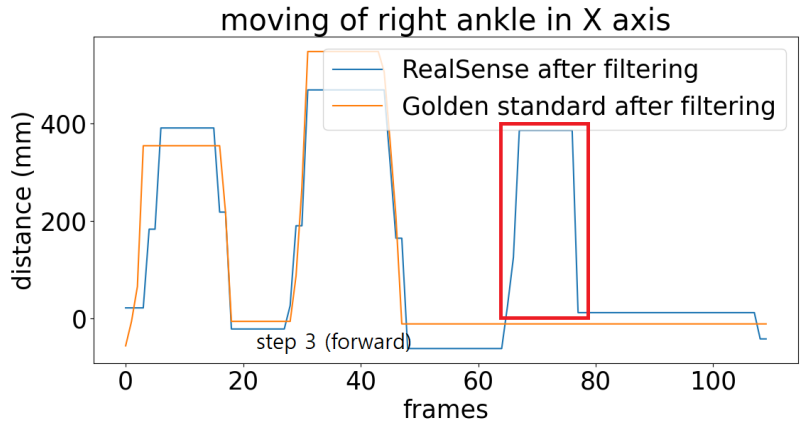


Figure 3.6: fake RealSense peak as an effect of occlusion in the forward step 4, subject 1 cycle 3

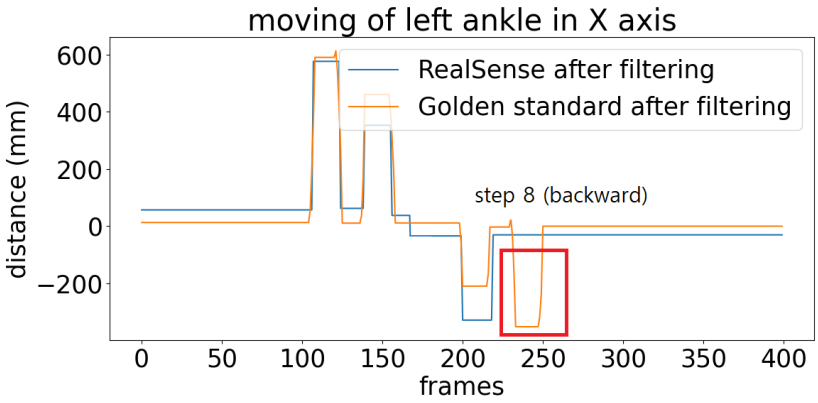


Figure 3.7: missing RealSense step as an effect of occlusion on the backward step 8, subject 3 cycle 3

As mentioned earlier, the occlusion issue needs to be investigated for steps 3, 4, 8 and 9. The findings are shown in Tables 3.2 and 3.3.

	subject 1	subject 2	subject 3	subject 4	subject 5	subject 6	times of occlusion in each step
step 1	0	0	0	0	0	0	0
step 2	0	0	0	0	0	0	0
step 3	0	0	1	0	4	0	5
step 4	5	0	5	5	5	3	23
step 5	0	0	0	0	0	0	0
step 6	0	0	0	0	0	0	0
step 7	0	0	0	0	0	0	0
step 8	4	0	5	2	3	1	15
step 9	5	1	5	2	3	0	16
step 10	0	0	0	0	0	0	0
times of occlusion in each subject	14	1	16	9	15	4	

Table 3.2: Occlusion times for each subject and each step.

	cycle 1 (percentage)					
	subject 1	subject 2	subject 3	subject 4	subject 5	subject 6
step 3	0	0	0	0	0	0
step 4	73	0	79.69	74.01	33.29	25.96
step 8	10.74	0	25.25	0	0	0
step 9	20.31	6.16	93.09	0	15.05	0
	cycle 2 (percentage)					
	subject 1	subject 2	subject 3	subject 4	subject 5	subject 6
step 3	0	0	8.55	0	20.18	0
step 4	74.86	0	82.28	46.26	71.18	0
step 8	11.59	0	26.82	0	23.01	0
step 9	4.6	0	47.21	7.66	2.04	0
	cycle 3 (percentage)					
	subject 1	subject 2	subject 3	subject 4	subject 5	subject 6
step 3	0	0	0	0	31.17	0
step 4	45.44	0	87.85	72.9	56.59	17.5
step 8	0	0	42.47	7.62	31.17	0
step 9	15.94	0	61.67	0	3.1	0
	cycle 4 (percentage)					
	subject 1	subject 2	subject 3	subject 4	subject 5	subject 6
step 3	0	0	0	0	25.82	0
step 4	55.31	0	73.81	82.9	56.87	30.67
step 8	8.75	0	29.85	4	5.11	6.75
step 9	8.48	0	44.63	15.07	0	0
	cycle 5 (percentage)					
	subject 1	subject 2	subject 3	subject 4	subject 5	subject 6
step 3	0	0	0	0	28.85	0
step 4	96.62	0	76.66	80.89	33.02	0
step 8	14.65	0	21.75	0	0	0
step 9	17.9	0	50.09	0	0	0

Table 3.3: Occlusion percentage for cycles of all subjects.

4 Discussion

The results are discussed here and the challenges that appeared in this project are explained. Finally, the recommendations for future work are presented.

As it can be seen from the results shown in Table 3.1, forward steps 3 and 4 have the highest accuracy among all steps when compared to the golden standard. Those are errors of 1.93% and 1.97%, respectively. As it also can be seen, the worst accuracy belong to backward steps of 7 and 8, which show the errors of 20.86% and 24.64%, respectively. Therefore, it can be concluded that although the results are satisfactory for forward steps, they are not proper for backward steps. Larger distance to the camera can be explained as the main reason for lower accuracy of backward steps.

Regarding RealSense accuracy in the previous works, however, no data can be seen in the literature survey about step position accuracy of D435i for pose analysis purpose, in the RealSense data-sheet, the depth accuracy of D series is explained to be less than 2% for the distance of 2 meters. Further, in the previous work performed by [9], the depth accuracy of 400 Series cameras (including RealSense D435i) is obtained to be between 2.5 mm to 5 mm when the object has the distance of 1 meter. In the mentioned study, when the distance goes further than 1 m, RealSense D435i shows more accuracy drifts.

As the existing noise on the output data effects the final accuracy, the following noise discussion is presented. OpenPose is based on the deep learning algorithm for human pose detection and is supervised by manually labeled data. Therefore, the existence of some noise is unavoidable due to inherent noise in the training data [14]. Generally, the output data noise consists of RealSense camera noise, OpenPose noise, and data acquisition system noise. SNR results, which is calculated using final data, are shown in the last column of Table 3.3. Considerably lower SNR is seen for steps 8 and 9. Again, the reason is the larger distance to the camera that these steps have.

The RealSense images are also investigated regarding the existence of outliers and dispersion of the error results in box plot. The box plot given in Figure 3.5 shows more sporadic error data for some subjects when compared to the others. For instance, subject 4 and subject 6 have obvious outlier error data in the forward steps 3 and 4, respectively. It is seen that the mentioned steps for these subjects are due to human error. The cause of error dispersion in backward steps is concluded to be larger distance.

Occlusion is reported previously by some other researchers, working on RealSense depth camera, as a source of error reference [26]. Here, it depends on if the steps are forward (steps 3 and 4) or backward (steps 8 and 9) different occlusion effects in the results are detected. In forward step 4, the occlusion effect appears as a fake peak in the time series graph. In forward step 3, none of the subjects show any remarkable occlusion. The undesirable effect of occlusion in forward step is due to wrong keypoint chosen by OpenPose. In other words, OpenPose chooses a point on the moving foot instead of the real stationary ankle keypoint which is occluded. In backward steps 8 and 9, occlusion effect appears as missing the peak for these steps. In this case, OpenPose can not see the occluded ankle keypoint, therefore missing its movement. Some examples of the above mentioned effects of occlusion are followed. For cycle 3 of subject 1, forward step 4 with occlusion of 73% shows false peak as seen in Figure 3.6. A case of occlusion effect in backward step can be seen in cycle 3 of subject 3, step 8, with occlusion of 42.47% shows the missing peak as seen in Figure 3.7. Those cycles of the subjects which resulted in fake peak or missing peak due to occlusions, have been left out in calculations of accuracy and SNRs. These unexpected results which are caused by the occlusion are not seen on the golden standard graphs, thanks to employing multi cameras.

In OpenPose, apart from the noise, there is a chance of considering other objects with shapes close to human anatomy as a person. To reduce the probability of OpenPose mistaking the human body with other objects in the environmental scene certainty setting is important. We let OpenPose set the certainty automatically and then the highest certainty is fed into back projection program.

The physical differences between the subjects, i.e. foot size, and also differences in their performance in taking the steps, could be considered as some parameters affecting measurements and resulted error values.

Limitation of frame frequency caused by OpenPose could also have a role in results error. For instance, as seen in Figure 4.1, low framing frequency has caused blurring in the image, which in turn resulted in wrong finding of the ankle point by OpenPose. The effect of low framing in the resulting errors is also referred to by some previous research works [17]. Due to the limitation in data volume that OpenPose can have, reducing image resolution would help to increase the rate of framing. This is recommended for further research.

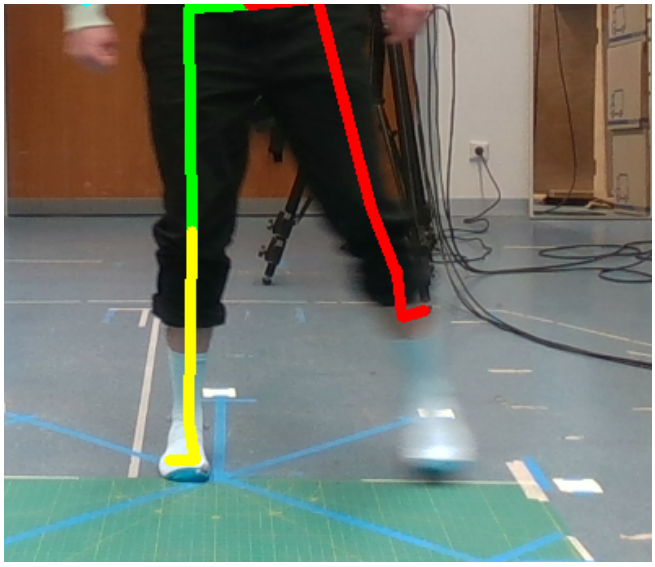


Figure 4.1: bad detection of depth due to the motion blur

Another recommendation for further study is to use another RealSense at the rear side of the subject to record the steps, which are called backward steps in this study. The latter camera can be dedicated to the backward steps (forward for the new camera). This would reduce the error seen in backward steps and make accuracy results more homogeneous. The RealSense manufacturer has already presented a setup for employing multiple cameras [13], which could be applied for this thesis special topic i.e. step location measurement.

Additionally, investigating the effect of subject distance from the camera, step size and stepping speed on the result accuracy is also recommended for further work.

5 Conclusion

The selected RealSense D435i depth camera has the basic capabilities and could act as a less expensive and less time-consuming sensor for measuring human movements. Therefore, use of the camera for pose analysis is feasible and the camera can be employed in the HERoS project. It is also concluded that better accuracy and higher SNR is obtained for measuring forward step locations.

Although the results show differences between step locations measured by Qualisys and RealSense, the setups of the two systems should be considered. Only one RealSense camera is used, imaging frontal plane, against 12 Qualisys camera, imaging 360 degrees. Thus, at least 2 RealSense cameras are needed if seeking efficient measurements in all directions. Occlusion effects can be totally avoided if certain rules are instructed to the subjects while imaging.

A Appendix A

A.1 Camera D435i

Stereo depth camera, which D400 Intel RealSense series are examples of that, contains two IR sensors with a certain distance from each other. As the distance between the two mentioned sensors is known, comparison of the images would result in depth information. Apart from the sensors, an IR projector is also provided on D400 series. Although these camera series work well in most lighting conditions, in the condition of low light the IR projector would torch the object and therefore still could the depth details be obtained [10].

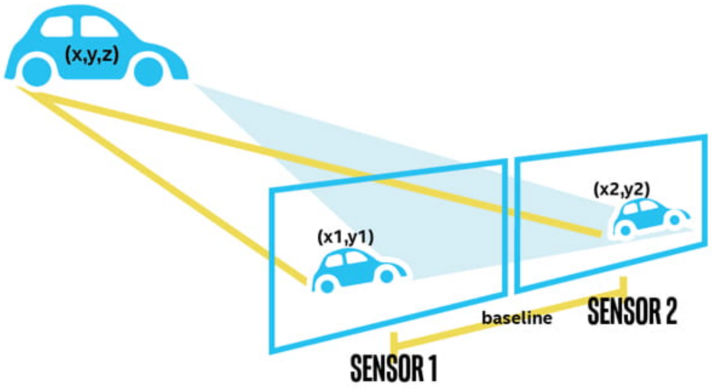


Figure A.1: stereo Depth used in D435i [2]

The Intel RealSense D400 series integrates vision processor, stereo depth module, RGB sensor with color image signal processing and inertial Measurement Unit (IMU). The series is supported with cross-platform and open source Intel RealSense SDK 2,0. In D435i version, shown below, the features include up to 1280x720 active stereo depth resolution and up to 1920x1080 RGB resolution. Depth diagonal field of view in D435i is over 90 degrees, it carries dual global shutter sensors for up to 90 FPS depth streaming and the installed IMU for 6 degrees of freedom. Further, D435i ranges 0.2m to over 10m varying with lighting conditions. D435i version has a dimension of 90 mm x 25 mm x 25 mm weighs 72 grams and System Interface Type is USB 3 [9].

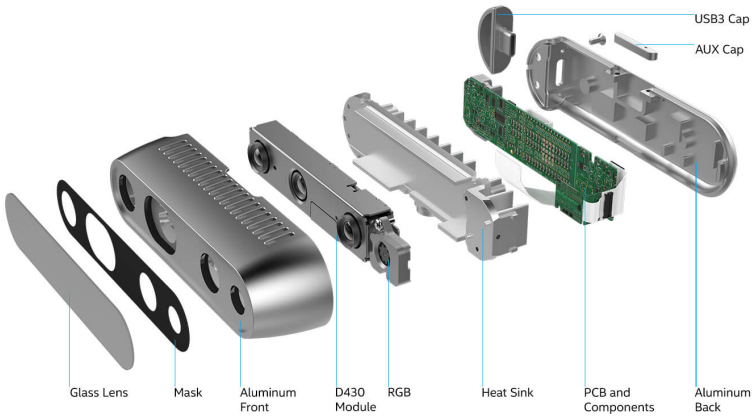


Figure A.2: the camera D435i [9]

In the following tables depth field of view and accuracy of D435i is compared with other versions of D400 series:

Format	D400/D410/D415	D420/D430/D435/D435i
Horizontal FOV (VGA 4:3)	48	74
Vertical FOV (VGA 4:3)	40	62
Diagonal FOV (4:3)	60	88
Horizontal FOV (HD 16:9)	64	86
Vertical FOV (HD 16:9)	41	57
Diagonal FOV (HD 16:9)	72	94

Figure A.3: field of view of D series-from data sheet-[9]

Metric	D400/D410/D415 (up to 2 Meters and 80% FOV)	D420/D430/D435/D435i (up to 2 Meters and 80% FOV)
Z-accuracy (or absolute error)	≤ 2%	≤ 2%
Fill rate	≥ 99%	≥ 99%
RMS Error (or Spatial Noise)	≤ 2%	≤ 2%
Temporal Noise (Pixel)	≤ 1%	≤ 1%

Figure A.4: depth specification of D series-from data sheet-[9]

Also, the X-Y dimensions shown below:

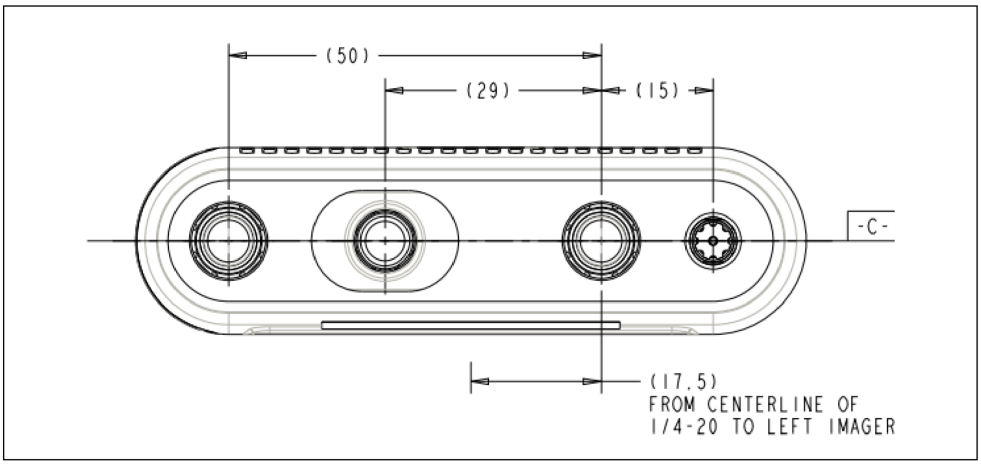


Figure A.5: X and Y in the camera-from the data sheet-[9]

A.2 Calibration of Intel RealSense D400 series

Dynamic calibration is in fact optimizing the extrinsic parameters. Intrinsic parameters are focal length, principal point and distortion while extrinsic parameters are rotationLeftRight, translationLeftRight, rotationLeftRGB and translationLeftRGB. Dynamic calibration is classi-

fied into two categories: Target dynamic Calibration (Depth Scale Calibration) and target-less Dynamic Calibration (Rectification Calibration).

Target Dynamic Calibration is aligning the depth frame to the changes in position of the optical elements [8]. This type of calibration is used in the present study. It is notable that cable with USB 3.1 connector must be used. This cable which connects the PC to the camera should be able to stand 3 A current and transferring 5 Gbit/s of data.

A.3 OpenPose

In post camera processing of data, most approaches have used a top-down strategy that first detects people and then have estimated the pose of each person independently on each detected region. In bottom-up method first the body parts are dealt with and then the people. Bottom-up system achieves higher accuracy and real-time performance when compared with top-down method.

OpenPose is an efficient bottom-up method for pose estimation with competitive performance. It is the first bottom-up representation that employs Part Affinity Fields (PAFs), a set of 2D vector fields that encode the location and orientation of limbs over the image domain. To obtain reliable local observations of body parts, Convolutional Neural Networks (CNNs) have been used, and the accuracy have been significantly boosted on body pose estimation. However, they had not enough accuracy when the foot only data set is considered [3].

Figure A.6: The bottom-up approach using Part Affinity Fields to detect the limb [3]

A.4 Back projection

When an ordinary camera records an image it converts the 3D real world object to a 2D image. In other words, it projects the object on a 2 dimensional plane. In a projected image all points carry an x and y representing their position on the plane. Obtaining the depth of points on the object, would enable us to associate a third coordinate Z to all the points and enables us to make a 3D picture as it originally was in real world. This latter process is called back projection or inverse projection.

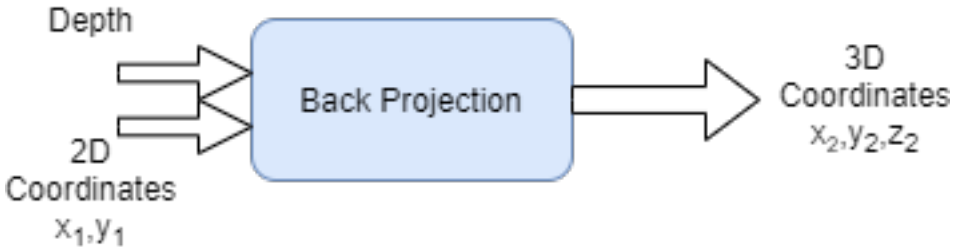


Figure A.7: back projection

It is essential in many computer vision applications to understand depth perception. For instance, for an autonomous vehicle finding the depth allows for better decision making as the agent is fully aware of the separation distances between other vehicles and pedestrians.

[2].

A.5 Exergaming

Fitness game, exergaming or exer-gaming is a term used for the games rely on the technology that track human movement and reaction ed based on seeing where the limb is located [4].

A.6 Mocap System

Capturing real life motion and showing it in 3D space has wide application in realistic simulation of human motion. One usual method for this Optical Motion Capturing (OPC) is to place markers in strategic locations on the body and employ several synchronized cameras to capture them. The markers locations are designed such that covers the required body joints and bone segments. In research performed by Gutemberg B et al [7] 49 white rectangular markers attached to black body suit contrasted with a white background as shown below:

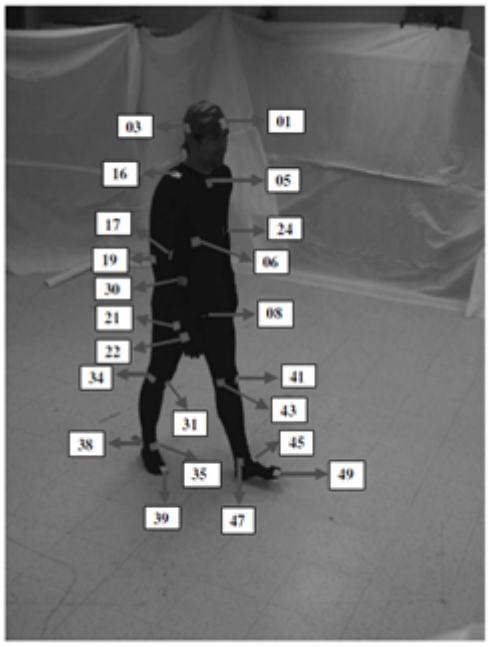


Figure A.8: Mocap system [7]

A.7 Comparing noise filters vs my filter

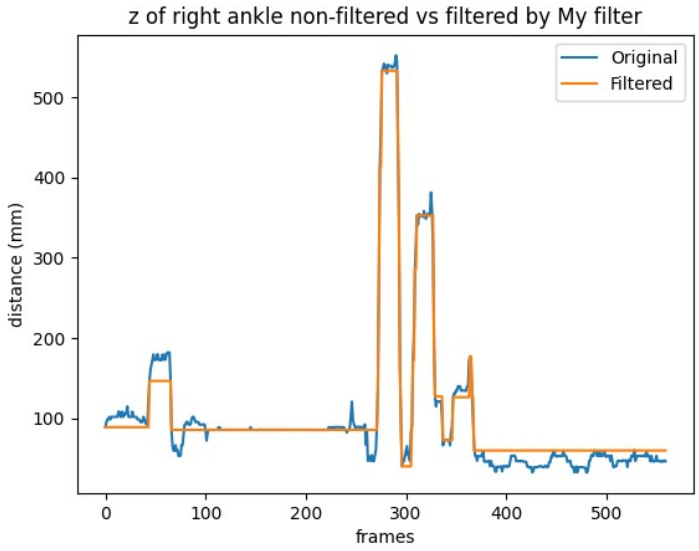


Figure A.9: My filter

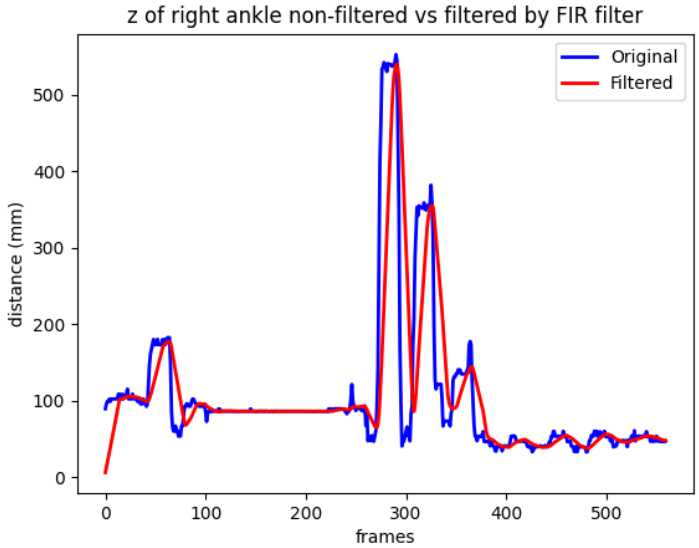


Figure A.10: finite impulse response filter (FIR)

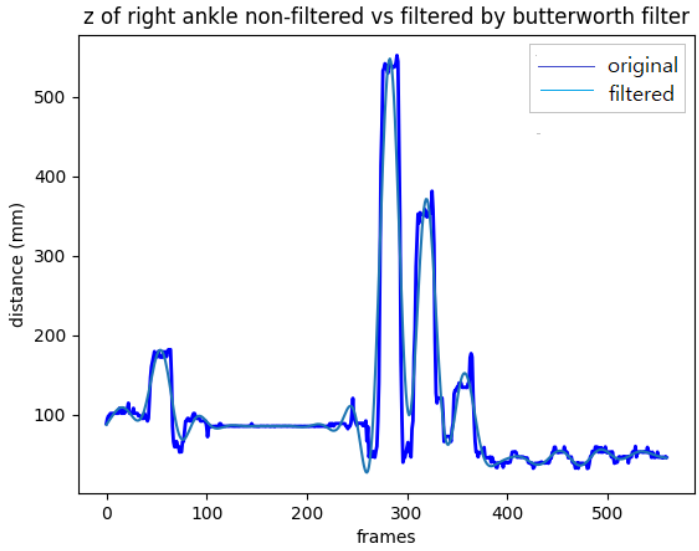


Figure A.11: butter-worth filter

In comparing the figures, figure (A.10) seem to be a bit delayed. Then we have tried the butterworth filter which is common in analysing movement kinematics. This filter is also a bit advanced and a bit not following the original curve.

A.8 Box plot

Median (Q2/50th Percentile): the middle value of the dataset.

First quartile (Q1/25th Percentile): the middle number between the smallest number (not the "minimum") and the median of the dataset.

Third quartile (Q3/75th Percentile): the middle value between the median and the highest value (not the "maximum") of the dataset.

Interquartile range (IQR): 25th to the 75th percentile.

Whiskers (shown in blue)

Outliers (shown as green circles)

"Maximum": Q3 + 1.5*IQR

"Minimum": Q1 -1.5*IQR

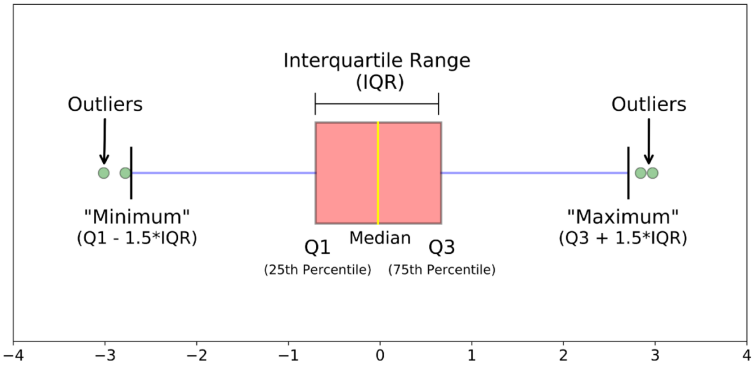


Figure A.12: different parts of a box plot [27]

Bibliography

[1] *Area Calculator App | Irregular Area Calculator | Calculate Irregular Shape Area.* URL: https://www.sketchandcalc.com/.

[2] *Beginner's guide to depth (Updated) – Intel® RealSense™ Depth and Tracking Cameras.* URL: https://www.intelrealsense.com/beginners-guide-to-depth/.

[3] Zhe Cao et al. "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.1 (2021), pp. 172–186. ISSN: 19393539. DOI: 10.1109/TPAMI.2019.2929257.

[4] *Exergaming - Wikipedia.* URL: https://en.wikipedia.org/wiki/Exergaming.

[5] Daphne Geerse et al. "Validation of Foot Placement Locations from Ankle Data of a Kinect v2 Sensor." In: *Sensors (Basel, Switzerland)* 17.10 (Oct. 2017). ISSN: 1424-8220. DOI: 10.3390/S17102301. URL: https://pubmed.ncbi.nlm.nih.gov/28994731/.

[6] Daphne J Geerse, Bert H Coolen, and Melvyn Roerdink. "Kinematic Validation of a Multi-Kinect v2 Instrumented 10-Meter Walkway for Quantitative Gait Assessments." In: (2015). DOI: 10.1371/journal.pone.0139913. URL: http://www.nwo.nl],.

[7] Gutemberg B Guerra-Filho1. "Optical Motion Capture: Theory and Implementation." In: ().

[8] User Guide. "Intel ® RealSense ™ Product Family D400 Series Calibration Tools." In: July (2020).

[9] Intel. "Intel® RealSense™ Camera D400 series Product Family Datasheet Rev. 01/2019." In: January (2019). URL: https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-D400-Series-Datasheet.pdf.

[10] Intel Realsense. "Beginner's guide to depth (Updated) – Intel® RealSense™ Depth and Tracking Cameras." In: (2019). URL: https://www.intelrealsense.com/beginners-guide-to-depth/.

[11] Mahvash Jebeli, Alireza Bilesan, and Ahmadreza Arshi. "A study on validating KinectV2 in comparison of Vicon system as a motion capture system for using in Health Engineering in industry." In: *Nonlinear Engineering* 6.2 (2017), pp. 95–99. ISSN: 21928029. DOI: 10.1515/nleng-2016-0017.

[12] Jeison David Mejia-Trujillo et al. "Kinect™ and Intel RealSense™ D435 comparison: a preliminary study for motion analysis." In: *2019 IEEE International Conference on E-Health Networking, Application and Services, HealthCom 2019* October (2019). DOI: 10.1109/HealthCom46333.2019.9009433.

[13] *Multiple depth cameras setup guide – Intel® RealSense™ Depth and Tracking Cameras.* URL: https://www.intelrealsense.com/multiple-depth-cameras-setup-guide/.

[14] Nobuyasu Nakano et al. "Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras." In: *Frontiers in Sports and Active Living* 2 (2020). DOI: 10.3389/fspor.2020.00050.

[15] *openpose extract only the skeleton - Stack Overflow.* URL: https://stackoverflow.com/questions/59288435/openpose-extract-only-the-skeleton.

[16] Karen Otte et al. “Accuracy and reliability of the kinect version 2 for clinical measurement of motor function.” In: *PLoS ONE* 11.11 (2016), pp. 1–17. ISSN: 19326203. DOI: 10.1371/journal.pone.0166532.

[17] Alexandra Pfister et al. “Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis.” In: *Journal of Medical Engineering and Technology* 38.5 (2014), pp. 274–280. ISSN: 1464522X. DOI: 10.3109/03091902.2014.909540.

[18] Pornphom Piraintorn and Vera Sa-Ing. “Stroke Rehabilitation based on Intelligence Interaction System.” In: *17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2020* (2020), pp. 648–651. DOI: 10.1109/ECTI-CON49241.2020.9158104.

[19] Alwin Poulose and Dong Seog Han. “Hybrid indoor localization using IMU sensors and smartphone camera.” In: *Sensors (Switzerland)* 19.23 (2019). ISSN: 14248220. DOI: 10.3390/s19235084.

[20] *Projection in Intel RealSense SDK 2.0*. URL: https://dev.intelrealsense.com/docs/projection-in-intel-realsense-sdk-20.

[21] *pyrealsense2 — pyrealsense2 2.33.1 documentation*. URL: https://intelrealsense.github.io/librealsense/python_docs/_generated/pyrealsense2.html.

[22] *pyrealsense2.align — pyrealsense2 2.33.1 documentation*. URL: https://intelrealsense.github.io/librealsense/python_docs/_generated/pyrealsense2.align.html.

[23] A. Ramnemark et al. “Fractures after stroke.” In: *Osteoporosis International* 8.1 (1998), pp. 92–95. ISSN: 0937941X. DOI: 10.1007/s001980050053.

[24] Vladimir Tadic et al. “Effects of the post-processing on depth value accuracy of the images captured by RealSense cameras.” In: *Contemporary Engineering Sciences* 13.1 (2020), pp. 149–156. ISSN: 13136569. DOI: 10.12988/ces.2020.91454.

[25] Waimei A. Tai, Jared Conley, and Lucy Kalanithi. “Cost-saving innovations for acute ischemic stroke and transient ischemic attack.” In: *Neurology: Clinical Practice* 4.5 (2014), p. 427. ISSN: 21630933. DOI: 10.1212/CPJ.0000000000000081. URL: /pmc/articles/PMC5765688/%20/pmc/articles/PMC5765688/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5765688/.

[26] Li Fen Tu and Qi Peng. “Method of Using RealSense Camera to Estimate the Depth Map of Any Monocular Camera.” In: *Journal of Electrical and Computer Engineering* 2021 (2021). ISSN: 20900155. DOI: 10.1155/2021/9152035.

[27] *Understanding Boxplots. The image above is a boxplot. A boxplot... | by Michael Galarnyk | Towards Data Science*. URL: https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51.

[28] Mike Van Diest et al. “Exergaming for balance training of elderly: State of the art and future developments.” In: *Journal of NeuroEngineering and Rehabilitation* 10.1 (2013), p. 1. ISSN: 17430003. DOI: 10.1186/1743-0003-10-101. URL: Journal%20of%20NeuroEngineering%20and%20Rehabilitation.

[29] Dave Van Opstal, Vivian Weerdesteyn, and Jolanda Roelofs. “Validity of Microsoft Kinect™, compared to Vicon, for measuring leg angles at foot contact in healthy volunteers during backward stepping.” In: (2015).

[30] Oliver Wasenmüller and Didier Stricker. “Comparison of Kinect v1 and v2 Depth Images in Terms of Accuracy and Precision.” In: ().

[31] David Webster and Ozkan Celik. “Experimental evaluation of Microsoft Kinect’s accuracy and capture rate for stroke rehabilitation applications.” In: *IEEE Haptics Symposium, HAPTICS* February 2014 (2014), pp. 455–460. ISSN: 23247355. DOI: 10.1109/HAPTICS.2014.6775498.

[32] Vivian Weerdesteyn et al. “Falls in individuals with stroke.” In: *Journal of Rehabilitation Research and Development* 45.8 (2008), pp. 1195–1214. ISSN: 07487711. DOI: 10.1682/JRRD.2007.09.0145.

references.bib