Development and Validation of a Smart Hallway for Human Stride Analysis using Marker-less 3D Depth Sensors

Vinod Gutta

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Applied Science

in

Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

© Vinod Gutta, Ottawa, Canada, 2020

Abstract

Human stride analysis is a viable tool for identifying mobility changes, classifying abnormal gait, estimating fall risk, monitoring progression of rehabilitation programs, and indicating progression of nervous system related disorders. Marker based motion analysis systems are accurate for stride analysis but are not feasible for monitoring larger populations since external markers or sensors must be attached on the body. These external markers or sensors may be awkward for the person and may affect their natural movements. Marker-less depth sensors have shown potential to capture human movement without external markers and, with automation, could be applied to larger populations with less human resource requirements. However, existing marker-less depth sensor research has reported poor foot tracking accuracy. Hence, the goal of this thesis was to determine an appropriate depth sensor, define a multiple depth sensor system for foot tracking that could work in an institutional hallway (Smart Hallway), develop a viable foot tracking algorithm with depth sensors data, and extract accurate stride parameters.

Six non-interfering Intel RealSense D415 were set up in an institutional hallway scenario and captured the lower leg of a walking person. The developed foot tracking algorithm was applied to the combined point cloud data from these sensors and the obtained stride parameters were compared with a reference gold standard Vicon system.

Results showed that this high frame-rate marker-less system (approximately 60fps) was in good agreement with the Vicon system for temporal stride parameters, with mean errors less than 10ms. For spatial related stride parameters, errors were observed for step width and foot angle due to random noise generated around the foot in the depth sensors data. This research supported the depth camera approach for stride timing analysis. Further research is required to improve noise reduction for better step width and foot angle measurement.

Table of Contents

List of Figures

List of Tables

Abbreviations and Definitions

2D	Two Dimensional
3D	Three Dimensional
3DMA	Three Dimensional Motion Analysis
ACM	Association for Computing Machinery
AP	Anterior-Posterior
CMBES	The Canadian Medical and Biological Engineering Society
CPG	Central Pattern Generator
CPU	Central Processing Unit
FOV	Field of View
FS	Foot Strike
FO	Foot Off
FPS	Frames per Second
GPU	Graphics Processing Unit
HSV	Hue Saturation Value
IEEE	Institute of Electrical and Electronics Engineers
IMU	Inertial Measurement Units
IR	Infrared
KNN	K-Nearest Neighbors
LED	Light Emitting Diode
LFS	Left Foot Strike
LFO	Left Foot Off
MDC	Minimum Detectable Change
ML	Medial-Lateral
MLP	Multi-layered Perceptron
MS	Multiple Sclerosis
NSERC	Natural Sciences and Engineering Research Council of Canada
OBB	Oriented Bounding Box
PCS	Physical Contact-based Systems
PCLS	Physical Contact-less Systems

POE	Power Over Ethernet
RANSAC	Random Sample Consensus
RCS	Room Coordinate System
RF	Random Forest
RFS	Right Foot Strike
RFO	Right Foot Off
RGB	Red Green Blue
SD	Standard Deviation
SVM	Support Vector Machine
V	Vertical

Acknowledgments

My sincere gratitude to my supervisors, Dr. Pascal Fallavollita, Dr. Edward D. Lemaire, and Dr. Natalie Baddour, for believing in my skills and providing me with this excellent research opportunity. I thank them for supervising and guiding me in all the difficulties throughout the thesis. I will be forever thankful for their invaluable guidance, motivation, and support. This thesis facilitated me to advance my knowledge and skills, provided me opportunities to interact with many professionals, and helped me to grow as a researcher.

I will forever be indebted to my parents, Veeraiah Gutta and Vani Gutta, my brother, Varun Gutta, and my uncle, Apparao Rayala, for their unconditional love and support. Thanks for all the encouragement and teachings throughout my life.

This thesis is part of CREATE-BEST program funded by Natural Sciences and Engineering Research Council of Canada (NSERC), my sincere acknowledgments for financing this research.

I want to acknowledge the “The Ottawa Hospital Rehabilitation Centre” for providing the infrastructure to develop and perform preliminary testing in the Rehab Technology laboratory. My acknowledgments to the METRICS laboratory at the University of Ottawa for its computation facilities. I thank Dr. Julie Nantel and her staff for their support with the Vicon motion capturing laboratory during our data collection.

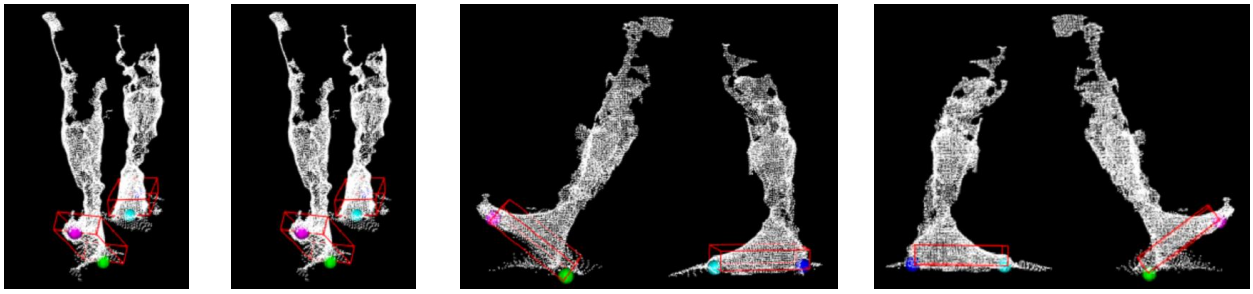
I would like to especially thank the volunteers who participated in the prototype testing and all the people that helped me throughout this thesis journey.

1 Introduction

Human motion capture is widely used in healthcare, robotics, entertainment, and education. In healthcare, gait analysis achieved with motion capturing enables quantitative patient evaluation and mobility risk documentation. Human stride analysis is a viable tool for identifying mobility changes; classifying abnormal gait; estimating fall risk; monitoring progression of Parkinson's disease [1], dementia [2], autism disorders [3], multiple-sclerosis (MS) disorders [4], and rehabilitation [5]. As technology advanced, motion capturing techniques evolved to track human joints accurately, and the latest trends move towards marker-less motion capture.

Marker-less systems can capture a person's natural gait efficiently, with no external sensors attached to their bodies. Depth sensing cameras with real-time human skeleton tracking opened a new dimension in marker-less approaches for healthcare applications. These low-cost depth sensor based methods can measure person's body position over time. However, limitations and gaps exist in marker-less motion capturing technologies to achieve an appropriate clinical level of accuracy, especially in foot tracking.

The goal of this thesis is to develop and evaluate a marker-less motion capture system with depth sensors that is optimized for use in institutional hallways, and to evaluate the system for foot tracking in human stride analysis (Figure 1.1). Patients, staff, and visitors constantly move through hospitals, long-term care facilities, and rehabilitation center hallways. This research focuses on hallways since these spaces could be smartly utilized for marker-less stride analysis to monitor patients or resident mobility progression without intrusion.



(a) **(b)** **(c)** **(d)**

Figure 1.1 Foot tracking results of an instance from different perspectives, (a) front, (b) back, (c) left, (d) right

1.1 Rationale

Marker-less motion capture for walking stride analysis typically use color cameras or depth sensors. Color cameras capture two-dimensional (2D) color images and depth sensors capture both 2D color images and three-dimensional (3D) information of a scene (i.e., point cloud).

Computer vision and artificial intelligence (AI) can be applied to 2D images to define key-points that represent the person's body. Images from different perspectives can be used to construct 3D information from 2D key-points. These systems are currently in the R&D phase and require substantial hardware and computational costs. AI based systems may need a specific image training set for optimal hallway measurements, which is challenging due to time and effort to label ground-truth key-points on the images. Also, poor lighting, shadows, or objects in the hallway could affect the results.

Depth sensors are a relatively low-cost and deliver 3D information in real-time. Infra-red based sensors can work in low lightening hallway conditions, are not affected by shadows, and are available commercially; therefore, this approach could be ideal for the hallway application. For example, the popular Kinect depth sensors can provide marker-less human motion capturing at 30 frames per second (fps) (captures data every 33.33ms). However, this sample rate may not be sufficient to appropriately identify stride parameter events for pathological gait [6] (i.e., foot off can occur between data samples), and significant errors were reported in the literature for some spatial stride parameters. Machine learning based skeleton tracking algorithms developed for Kinect depth sensors showed poor performance in foot tracking, with joint locations sometimes falling out of the body shape and results highly dependent on the viewing angle of sensors. Simultaneous data capture with multiple depth sensors would be required to avoid occlusions in the narrow hallway spaces, but these structured light and time-of-flight based depth sensors interfere with each other when used together.

The latest depth sensors with promising specifications of high resolution, speeds at or above 60fps, and long-range capabilities could improve foot tracking accuracy over other depth sensors. Whole foot surfaces obtained with high resolution (i.e., more 3D points on the surface) could enable accurate 3D foot identification. When combined with 60fps or faster data capture rates, gait events can be accurately identified and used to calculate clinically useful stride parameter outcomes.

1.2 Objectives

This thesis developed and validated a novel marker-less foot tracking system using 3D depth sensors. The objectives were:

1. Select an appropriate depth sensor for motion capturing in an institutional hallway.
2. Design the physical setup with multiple depth sensors in a hallway scenario and synchronize all the depth sensors.
3. Develop an algorithm to track human foot in the 3D point cloud data.
4. Validate the foot tracking algorithm by calculating and comparing the stride parameters with the gold standard Vicon system.

1.3 Thesis Contributions

This thesis resulted in many contributions related to depth sensor based marker-less motion capture. These contributions include:

- Characterization and comparison of current state-of-the-art depth sensors from time-of-flight, structured light and stereoscopic technologies.
- Development of a viable hallway-based setup with depth sensors for 3D point cloud data collection of human gait.
- Development of a foot tracking and stride parameters extraction algorithm from the walking human point cloud data.
- Quantitative evaluation of stride parameter performance compared to the gold standard Vicon system.

1.4 Thesis Outline

This thesis is divided into six chapters and one appendix. Chapter 2 is a literature review that discusses human motion capturing technologies, contact-based and non-contact-based technologies used for gait analysis, depth sensor evolution, and limitations in marker-less foot tracking techniques.

Chapter 3 compares three different depth-sensing technologies: Microsoft Kinect V2 sensor with time-of-flight technology, Orbbec Astra pro sensor with structured light, and stereoscopic based Intel RealSense D415 sensors. The contents of this chapter were published in [7].

Chapter 4 discusses the novel 3D animation-based simulations developed to quantify the visibility of human foot surface while walking through an institutional hallway. This research details were published in [8].

Chapter 5 presents a novel foot tracking algorithm, data collection protocol, data comparison analysis between marker-less system and a Vicon system.

Chapter 6 presents the thesis summary, thesis conclusions, and suggestions for future work.

Appendix shows the University of Ottawa Ethics approval certificate.

2 Literature Review

Gait analysis is the study of human movement and can be used to measure changes in mobility, progression of neurological disorders (Parkinson's, dementia, autism, MS diseases), ageing related issues, and rehabilitation. Monitoring gait over the time can provide early disease identification, enabling health staff to provide required treatments.

Gait analysis techniques can be classified as semi-subjective and objective [9]. In the semi-subjective method, the quality of a person's gait is observed by a specialist then sometimes followed by a patient survey to provide a subjective evaluation about their gait. Because of the subjectivity, this method affects long-term monitoring accuracy, and can negatively affect diagnosis. The most common semi-subjective techniques performed in clinical environments are:

1. **Timed 25-Foot Walk (T25-FW):** A specialist measures the time for 7.5 meters straight line walk [10].
2. **Multiple Sclerosis Walking Scale (MSWS-12):** 12 parameters describe the impact of MS on patient's gait [11].
3. **Tinetti Performance-Oriented Mobility Assessment (POMA):** Evaluates elderly person's balance and gait disorders by observing a series of maneuvers [12].
4. **Timed Get up and Go (TUG):** Balance measurement timed test in which the patient stands up from the chair, walks 3m forward, turns around, walks back, and sits down on the chair [13].
5. **Gait Abnormality Rating Scale (GARS):** Video-based analysis of 16 gait characteristics (5 general, 4 lower limbs, 7 related to head, upper limbs, and trunk) to study fall risk [14].
6. **Extra-Laboratory Gait Assessment Method (ELGAM):** Evaluates unstable gait by studying step length, speed, initial gait style, ability to turn the head while walking, and static balance [15].

In the objective analysis category, advanced sensors, devices, and technologies are used to measure human motion. Some objective techniques can provide more information on a patient's gait. Objective measurements can be more accurate and reliable than semi-subjective methods. Human motion capture systems for objective gait analysis can be broadly classified as physical contact-based systems (PCS) and physical contact-less systems (PCLS). In PCS, either sensors or

markers are physically attached to users or users must touch the sensors or markers for gait measurement. In PCLS, the user does not have to wear sensors, markers, and is not required to touch or be touched for data collection. With sensors or markers affixed to the body, PCS has the advantage of delivering highly accurate results. PCLS has the benefit of easier and more time-efficient movement measurement.

2.1 Stride Parameters

In quantitative gait analysis, stride parameters are outcome measures related to step timing and step distances. These parameters are useful for describing how a person walks and can be related to other issues such as fall risk [16], balance confidence [17], and mental capacity [18]. Stride parameters are related to gait cycle, the time interval or series of motions between two consecutive initial contacts of same foot [19]. Figure 2.1 shows different events, periods, tasks, and phases in a right leg's gait cycle. A gait cycle is divided into two phases. Stance phase is from the initial contact event (foot strike) of a leg to its immediate toe off event (foot off). From foot off event to the same leg's immediate foot strike event is known as swing phase. Foot strike and foot off events are crucial in calculating the stride parameters.

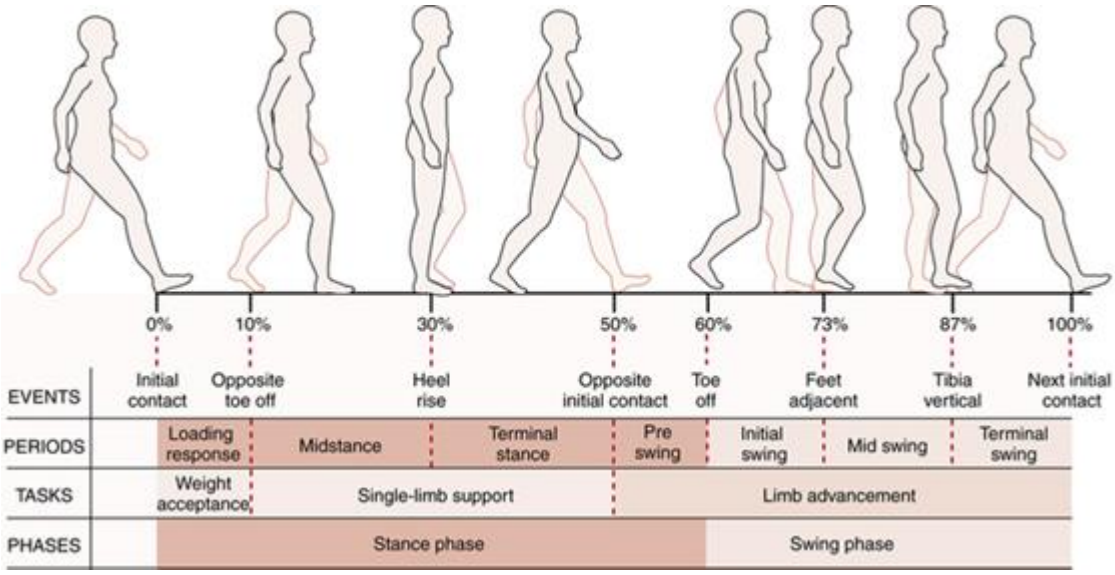


Figure 2.1 Overview of a gait cycle (modified from [19])

Based on the events in the gait cycle, many useful stride parameters can be calculated. Table 2-1 shows the major stride parameters used for clinical applications [9].

Table 2-1 Stride parameters used in clinical applications

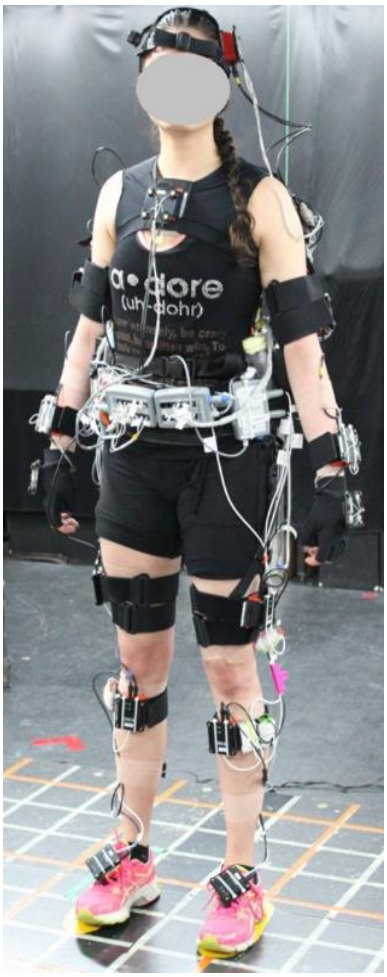
Parameter	Description
Stride length	Distance covered by the foot between two consecutive foot strike events (same leg) along the sagittal plane (line of progression)
Stride time	Time between two consecutive foot strike events (same leg)
Stride speed	Stride length divided by stride time
Step length	Sagittal distance between right and left heels at foot strike
Step width	Frontal distance between right and left heels at foot strike
Step time	Time between foot strike event to opposite leg foot strike
Cadence	Steps per minute
Stance time	Stance phase time
Swing time	Swing phase time
Foot angle	Angle of foot midline with sagittal plane when the foot is entirely in contact with the floor (mid-stance period).
Double support time	Time between foot strike and contralateral foot off

2.2 Physical Contact-based Systems (PCS)

Physical Contact-based Systems (PCS) include optical motion capture technology using cameras and markers [20], [21]; inertial motion capture technology using accelerometers, magnetometers and gyroscopes [22], [23]; magnetic motion capture technology using electromagnetic transmitter and receivers [24]; electromechanical motion capture technology using exoskeleton, gyroscopes, and potentiometers [25]; and acoustic motion capture technology using ultrasonic sensors [26] (Figure 2.2). For wearable based systems, the person wears active or passive markers or sensors on the body. Some sensors may cause discomfort or change the natural way a person moves [27]. Newer PCS technologies include textile-based transducers incorporated into socks to detect critical gait events during stance phase [28]; non-wearable GAITRite [29] that uses a pressure sensor carpet to measure stride parameters and plantar pressures; and shoe insoles embedded with an array of pressure sensors [30].

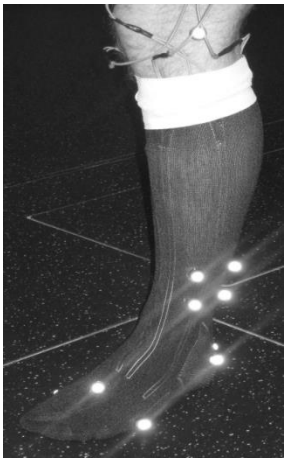
(a)

(b)



(c)

(d)



(e)

(f)

(g)

Figure 2.2 (a) Markers for Vicon system (optical motion capture) [20], (b) Electro-mechanical motion capture system [25], (c) Inertial motion capture technology [23], (d) Ultrasonic acoustic technology [26], (e) Instrumented sock [28], (f) GAITRite carpet [29], (g) Instrumented insoles [30].

Each of these PCS methods has advantages and disadvantages. For example, optical based systems such as Vicon and Optitrack delivers gold standard accurate outputs, but for large sample populations gait parameters measurement requires substantial resources such as data analysis time and technical assistants to prepare the person. User preparation effort is required for all wearable technologies. With the GAITRite's non-wearable technology, a limitation is that body kinematics cannot be measured (i.e., limb movement, foot clearance, etc.).

In the literature, Vicon is considered a gold standard for tracking human motion. In this optical technology, reflective passive markers are attached to the human body and these markers are tracked with multiple high-speed cameras from different angles, resulting in 3D coordinates. The Vicon motion capturing system needs an expensive, sophisticated hardware setup and dedicated lab space, which often remains unused when data collection sessions are not in progress. These factors introduce challenges for establishing a motion capture lab in healthcare clinics [31].

2.3 Physical Contact-less Systems (PCLS)

Human motion capture based on RGB cameras and depth cameras are popularly known as marker-less systems since they do not require markers or sensors affixed on the person's body. Hence, they are a more flexible approach for capturing natural human motion.

Depth cameras, unlike traditional cameras, directly capture 3D information in real-time. This reduces the quantity of required hardware and computation power for calculating 3D information of a captured scene. With these advantages over regular RGB cameras with AI or photogrammetry processing, depth sensors are further investigated in this thesis for performing stride analysis.

2.4 Depth Sensors

Time-of-flight, structured light, and stereo-vision are three viable depth sensing technologies [32]. This section describes the chronological evolution and product details about common depth sensors used in movement analysis research and other applications. Depth sensor technologies have been used in different domains; such as, entertainment, health, robotics, and education.

2.4.1 Microsoft Kinect Xbox 360 [33], [34]

- **Sensor Type:** Infra-red
- **Depth Technology:** Structured Light
- **Range:** 0.4m – 4.5m
- **Color Image:** 640 x 480 at 30fps
- **Depth Image:** 320 x 240 at 30fps
- **Depth FOV:** 57° x 43°
- **Dimensions:** 280mm x 64mm x 38mm
- **Connectivity:** USB 2.0
- **Environment:** Indoors
- **Release Year:** 2010

© Public Domain

2.4.2 Asus Xtion Pro Live [35]

- **Sensor Type:** Infra-red
- **Depth Technology:** Structured Light
- **Range:** 0.8m – 3.5m
- **Color Image:** 1280 x 1024
- **Depth Image:** 640 x 480 at 30fps
  : 320 x 240 at 60fps
- **Depth FOV:** 58° x 45°
- **Dimensions:** 457mm x 90mm x 127mm
- **Connectivity:** USB 2.0
- **Environment:** Indoors
- **Release Year:** 2011

© 2015 Plazathemes.com

2.4.3 Microsoft Kinect for Xbox One [34], [36]

- **Sensor Type:** Infra-red
- **Depth Technology:** Time-of-flight
- **Range:** 0.4m – 4.5m
- **Color Image:** 1920 x 1080 at 30fps
- **Depth Image:** 512 x 424 at 30fps
- **Depth FOV:** 70° x 60°
- **Dimensions:** 250mm x 66mm x 67mm
- **Connectivity:** USB 3.0
- **Environment:** Indoors
- **Release Year:** 2013

© Public Domain

2.4.4 Intel RealSense F200 [37], [38]

- **Sensor Type:** Infra-red
- **Depth Technology:** Coded Light
- **Range:** 0.2m – 1.2m
- **Color Image:** 1920 x 1080
- **Depth Image:** 640 x 480 at 30fps
- **Depth FOV:** 73° x 59°
- **Dimensions:** 110mm x 12.5mm x 3.75mm
- **Connectivity:** USB 3.0
- **Environment:** Indoors
- **Release Year:** 2014

© Intel Corporation

2.4.5 Intel RealSense R200 [39]

- **Sensor Type:** Infra-red
- **Depth Technology:** Active Stereo
- **Range:** 0.2m – 3.5m
- **Color Image:** 1920 x 1080 at 30fps
- **Depth Image:** 480 x 360 at 60fps
- **Depth FOV:** 59° x 46°
- **Dimensions:** 101.6mm x 9.6mm x 3.8mm
- **Connectivity:** USB 3.0
- **Environment:** Indoors/Outdoors
- **Release Year:** 2015

© Intel Corporation

2.4.6 ZED Stereo Camera [40]

- **Sensor Type:** RGB
- **Depth Technology:** Stereo
- **Range:** 1.5m – 20m
- **Color and Depth Image:** 4416 x 1241 at 15fps
  : 3840 x 1080 at 30fps
  : 2560 x 720 at 60fps
  : 1344 x 376 at 120fps
- **Depth FOV:** 90° x 60°
- **Dimensions:** 175mm x 30mm x 33mm
- **Connectivity:** USB 3.0
- **Environment:** Indoors/Outdoors
- **Release Year:** 2015

© 2020 Stereolabs Inc.

2.4.7 Orbbec Astra Pro [41]

- **Sensor Type:** Infra-red
- **Depth Technology:** Structured Light
- **Range:** 0.6m – 8m
- **Color Image:** 1280 x 720 at 30fps
- **Depth Image:** 640 x 480 at 30fps
- **Depth FOV:** 60° x 49.5°
- **Dimensions:** 165mm x 30mm x 40mm
- **Connectivity:** USB 2.0
- **Environment:** Indoors
- **Release Year:** 2015

© 2019 Orbbec 3D

2.4.8 Orbbec Presee [42]

- **Sensor Type:** Infra-red
- **Depth Technology:** Structured Light
- **Range:** 0.6m – 8m
- **Color Image:** 1280 x 720 at 30fps
- **Depth Image:** 640 x 480 at 30fps
- **Depth FOV:** 60° x 49°
- **Dimensions:** 172mm x 63mm x 56mm
- **Connectivity:** USB 2.0
- **Environment:** Indoors
- **Release Year:** 2016

© 2019 Orbbec 3D

2.4.9 Intel RealSense SR300 [43]

- **Sensor Type:** Infra-red
- **Depth Technology:** Coded Light
- **Range:** 0.3m – 2m
- **Color Image:** 1920 x 1080 at 30fps
  : 1280 x 720 at 60fps
  : many other configurations
- **Depth Image:** 640 x 480 at 60fps
- **Depth FOV:** 73° x 59°
- **Dimensions:** 110mm x 12.5mm x 3.75mm
- **Connectivity:** USB 3.0
- **Environment:** Indoors
- **Release Year:** 2016

© Intel Corporation

2.4.10 Intel RealSense D415 [44]

- **Sensor Type:** Infra-red
- **Depth Technology:** Active Stereo
- **Range:** 0.3m – 10m
- **Color Image:** 1280 x 720 at 30fps
  : 848 x 480 at 60fps
  : many other configurations
- **Depth Image:** 1280 x 720 at 30fps
  : 848 x 480 at 60fps
  : many others up to 90fps
- **Depth FOV:** 65° x 40°
- **Dimensions:** 99mm x 20mm x 23mm
- **Connectivity:** USB 3.0
- **Environment:** Indoors/Outdoors
- **Release Year:** 2018

© Intel Corporation

In addition to the above-mentioned depth sensors, depth sensors are also available from Duo3d [45], e-con systems (Tara depth sensor) [46], Occipital [47], and newer depth sensors from Microsoft [48], Intel [49], Orbbec [50]. These depth sensors can have smaller physical dimensions and power, data transfer over USB, high resolution, and higher frame rates. Any one sensor is not suitable for all applications, thus no universal solution exists at this time.

2.5 Stride Analysis using Depth Sensors

Most of the research on depth sensors for human movement analysis involved Microsoft Kinect, because of the potential to track user joints with ease. However, little research on stride analysis using other depth sensors have been published, other than Asus Xtion Pro Live [51]. Therefore, the following review of depth camera systems that track lower body motion (i.e., knees, ankles, foot, and stride) is predominately on Kinect sensors.

Other depth sensors have been investigated for measuring geometric foot dimensions and foot arch curves [52], gesture recognition [53], home physiotherapy rehabilitation assistance [54], emotion detection systems [55], virtual-reality based physical rehabilitation games [56], upper body rehabilitation systems [55], and visually impaired people assistance systems [57], [58].

2.5.1 Abnormal Gait Classification

Kinect Xbox 360 (Kinect V1) skeleton tracking data has been validated to detect abnormal gait patterns, which subsequently helped identify fall detection, nervous system related autism

disorders, multiple sclerosis, and Parkinson's disease. The Kinect V1 machine learning algorithm tracked 18 skeleton joints per person (maximum of two people tracked simultaneously).

In one study, five walking styles (normal, normal with smaller steps, waddling, left foot pain, right foot pain) were used to identify abnormal gait patterns [59]. Hip progression line, footsteps, and knee rotations were also considered to identify abnormalities using a Support Vector Machine (SVM) classifier. 88 percent accuracy was achieved in gait classification. People with Parkinson's disease have also been classified with greater than 90 percent accuracy [60], based on normalized stride length (with respect to leg length). Tip-toe walking classification from normal walking was developed to identify autism disorders [61]. In this research, hip, knee, and toe angles were outputs of a Central Pattern Generator (CPG), and its CPG's parameters were estimated (best achieved with two oscillators). These CPG parameters were provided to a classifier to identify tip-toe walking. K-Nearest Neighbors (KNN), SVM classifier, Multi-layered perceptron (MLP), and Random Forest (RF) classifiers were implemented and compared. MLP outperformed the other classifiers in both without joint coupling (joints were independent, accuracy = 75.07%) and with joint coupling (joints were dependent, accuracy = 79.7 %).

An attempt to classify brain concussions based on abnormal gait patterns was made with athletes having a potential risk of concussion [62]. This study did not yield acceptable results because of a lack of non-normal gait data. Real-time fall detection of people with multiple sclerosis at their home was investigated but failed due to a limited monitoring area with a single sensor [63].

Spatiotemporal (stride time, length, velocity) and kinematic variables (hip and knee flexion) of 257 healthy individuals with wide age range of 10-80 years were studied to identify age related gait differences with Kinect V2 skeleton data. Kinect V2 can track 25 skeleton joints per person (maximum of six people tracked simultaneously). Results had a similar trend as laboratory grade systems but limited capture range affected the user's ability to reach their comfortable normal walking speed [64] and depth occlusion while walking caused errors in finding foot locations and step length [65]. These problems could be addressed by using multiple sensors.

Algorithms based on Kinect skeleton tracking data delivered reasonable results in abnormal gait classification, and the limitations in capturing a large volume area were solved by using multiple sensors simultaneously. In a study with four Kinect V1 sensors, accuracy for properly identifying two peoples as they transitioned between pairs of cameras was 90.61% [66].

2.5.2 Stride Parameters

Methods for extracting stride parameters from Kinect data, with comparison to ground truth systems, are summarized in Table 2-2. Kinect V1 skeleton data had poor foot landmark detection during toe-off. Additionally, skeleton joints were only detected when a person faced the sensor. Skeleton tracking algorithm sampling rate irregularities and inability to provide foot key points (e.g., metatarsophalangeal and calcaneus) contributed to errors in for gait event detection [67]. Detection of valid motion periods and postprocessing limb orientation data using 5th order polynomial fitting (Savitzky-Golay) improved tracking accuracy [68].

Only two studies used the Kinect 3D point cloud [69] [70]. Both studies calculated stride parameters with the camera in front of the person and then perpendicular to the walking path. In [69], the perpendicular camera placement achieved better results than having the camera in front. Contrarily, in [70] the front position achieved better results. In both studies, point cloud 3D points were more accurate when the person was closer to the sensor, irrespective of walking direction. Since mean errors were greater than gait variability minimum detectable change (MDC) [71], system accuracy must improve further before being acceptable in clinical practice.

The Kinect V2 sensor has higher depth resolution, wider field of view (FOV), and improved skeleton tracking (tracks 25 joints), but the capture frequency remains at 30fps. Temporal gait parameter accuracy is highly dependent on the frame rate. For example, for a person walking at 0.67 m/s, considered a slow walking speed [72], the time between pre-swing phase and the initial swing phase would be 0.1s, in which foot-off occurs (0.03 s duration). Therefore, a 30fps frame rate (0.033s per frame) would make accurate detection of foot-off theoretically not possible [6].

Multiple Kinect V2 sensors have been used to capture longer ranges and overcome viewing angle sensitivity issues. In a 10-meter walking test study, four Kinect V2 sensors were placed on the same side of a walking pathway and were angled to capture the person’s front. Results were in good agreement with an Optotrak system [73]. The ankle farther from the sensors showed more inconsistency in the mediolateral direction.

Table 2-2 Stride parameters using Kinect sensors

Ref	Ground truth	Methodology	Stride parameters	Error results
[67]	Marker based 3DMA system	Kinect V1 skeleton data	Step length, Step time, Speed, Stride length, Stride time, Foot swing velocity	**Mean error** Step length: 11.5mm Step time: -0.17s Speed: -0.01m/s Stride length: -4.2mm Stride time: -0.20s Foot swing velocity: 0.43m/s
[69]	GAITRite	Two experiments with a single Kinect V1 1. Perpendicular to walking pathway 2. In front of walking pathway Gait parameter extraction based on vertical center of mass displacement Normal walking, small steps, walking in long skirt cases in the perpendicular view Normal walking alone in the front view	Step length Step width Speed	**Error percentage (Mean ± SD)** **Normal walk** Step length: 4.6 ± 3.8 Step time: 6.6 ± 4.9 Speed: 3.69 ± 1.81 **Small steps** Step length: 4.7 ± 4.2 Step time: 6.1 ± 4.5 Speed: 3.28 ± 2.28 **Long skirt** Step length: 4.7 ± 3.3 Step time: 4.7 ± 5.5 Speed: 4.37 ± 3.19 **Kinect in the front view** Step length: 12.5 ± 12.6 Step time: 11.5 ± 8.3 Speed: 10.04 ± 7.33
[70]	Vicon	Two independent Kinect V1 sensors, sagittal and frontal plane Projections of normalized point cloud onto floor (points in bottom 20 inches)	Stride length Stride time Stride velocity	**Sagittal Kinect (mean ± SD)** Length (cm): -0.7±2.44 Time (ms): 9.39±190.54 Velocity (cm/s): -0.95±1.54 **Front Kinect (mean ± SD)** Length (cm): -0.32±1.0 Time (ms): 7.47±62.2 Velocity (cm/s): -0.42±0.64
[73]	Optotrak	Four Kinect V2 sensors skeleton data	Walking speed (cm/s) Cadence(steps/min) Step length (cm) Stride length (cm) Step width (cm) Step time (s) Stride time(s)	**Bias (95% LoA)** Walking speed: 1.1 (0.1, 2.1) Cadence: -0.9 (-3.0, 1.2) Step length: -0.1 (-1.4, 1.2) Stride length: 0.1 (-0.7, 0.9) Step width: -1.3 (-5.2, 2.2) Step time: 0.01 (-0.02, 0.03) Stride time: 0.01 (-0.02, 0.04)
[74]	Stride length set to 0.62m	Kinect V1skeleton data	Step width Stride length Speed	**Stride length error (m) (mean ± SD)** Left leg: 0.43 ± 0.7 Right leg: 0.54 ± 0.74

[75]	GAITRite	Tracking based on particle filtering formulated as Hidden Markov Model. Single Kinect V1 closer to the foot, region of interest 1.5m	Step length Step width	**Mean and SD ranges** Step length (cm): Mean error (2.73, 3.87) SD (2.02, 2.31) Step width (cm): Mean error (1.75, 3.04) SD (1.17, 2.52)
[76]	GAITRite	Single Kinect V1 perpendicular to walking pathway Center-of-mass based step frame detection	Step length Step width	**Error (mean ± SD)** Step length (cm): 8.25 ± 5.13 Step time (s): 0.29 ± 0.17
[77]	Tracking attached colored markers on key locations	Kinect V1 skeleton data Color segmentation with two camrecorders	Asymmetry index of elbow, knee, ankle	Cam recorders system outperformed Kinect skeleton tracking
[78]	Stereo photogrammetry (6 cameras)	3D colored path attached to foot and tracked based on color segmentation Treadmill walking	Foot position and orientation	**Foot position average RMSD error range (mm)** Medio-lateral: (4.9, 12.1) Vertical: (4.9, 12.1) Anterior-posterior:(19.4,26.5) **Foot orientation average RMSD error percentage** Medio-lateral: (5.6, 8.8) Vertical: (5.6, 8.8) Anterior-posterior:(15.5,18.6)
[79]	Shoe insole pressure sensors	Kinect V1 skeleton data from front and 45° side views	Stride time Stance time Swing time	**Front view (Mean ± SD)(ms)** Left stride time: 8 ± 62 Right stride time: 2 ± 46 Left stance time: -8 ± 110 Right stance time: -20±90 Left swing time: 6 ± 115 Right swing time: 27±104 **45° view (Mean ± SD) (ms)** Left stride time: 18.1±112 Right stride time: 27.3±52 Left stance time: 5.4±123 Right stance time 8±128 Left swing time: -8.3±106 Right swing time: 11.7±112
[80]	Vicon	Kinect V2	Ankle angle Knee angle Hip angle	**RMSE (degrees)** Ankle angle: 22.03 Knee angle: 5.33 Hip angle: 5.29
[81]	Vicon	Six Kinect V2 sensors	Walking speed (cm/s) Step time (s) Step length (cm) Step width (cm) Stride length (cm)	**Bland-Altman(Bias, 1.96SD)** Walking speed (0.53, 0.8) Step time (0.00, 0,00) Step length (0.36, 1.12) Step width (0.66, 1.49) Stride length (0.04, 1.53)

Another research with Vicon as ground truth showed significantly greater ankle angle errors (22.03°) with Kinect V2 than hip (5.29°) and knee (5.33°) [80]. Skeleton tracking algorithms were sensitive to the viewing angle, so step width accuracy (mediolateral errors) improved by using Kinect sensors on two sides of the walking pathway and then averaging the tracked values [81]. Nevertheless, Kinect V2 needs improvements to be clinically viable.

2.6 Conclusion

The popular Kinect based systems can identify and track most gait events with reasonable accuracy, but gaps and limitations exist, mainly with ankle and foot tracking. Skeleton tracking algorithms provided with the Kinect sensor are not reliable at the foot and are dependent on the viewing angle (to avoid depth occlusions), which requires multiple sensor setups for tracking. By considering these limitations, instead of just depending on skeleton joints tracking data, utilizing the foot's whole 3D point cloud has higher potential to achieve accurate tracking. The evolution of the latest depth sensors has overcome major hardware limitations such as depth resolution, capture speed, and capture range. These advanced depth sensors have the capability to provide better results than current Kinect systems.

3 Depth Sensor Selection

3.1 Overview

This chapter presents a comparison of three major depth sensing technologies: time-of-flight, structured light, and stereoscopic. In this chapter, experiments were performed with Microsoft Kinect V2 (time-of-flight), Orbbec Astra Pro (structured light), and Intel RealSense D415 (stereoscopic) sensors to compare the quality of depth images and 3D point clouds. In addressing Objective1, the outcomes of this chapter showed that the stereoscopic Intel RealSense D415 depth sensor is suitable for this application of motion capturing in a hallway scenario since multiple sensors are required to avoid occlusions while walking.

The contents of this chapter were published in the 42th conference of "The Canadian Medical and Biological Engineering Society" held at Ottawa, Ontario, Canada in May 2019.

V. Gutta, E. D. Lemaire, N. Baddour, and P. Fallavollita, "A Comparison of Depth Sensors for 3D Object Surface Reconstruction," in CMBES, 2019.

3.2 Abstract

The ability of depth cameras like Kinect to capture a scene's depth information in three-dimensions, along with 2D color RGB images, in real-time makes marker-less human motion capture a potential option for applications such as rehabilitation, robotics, education, etc. Various depth sensor technologies are commercially available, and selecting the appropriate depth sensor is highly dependent on the desired application. This research chapter compared Microsoft Kinect V2, Orbbec Astra Pro, and Intel RealSense D415 depth sensing technologies for object surface reconstruction within an interior daily living environment. Intel RealSense D415 was resistant to interference with multiple sensors and point cloud data at 1m range was more accurate than Microsoft Kinect V2 and Orbbec Astra Pro.

3.3 Introduction

Human motion capture with optical systems can be classified into two categories, marker-based and marker-less. For marker-based systems, a person wears either active markers that emit

light or passive markers that reflect light, and 3D marker positions are tracked in real-time. Examples of marker-based optical tracking systems are Vicon, Qualisys, and Codamotion.

For marker-less motion capture no markers are placed on the human body. Images are captured while a person moves in the system's field of view. Motion capture is based on processing these captured images.

Marker-less approaches typically use multiple RGB cameras or depth cameras. Finding 3D human body information from multiple RGB cameras requires extensive image processing and computation. Depth sensors based on infra-red light are independent of ambient lighting conditions in the scene and can also provide 3D data in real-time. These depth sensors may be convenient and easier to use than RGB cameras. Kinect for Xbox One (Kinect V2) has been used in the literature for marker-less gait analysis, but has not progressed to use in practice.

Human stride duration depends on walking speed [82]. Based on analysis from [72], 0.5 m/s is considered slow walking and speeds greater than 1.6 m/s are fast walking. Gait phases can have a minimum duration of 0.12s (fast walking at 1.75m/s, loading response and "pre-swing") [83]. Depth sensors such as the Kinect V2 operate at 30fps, which implies that a frame is captured every 33.33ms. Depth sensors working at 30fps can identify all gait sub-phases, even during fast walking. However, proper identification of foot-off, which happens in 30ms for a 0.671m/s walking speed, is theoretically not possible [6]. Even at low walking speeds, the Kinect V2 cannot consistently track foot and toe-off moments [6], [80]. Furthermore, to avoid occlusion while walking and increase the accuracy of tracking, multiple sensors are required [84].

The purpose of research in this chapter was to find a suitable depth sensor to overcome the limitations of Kinect V2 sensor for human foot 3D construction during fast walking. As a part of this study, three depth sensors were investigated: Kinect V2 (time-of-flight), Orbbec Astra Pro (structured light), Intel RealSense D415 (stereoscopic). The research outcomes provide a basis for developing new marker-less human movement analysis approaches.

3.4 Kinect V2, Astra Pro, RealSense D415

Table 3-1 shows typical specifications among Kinect V2, Astra Pro, and RealSense D415 sensors. Most D415 specifications are superior to those of Kinect V2 and Astra Pro. RealSense D415 can be configured to run at various resolutions and speeds, and has a high color resolution

of 1920 × 1080 and depth resolution of 1280 × 720. In this research, the D415 sensors were configured with 848 × 480 resolution, to achieve 60fps capture speed at the highest resolution.

Table 3-1 Depth sensors specification comparison

Features	Kinect V2	Astra Pro	RealSense D415
Technology	Time-of-flight	Structured light	Stereoscopic
Dimensions	Length: 250mm Width: 66mm Height: 67mm	Length: 165mm Width: 30mm Height: 40mm	Length: 99mm Width: 20mm Height: 23mm
Color Resolution	1920 x 1080 at 30fps	640 x 480 at 30fps	848 x 480 at 60fps
Depth Resolution	512 x 424 at 30fps	640 x 480 at 30fps	848 x 480 at 60fps
Maximum depth speed	30fps	30fps	Up to 90fps
Depth range	0.5m to 5m	0.6m to 8m	0.3m to 10m
Field of view	70.6° x 60°	60° x 49.5°	65° x 40°
Multiple sensors to a single computer	No	Yes	Yes
External power supply	Yes	No	No

3.4.1 Microsoft Kinect V2

Kinect V2 has a wider field of view and uses time-of-flight technology. In brief, Infrared (IR) rays are projected onto an object and these rays reflect to a camera array after hitting an object. The object's depth is calculated based on IR time-of-flight [85].

3.4.2 Orbbec Astra Pro

Astra Pro uses structured light technology to determine depth information. Known features are projected onto an object and observed using an IR camera. 3D construction is based on image correlation and triangulation [86].

3.4.3 Intel RealSense D415

Active IR stereoscopic technology is used to find depth data. An IR projector projects a texture pattern onto the scene, to find more matching features between stereo IR images. Depth is estimated based on the disparity between key-points in stereo images [87].

3.5 Methodology

Experiments were performed with Kinect V2, Astra Pro, and RealSense D415 to compare depth image quality and 3D point cloud quality. Point clouds were generated directly from the depth cameras.

3.5.1 Single Sensor Depth Images

A standard-size basketball of radius 119.3mm and two boxes with dimensions of 342mm x 162mm x 115mm were placed on the ground plane. Depth sensors were positioned approximately 1m from these objects (Figure 3.1). Depth images of these objects were captured with a single sensor and 100 frames were averaged. Sphere models to fit point cloud data are readily available in PCL libraries [88], so this chapter focused on sphere analysis.

Figure 3.1 Single sensor setup

3.5.2 Multiple Sensor Depth Images

The experimental setup remained the same as for the single sensor assessment, but with one additional depth sensor, opposite the first sensor, placed 1m from the objects (Figure 3.2). Depth images were captured with both sensors at the same time and averaged over 100 frames.

3.5.3 Sphere Detection from Point Cloud

Background removal was necessary to segment the ball from the scene. The minimum background method was applied using 500 depth frames captured without the ball and box objects in the scene [89]. Then, the basketball was placed in front of sensor on the ground plane.

Hundred depth frames were captured and background subtraction was applied to the depth data. A point cloud was created from each frame, containing only points belonging to the basketball. These points were fitted to a RANSAC sphere model [88].

Variables included the actual basketball radius (R), estimated radius (R_{est}), number of points in point cloud (N), number of points inliers to RANSAC sphere (N_m), estimated center (C_{est}), number of points farther than R from C_{est} (N_+), number of points nearer than R from C_{est} (N_-), distance of i^{th} model inlier point from C_{est} (d_m^i), distance of i^{th} farther point from C_{est} (d_+^i), and distance of i^{th} nearer point from C_{est} (d_-^i). σ_+ is standard deviation towards the exterior of a sphere, σ_- is standard deviation towards the interior of sphere, and R_{rms} is the root mean square error. The standard deviations and root mean square error are calculated using (3.1), (3.2), and (3.3). Results provided in Table 3-2 were the average of 100 iterations.

Figure 3.2 Two sensors setup

$$\sigma_+ = \sqrt[2]{\frac{\sum_{n=1}^{N_+}(R - d_+^n)^2}{N_+}} \quad \textbf{(3.1)}$$

$$\sigma_- = \sqrt[2]{\frac{\sum_{n=1}^{N_-}(R - d_-^n)^2}{N_-}} \quad \textbf{(3.2)}$$

$$R_{rms} = \sqrt[2]{\frac{\sum_{n=1}^{N_m}(R - d_m^n)^2}{N_m}} \quad \textbf{(3.3)}$$

3.6 Results

3.6.1 Depth Images

Depth pixel values greater than 2000 mm were excluded. Figure 3.3 – Figure 3.7 show depth camera images, where each pixel is a depth, scaled from 0-2000 mm to 0-255 for display. Black pixels indicate either no depth data available or depth data range is greater than 2000 mm.

The depth images when using a single Kinect V2 (Figure 3.3) had clean, sharp edges for both the rectangular shaped object and spherical object. Astra pro's depth images had many depth data gaps at object edges. RealSense D415's depth images had less missing depth data than Astra pro.

For two sensors capturing data simultaneously, Astra pro had extensive interference, resulting in missing depth data (i.e., dark pixels in Figure 3.4). Kinect V2 had less interference with two sensors but the interference region was not consistent (Figure 3.5). When averaged over 100 frames, the black pixel areas were reduced for Kinect (Figure 3.6). Averaging would only be applicable for static objects. The stereoscopic-based Intel RealSense D415 sensors did not have interference, even when both sensors capture depth data simultaneously (Figure 3.7).

3.6.2 Point Cloud

Intel RealSense D415's delivered a denser point cloud because of its high depth resolution. Error in estimated sphere radius was less than 1mm for both Astra Pro and RealSense D415 depth sensors. Intel RealSense D415 point cloud quality outperformed the other depth sensors in every aspect, other than the standard deviation of points towards the interior of sphere (Table 3-2).

(a) **(b)** **(c)**

Figure 3.3 (a) Single Microsoft Kinect V2 depth image, (b) Single Orbbec Astra pro depth image, (c) Single Intel Realsense D415 depth image

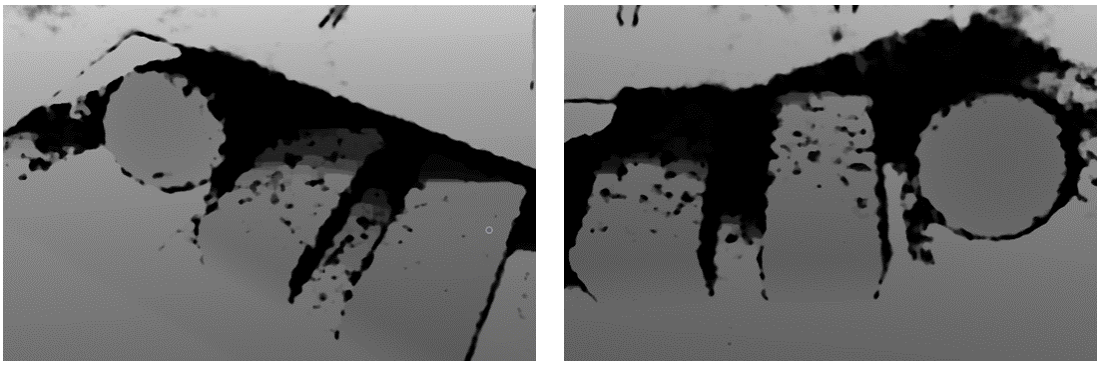


Figure 3.4 Averaged depth images from two Astra Pro sensors

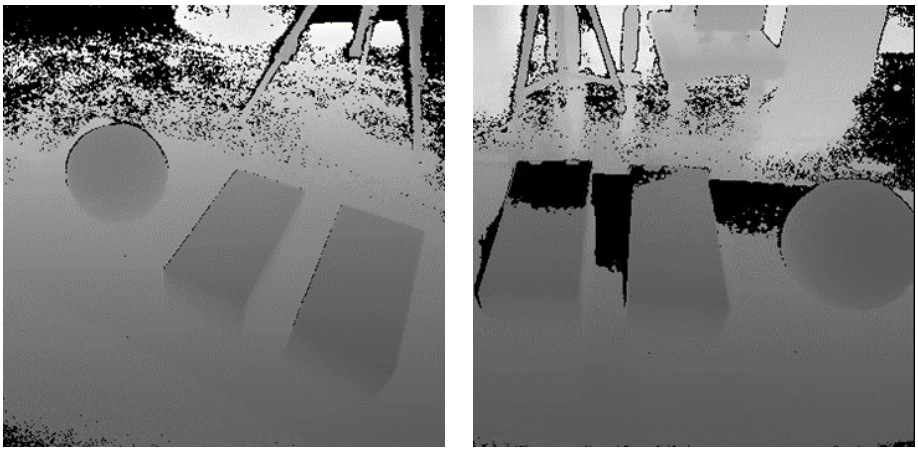


Figure 3.5 Depth images from two Kinect V2 sensors

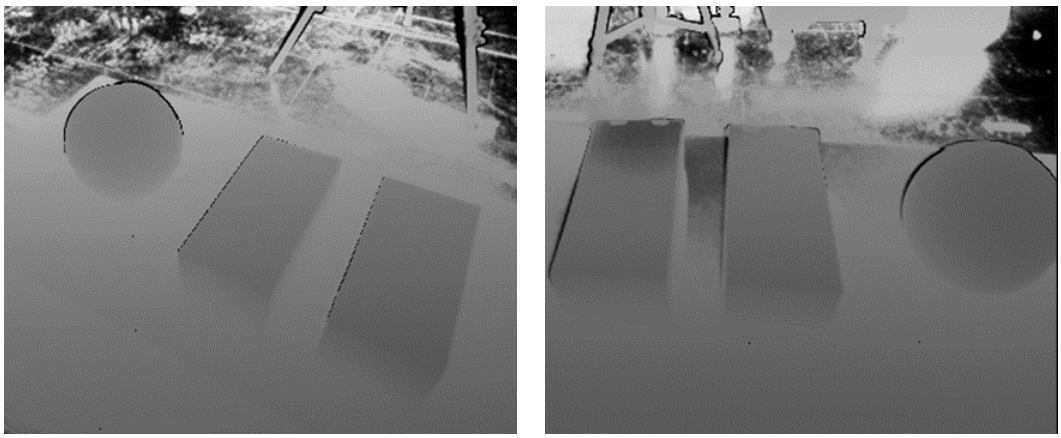


Figure 3.6 Averaged depth images from two Kinect V2 sensors

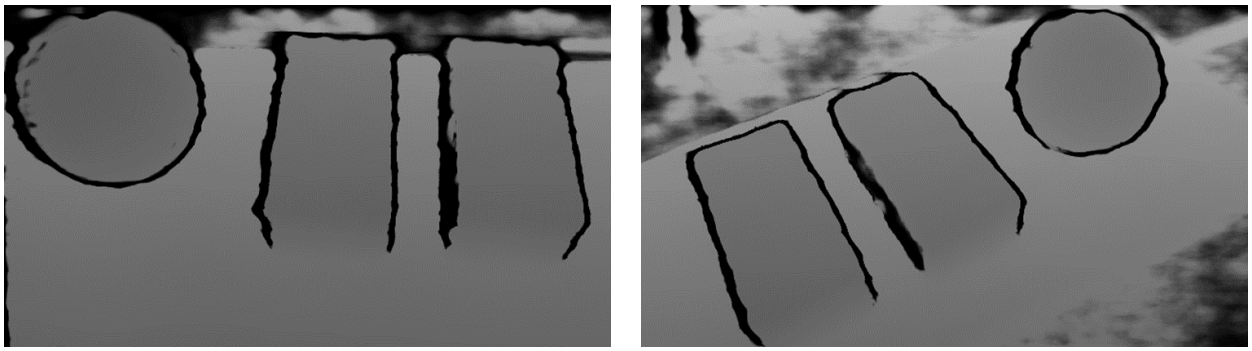


Figure 3.7 Averaged depth images from two RealSense D415 sensors

Table 3-2 Sphere detection

	Kinect V2	**Astra Pro**	**RealSense D415**
N	8132.33	18589.8	26708
N_m	7654.43	18482.6	26708
R_{est} (mm)	113.83	120.25	118.74
R_{rms} (mm)	7.15	1.88	1.68
σ_+ (mm)	5.79	1.86	1.37
σ_- (mm)	7.27	0.93	1.78

3.7 Conclusion

To construct 3D objects, data from multiple sensors are simultaneously required in order to avoid occlusion. The Kinect V2 and the Astra Pro sensors had interference when multiple sensors were used at the same time. The Intel RealSense D415 sensor did not display evidence of interference when multiple sensors were used simultaneously, and delivered a dense point cloud at 60fps. Therefore, the RealSense D415 is a feasible technology to construct 3D objects and also delivered less point cloud errors than the Astra Pro and Kinect Xbox One sensors at a range of 1m. The RealSense D415 dimensions, no external power supply requirement, and ability to use multiple sensors per computer are other advantages for a marker-less human movement analysis application.

4 Hallway Setup and Synchronization

4.1 Overview

This chapter discusses the physical setup of multiple depth sensors in an institutional hallway scenario, which was based on the 3D animated simulations. The details of temporal and spatial synchronization of these multiple sensors are presented. This chapter addresses Objective2 of this research.

The contents of this chapter were published in the "IEEE/ACM 1st International Workshop on Software Engineering for Healthcare" held at Montreal, Quebec, Canada in May 2019. Paper details are:

V. Gutta, N. Baddour, P. Fallavollita, and E. D. Lemaire, "Multiple Depth Sensor Setup and Synchronization for Marker-less 3D Human Foot Tracking in a Hallway," in IEEE International Workshop on Software Engineering for Healthcare, 2019, pp. 77–80 (© 2019, IEEE).

4.2 Abstract

In this chapter, an animation was designed to determine an appropriate physical setup for depth sensors in an institutional hallway, for human foot surface capture and tracking. The sensor configuration was tested by creating the setup in a laboratory with Intel RealSense D415 depth cameras, synchronizing multiple depth sensors (spatially and temporal), collecting depth camera data, and generating a point cloud of a person walking in the capture area. Simulations showed that 96% of a walking human's foot surface can be reconstructed from depth sensors data, using sets of four diagonal corner locations for each instance.

4.3 Introduction

Motion capture systems for human stride analysis can be broadly classified into marker-based and marker-less systems. In marker-based systems, a person wears markers or sensors on their body for tracking. Inertial marker-based systems use accelerometers and gyroscopes to calculate limb speed and orientation. These external sensors could affect the person's gait [27]. Marker-less systems track human motion without any external devices or markers attached to the

body. These systems can capture the person's natural gait and could enable movement analysis for larger populations. Depth sensors are well-known for real-time marker-less motion capture. The Kinect depth sensor was able to track most of human joints within acceptable clinical accuracy, but was poor at foot tracking [80].

Patients, staff, and visitors move through hospital, rehabilitation center, and long-term care hallways [90]. This space can be smartly utilized for stride analysis by implementing marker-less tracking. A single depth sensor is not sufficient to capture the entire human surface, especially for narrow areas like hallways. Foot tracking enables stride parameter analysis, which is a viable tool for identifying changes in mobility [4], fall risk [16], and mental capacity [18].

The Intel RealSense D415 depth camera can be considered for the foot tracking application. This camera uses active infrared stereo technology and was chosen because of its high resolution (848 × 480), high speed (60fps), range (10m) and no interference while multiple cameras are used simultaneously [44].

This research used software simulations to determine the required number of depth sensors and their set up in a hallway for foot tracking. The depth sensor setup was then tested in a lab environment.

4.4 Simulation

3D simulation of human walking in a hallway was developed using Blender software [91]. All objects in the simulation were designed with realistic dimensions. A light source with a field-of-view (FOV) similar to Intel RealSense D415 [44] was used as the depth sensor (Table 4-1). In simulations, we used the Intel D415 RGB camera's FOV (69.4°×42.5°). Based on the insights from the simulation results, the physical setup was built with the sensor's depth FOV (65°×40°).

In this simulation, three different setups were investigated (Figure 4.1). Setup1 had two depth sensors at opposite corners of hallway and Setup2 had four sensors at each corner. Light sources had a 7.5m center range, such that maximum diagonal range was within 10m.

Setup3 used a smaller capture area and six lower range sensors (5m). Since RealSense D415 depth error is proportional to the square of distance [92], improved accuracy requires a lower range. Reducing the range requires more sensors to provide four camera views at each instance in the capture volume.

Table 4-1 Animation objects details

Object	Dimensions
Hallway	Length: 10m Width: 1.8m Height: 2.7m
Human model (Figure 4.2)	Height: 1.78m
Light Source (2 and 4 sensors setups)	FOV: 69.4° × 42.5° × 76.9° Range: 7.5m
Light Source (6 sensors setup)	FOV: 69.4° × 42.5° × 76.9° Range: 5m

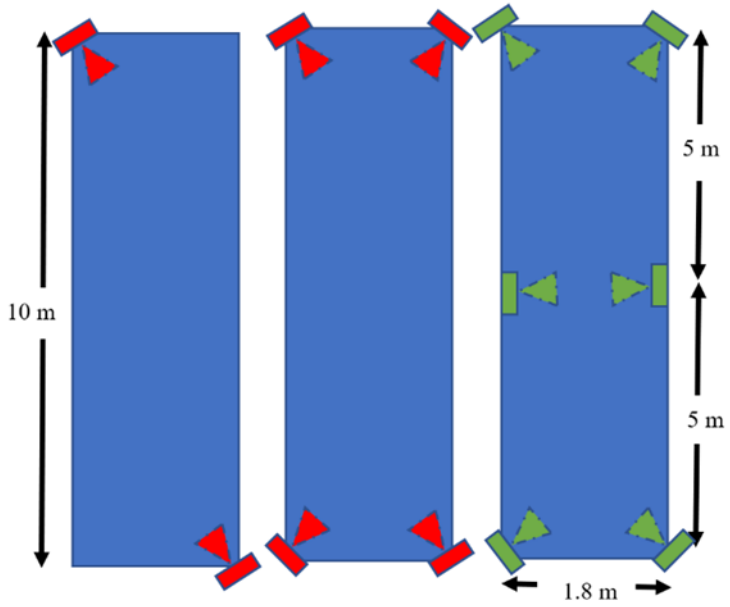
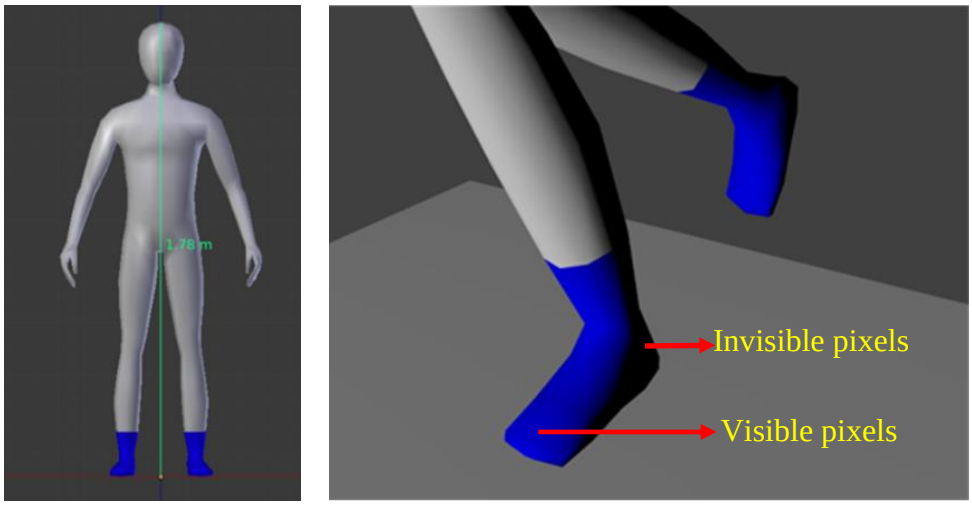


Figure 4.1 Hallway simulation setups with two, four, and six sensors

4.4.1 Experiment

The human model (Figure 4.2) was animated to walk through each setup. The target foot area was textured in blue and the light source used as a depth sensor. The region where light contacted the human was brighter; therefore, bright target pixels were considered visible (i.e., depth data would be captured). At each instance, the target area was captured from 12 perspectives using virtual cameras (Figure 4.3).



(a) **(b)**

Figure 4.2 (a) A 3D Human model, (b) Textured target area and visible bright areas

Every target pixel was captured in at least one of these 12 frames. For the i^{th} camera frame at an instance, target pixels (T_i), visible target pixels (V_i), and invisible target pixels (I_i) were used to calculate visibility percentage (P) at that instance using equations (4.1) and (4.2).

All target pixels were converted into hue saturation value (HSV) color spaces. Using the "Value" channel, also known as "Intensity" or "Brightness", target pixels with "Value" channel of at least 30% of maximum were categorized as visible. Pixels with lower brightness may result in poor capture quality and were categorized as invisible pixels.

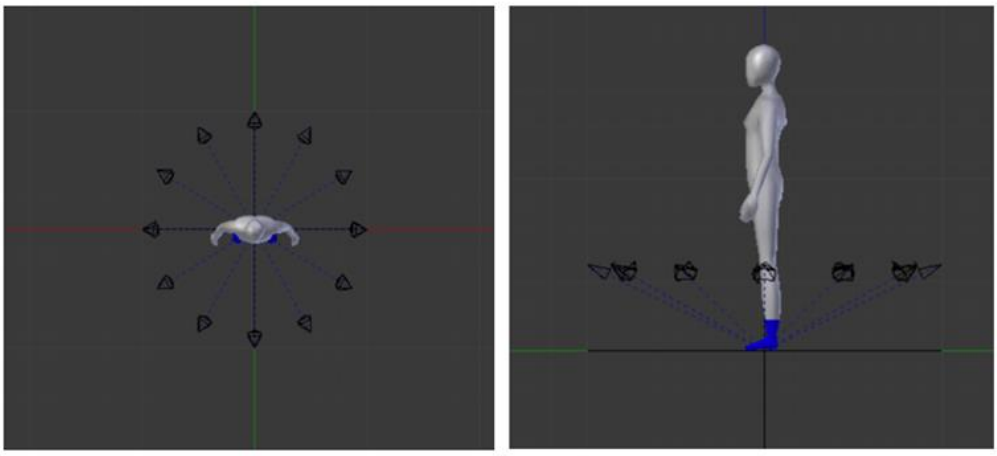


Figure 4.3 Camera setup around the target area

$$T_i = V_i + I_i \tag{4.1}$$

$$P = \frac{\sum_{i=1}^{12} V_i}{\sum_{i=1}^{12} T_i} \times 100 \tag{4.2}$$

4.4.2 Results

The target area's visibility percentage while walking was plotted for each instance along the hallway (Figure 4.4). Visibility with two sensors was less than the other setups. Four sensor and six sensor setups had similar visibility. The capture hallway length for six sensors setup was greater than other setups (Table 4-2). The sensors were operating at lower range for six sensors setup, so the accuracy would be better than other two setups.

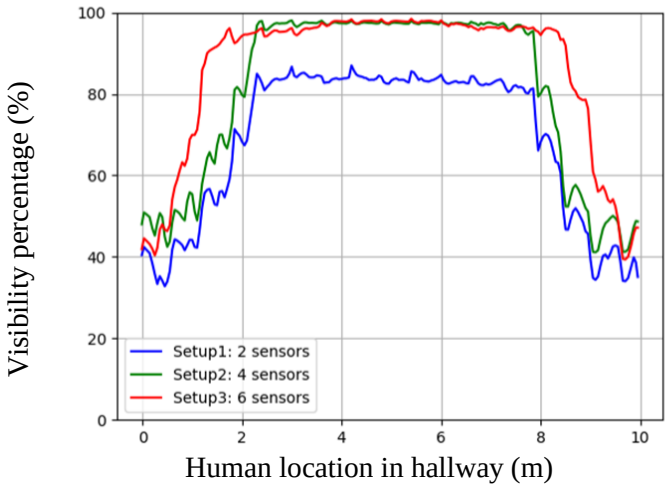


Figure 4.4 Target visibility percentage

Table 4-2 Simulation results

Setup	Capture hallway length	Average foot visibility
2 sensors	5.6m	83.33 %
4 sensors	5.65m	97.07 %
6 sensors	7.15m	96.18 %

4.4.3 Physical setup

Six Intel RealSense D415 sensors were placed in the same configuration as Setup3 (Figure 4.5) at 0.8m from the floor. The horizontal distance between sensors (i.e., distance between pairs (1,6), (2,3), (3,4), (5,6)) was 1.4m and vertical distance (i.e., between pairs (1,2), (3,6), and (4,5)) was 1.8m. Intel D415 units had noisy and inconsistent data at a 5m range. Therefore, the camera setup was brought down from the ceiling to waist height.

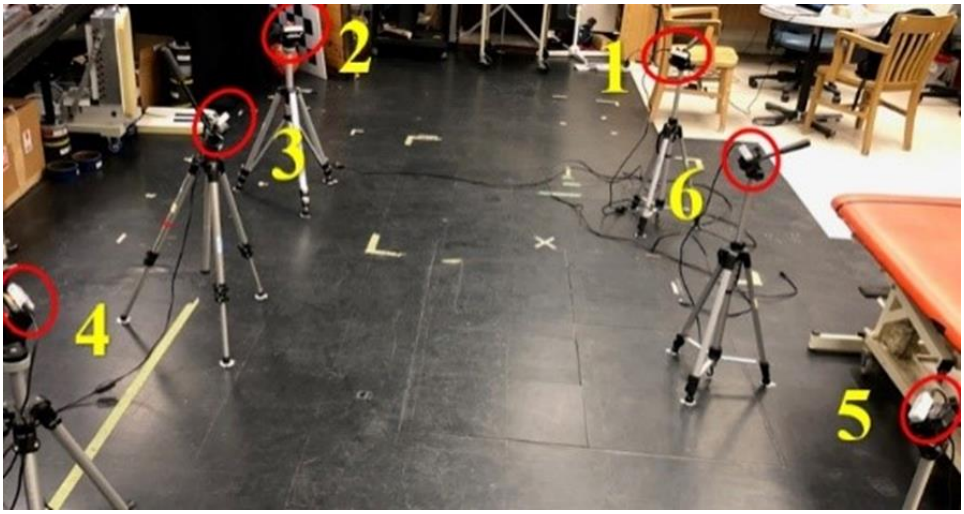


Figure 4.5 Physical setup of six sensors

4.5 Synchronization

4.5.1 Time Synchronization

Time Synchronization used a server-client approach with each sensor connected to a laptop that acted as a server. These laptops were connected to another laptop (client) through Ethernet.

All server laptops were triggered by the client at the same time. Server laptops captured scene data (depth image and RGB image, 848 × 480 pixels, 60fps) and timestamp, parallel down-sampled the depth image to half the resolution (424 × 240), and sent a message to the client. After signals were received from all servers, the client sent a message to capture the next frame data. All servers were forced to capture data within 30ms. If any server was not able to capture, server data for that frame was ignored.

Depth image down-sampling used a non-zero median of each group of four (2 × 2) pixels [93]. Down-sampling increased accuracy and reduced computation time.

4.5.2 Point Cloud Data from Depth Images

Each RealSense D415 unit had color and depth coordinate systems. Intrinsic parameters (focal length, principal point) and extrinsic parameters (rotation and translation between coordinate systems) were available from the sensor.

For focal length (f_x, f_y) and principal point (c_x, c_y) intrinsic parameters, a 3D point (x, y, z) for a pixel value (d) in depth frame at row (r) and column (c) was calculated using (4.3), (4.4), and (4.5).

$$x = \frac{(c - c_x) \times d}{f_x} \quad \textbf{(4.3)}$$

$$y = \frac{(r - c_y) \times d}{f_y} \quad \textbf{(4.4)}$$

$$z = d \quad \textbf{(4.5)}$$

4.5.3 Spatial Synchronization

Spatial synchronization identified sensor physical location with respect to a reference sensor coordinate system (transformation matrix between the reference and sensor coordinate systems), using a traditional chessboard stereo calibration method [94]. An 8 x 6 chess board with each square side 108mm was used as calibration tool. A screenshot of detected corners on the chessboard while stereo calibration between two sensors was shown in Figure 4.6. These corners pixels were the key points and accuracy of calibration results depends on them significantly, so high resolution (1280 × 720) color images were used and the result was a transformation matrix between two sensor's color coordinate systems.

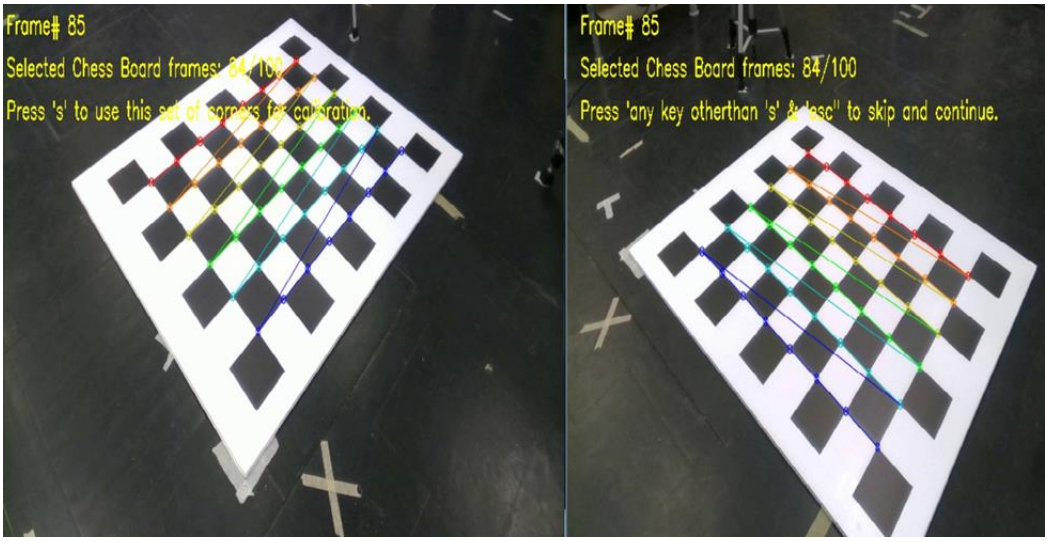


Figure 4.6 Corners detected on a chessboard for stereo calibration

4.5.4 Merging Six Sensor's Data into a Point Cloud

A time synced point cloud in the depth coordinate system was obtained for each sensor and transformed into a color coordinate system. All point cloud data were transformed into the first sensor's color coordinate system. Transformation matrix from j^{th} to i^{th} sensor's color coordinate system was represented as $T_{i\leftarrow j}$ and transformation matrix from j^{th} to the first sensor was calculated using (4.6). A point (Pt_j) from j^{th} sensor's point cloud was transformed to point (Pt_1) in the first sensor's coordinate system with (4.7). The point cloud was statistical outlier filtered [95], smoothed [96], and down-sampled [97] using Point Cloud Libraries.

$$T_{1\leftarrow j} = \prod_{k=2}^{j} T_{1\leftarrow k}\,; j > 1 \qquad \textbf{(4.6)}$$

$$Pt_1 = T_{1\leftarrow j} \times Pt_j; j > 1 \qquad \textbf{(4.7)}$$

4.5.5 Evaluation

A 500 continuous data set (500 x 6 servers = 3000 depth and 3000 RGB frames) was captured at 60 fps. Speed was consistent for all sensors, indicating time sync for all the six sensors (Figure 4.7).

A basketball with a radius (R) of 119.6mm was placed at the center of capture volume. The background subtracted [89] merged point cloud of size $'N'$ was fitted to a RANSAC sphere model [88]. From the RANSAC model, the estimated sphere radius (R_{est}), estimated center (C_{est}) and the number of inlier points to RANSAC sphere (N_m) were calculated.

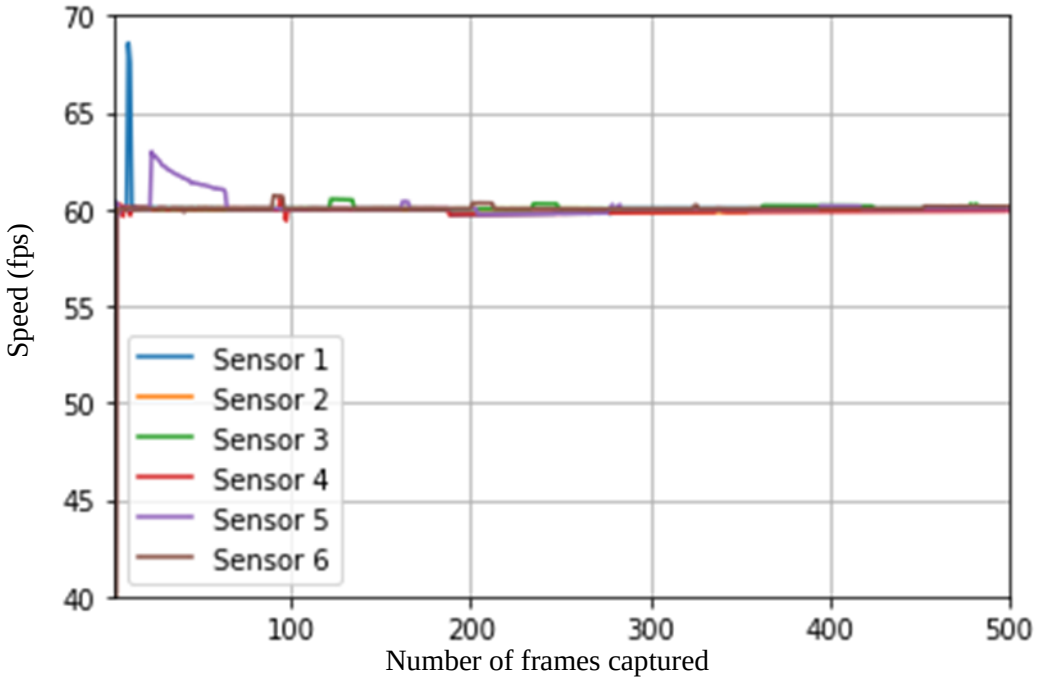


Figure 4.7 Time sync results showing frames captured vs speed

Points farther than 135mm from C_{est} were removed (i.e., noise) then point cloud (size N_e), number of points farther than R from C_{est} (N_+), number of points nearer than R from C_{est} (N_-), distance of i^{th} point from C_{est} (d_m^i), distance of i^{th} farther point from C_{est} (d_+^i), distance of i^{th} nearer point from C_{est} (d_-^i), and distance of i^{th} point from C_{est} (d^i) were calculated. Standard deviation towards the sphere's exterior (σ_+), standard deviation towards the sphere interior (σ_-), root mean square error (R_{rms}), and average distance of points from C_{est} (r_{avg}) were calculated using (4.8) – (4.11). Averaged values over 100 frames were shown in Table 4-3.

$$\sigma_+ = \sqrt[2]{\frac{\sum_{n=1}^{N_+}(R - d_+^n)^2}{N_+}} \quad \textbf{(4.8)}$$

$$\sigma_{-} = \sqrt[2]{\frac{\sum_{n=1}^{N_{-}}(R - d_{-}^{n})^2}{N_{-}}} \quad \textbf{(4.9)}$$

$$R_{rms} = \sqrt[2]{\frac{\sum_{n=1}^{N_m}(R - d_{m}^{n})^2}{N_m}} \quad \textbf{(4.10)}$$

$$r_{avg} = \frac{\sum_{n=1}^{N_e} d^{i}}{N_e} \quad \textbf{(4.11)}$$

Table 4-3 Spatial synchronization results

Variable	Value
Ball radius (R)	119.26mm
Number of point cloud points (N)	2186.76
Number of sphere inlier points (N_m)	1945.15
Calculated sphere radius (R_{est})	111.26mm
Average distance from center (r_{avg})	110.58mm
Exterior standard deviation (σ_{+})	6.9mm
Interior standard deviation (σ_{-})	12.24mm
Root mean square error (R_{rms})	11.83mm

The error in values from the merged point cloud was consistent with the theoretical limitation of approximately 10mm error at 2.5m from the sensor [92]. A sample merged human walking point cloud is shown in Figure 4.8.

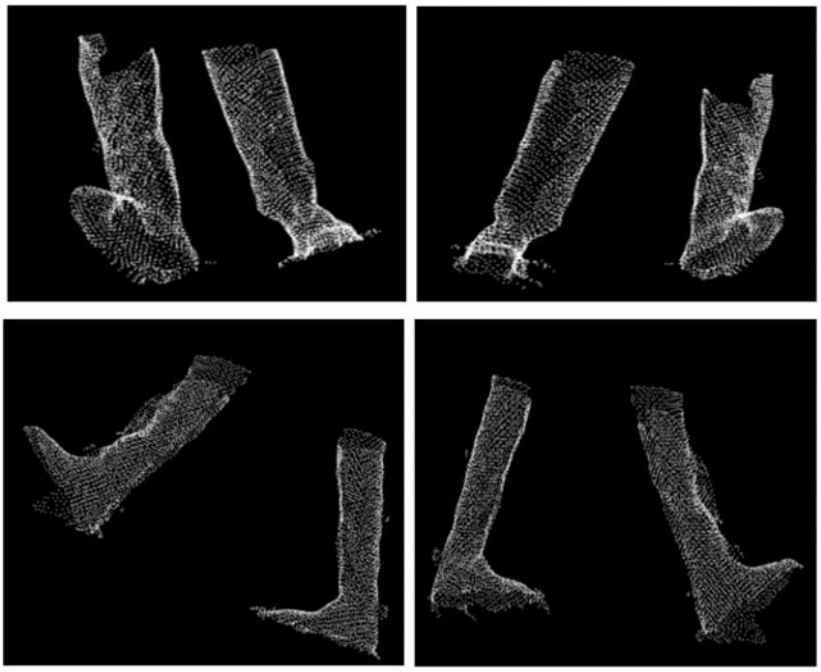


Figure 4.8 Front (top left), back (top right), left (bottom left) and right (bottom right) views of a merged point cloud of human walking

4.6 Conclusion

From the simulation results, capturing the foot surface of a walking person would need a minimum of four sensors positioned diagonally to the person.

Intel RealSense D415 depth sensors were suitable for this multiple sensor application because of non-interference when simultaneously capturing depth data. Due to limits of distance versus accuracy for stereo technology-based depth sensors, sensors should be as close as possible to the scene or object. The method for time synchronization showed consistent performance at 60fps.

5 Foot Tracking and Stride Analysis

5.1 Overview

Addressing Objective3 and Objective4, this chapter presents a novel algorithm to track the foot in a walking human's 3D point cloud data obtained from the depth sensors in a hallway setup. Using the tracking results, stride parameters were calculated and compared with the gold standard Vicon system. Modified contents of this chapter were submitted to the IEEE Transactions in Human-Machine Systems as a manuscript titled as "Development of a Smart Hallway for Marker-less Human Foot Tracking and Stride Analysis".

5.2 Abstract

In this chapter, a marker-less 'smart hallway' is proposed where stride parameters are computed as a person walks through an institutional hallway. A novel algorithm was developed to track a human foot using combined point cloud data obtained from multiple Intel RealSense D415 sensors. A method was implemented to separate the left and right leg point cloud data, then find the average foot dimensions. Foot tracking was achieved by fitting a box with average foot dimensions to the foot, with the box's base on the foot's bottom plane. Smart hallway results were validated by comparing marker-less system stride parameters with Vicon system output. With the smart hallway system frame rate at approximately 60fps, temporal stride parameter differences between systems were less than 10ms. Random noise around the foot's point cloud was observed, especially during foot strike phases. This caused errors in medial-lateral axis dependent parameters such as step width and foot angle. This novel marker-less smart hallway approach delivered promising results for stride analysis with minimal errors for temporal stride parameters and reasonable errors for medial-lateral spatial parameters.

5.3 Introduction

Human stride analysis in clinical settings is often performed with optical marker tracking systems such as Vicon, Optitrack, and Qualisys requiring expensive setup and dedicated laboratory space. Passive markers are placed on the body to track limbs. Active markers such as light-emitting diodes (LED) can be also used with cameras for human gait acquisition and

characterization [98]. Inertial Measurement Units (IMU) can also be attached to body parts to enable inertial motion capturing technologies [23], [99]–[101] to record kinematic gait data. Affixing external sensors on the human body may cause discomfort to patients and substantially change their natural gait [27]. An additional factor is that these systems require technical expertise for attaching markers and conducting experiments.

The release of low-cost Kinect depth sensors for gaming showed a vast potential and has been well researched for human gait-related health care applications; such as, fall risk [69], [70], [74], [76], Parkinson's disease movement assessment [60], fall detection of people with multiple sclerosis-related to nervous system [63], autism disorder identification [61], classification of abnormal gait [59], [62], virtual gait training programs [78], diagnosis, monitoring, and rehabilitation [79].

Depth sensors capture both depth and color images. Depth data contains distances, at each pixel, between the depth sensor and objects in the capturing scene. Popular Kinect V2 systems can identify and track the majority of human joints by defining joint points that constitute a human skeleton-model or building a human model from the scene's point cloud. However, gaps and limitations exist with lower body tracking, especially at the ankle and foot [6], [80]. Kinect's machine learning-based skeleton tracking is not reliable, with tracking points sometimes moving outside the body and tracking varying with viewing angle [81]. Given these limitations, approaches that use the whole point cloud could provide better foot tracking results.

Patients, staff, and visitors typically move through similar hallways in hospitals, rehabilitation centers, or long-term care facilities [90]. This space could be utilized in an intelligent way by performing marker-less stride analysis as patients walk through the hallway. Measuring patients every time they walk through the hallway, without intervention, could help identify changes in their movement status. Due to narrow hallway spaces and to avoid occlusions while a person is walking, multiple depth sensors would be required. Stereoscopic based Intel RealSense D415 depth sensors are suitable for this application since they did not experience interference when multiple sensors were used simultaneously [7]. Furthermore, these sensors capture data at 60fps, which can provide sufficient data for analyzing temporal related stride parameters.

To explore depth sensing technology for stride parameter analysis within an institutional hallway environment, this research developed and prototyped a novel point-cloud-based marker-less

system and assessed system performance by comparing stride parameter output with industry standard marker-based motion analysis. Successful implementation of this smart hallway concept would introduce unobtrusive movement status assessment that can guide clinical decision-making, without introducing unsustainable human resource requirements. This would also become the basis for future data analytics applications for predicting changes in dementia, fall risk, or other aging-related conditions.

5.4 System Setup and Synchronization

In our previous work [8], a physical setup with six Intel RealSense D415 was configured in the lab to replicate a hallway scenario (Figure 5.1). All sensors focused downwards to capture the foot and lower leg, with 4 cameras capturing the leg at all times. Data were trimmed at the both ends of this setup because at least four sensors did not capture the foot. The middle 1.5m was defined as the capture zone. Techniques from our previous work were used for temporal synchronization (client-server approach) and spatial synchronization (stereo chess board method).

This system captures 848 × 480 size depth images and color images at approximately 60fps, with a timestamp for each frame, from all the six sensors. To increase accuracy and reduce computation time, a non-zero median filter for every 2 × 2 pixels was applied to remove “spikes” in the depth data and down-sample to half resolution (424 × 240) [93].

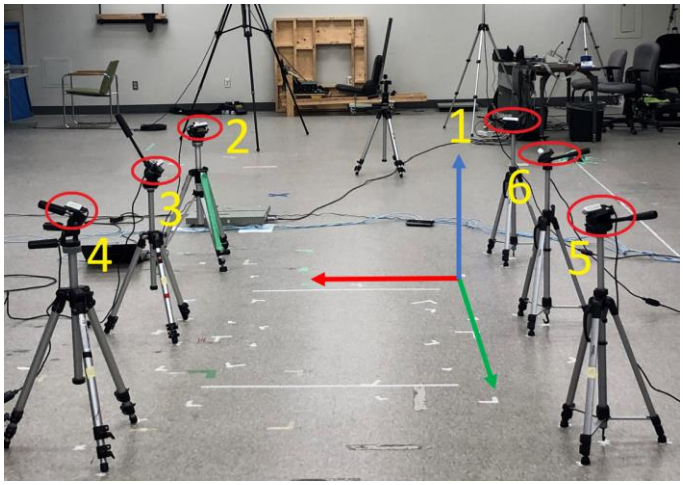


Figure 5.1 Physical setup of marker-less system using hallway dimensions (Red: x-axis, Green: y-axis, and Blue: z-axis)

5.4.1 Sensor Parameters

Intrinsic parameters such as focal length (f_x, f_y) and principal point (c_x, c_y) for Intel RealSense D415 depth and color cameras were obtained from the manufacturer. Two coordinate systems (depth, color) and extrinsic parameters (rotation and translation, to transform data between depth and color coordinate systems) were also obtained. This marker-less six depth sensors setup was spatially synchronized in the color coordinate system, such that the combined output point cloud was in the first sensor's color coordinate system.

5.4.2 Room Coordinate System

A reference coordinate system on the floor plane (Figure 5.1) was designed, with x-axis in the medial-lateral (ML) walking direction, y-axis parallel to the pathway (anterior-posterior; AP), and z-axis outwards to the floor (Vertical; V). This reference coordinate system was labelled as the Room Coordinate System (RCS). Point cloud data from the first sensor's color coordinate system was transformed into the RCS by determining a transformation matrix '$T_{R\leftarrow 1}$'.

A chessboard (8 × 6) was placed on the floor with the horizontal edge parallel to the RCS x-axis and vertical edge parallel to the y-axis. The board's depth and color images were captured with the first sensor at 1280 × 720 resolution. The depth image was down-sampled to half resolution using median filter, 3D points were calculated from the depth image using depth intrinsic parameters, and then points were transformed into the color coordinate system using extrinsic parameters.

A new color image was constructed using 3D points and corresponding projection pixels onto the captured color image. For a 3D point (x, y, z) in the color coordinate system corresponding to row 'r_d' and column 'c_d' in the depth image, the projected pixel location ('r_c','c_c') in the captured color image was found using (5.1) and (5.2) (color camera's intrinsic parameters used in the equations). The red, green, and blue channel values at row 'r_d' and column 'c_d' in the constructed color image (Figure 5.2) were the values at row 'r_c' and column 'c_c' in the captured color image and the values of pixels corresponding to invalid 3D data (0,0,0) were set to zeroes.

Three key points k_1, k_2, and k_3 were identified in 100 frames. For each key point, 100 instants of its 3D location obtained in the color coordinate system. Each dimension value (x, y, and z) of these 100 3D points was sorted separately and the middle 50 values were averaged.

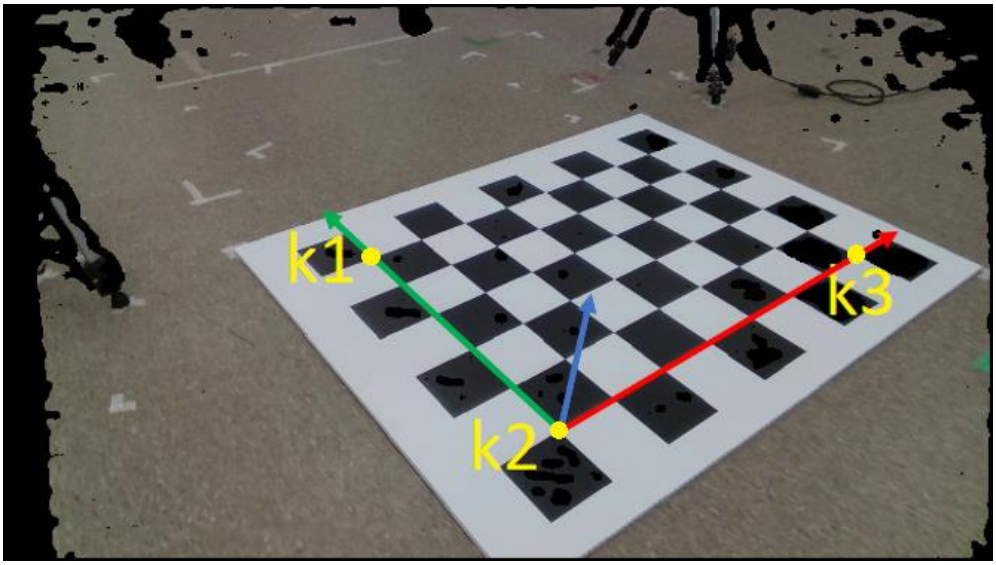


Figure 5.2 Key points on the chessboard for room coordinate system (Red: x-axis, Green: y-axis, and Blue: z-axis)

$$r_c = \left\lfloor \left(\frac{y \times f_y}{z} \right) + c_y \right\rfloor \quad \textbf{(5.1)}$$

$$c_c = \left\lfloor \left(\frac{x \times f_x}{z} \right) + c_x \right\rfloor \quad \textbf{(5.2)}$$

The RCS origin was at point $k_2(x_0, y_0, z_0)$, the unit vectors were $\widehat{x_R}$ $(x_R^x,\ y_R^x, z_R^x)$ from k_2 to k_3, $\widehat{y_R}$ $(x_R^y,\ y_R^y, z_R^y)$ from k_2 to k_1, and $\widehat{z_R}$ $(x_R^z,\ y_R^z, z_R^z)$ the cross product of $\widehat{x_R}$ and $\widehat{y_R}$ with respect to the first sensor color coordinate system, whose origin was at the point (0,0,0) and corresponding unit vectors were $\hat{x}(1,0,0)$, $\hat{y}(0,1,0)$, and $\hat{z}(0,0,1)$ respectively. The transformation matrix from the first sensor to the RCS was obtained using (5.3).

$$T_{R \leftarrow 1} = \begin{bmatrix} x_R^x & x_R^y & x_R^z & x_0 \\ y_R^x & y_R^y & y_R^z & y_0 \\ z_R^x & z_R^y & z_R^z & z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \quad \textbf{(5.3)}$$

5.5 Point Cloud

A point cloud was generated from the six depth sensors. The process involved generating a background depth image from a static scene; subtracting the background information from the depth images; constructing walking human point cloud data for each sensor from the background subtracted depth images; and merging and transforming point clouds from the six sensors to RCS. The combined point cloud was filtered and smoothed to reduce noise.

5.5.1 Background Frame

From each sensor, 1000 depth frames of background data (without any objects) were captured. The pixel value at row 'y', column 'x' of these background frames was represented as $bf_{yx}^{(j,i)}$ for the i^{th} frame of j^{th} sensor. This system was designed to work in the range of 200mm to 5000mm. All background frame pixels for the j^{th} sensor $\left(BF^{(j)}\right)$ were initialized with 5000 as shown in (5.4), then the pixel value at row 'y', column 'x' ($BF_{yx}^{(j)}$) was updated with the minimum of $BF_{yx}^{(j)}$ and $bf_{yx}^{(i)}$, iterating through 1000 frames ($i = 1$ to 1000) using (5.5).

$$BF_{yx}^{(j)} = 5000; 1 \leq y \leq 424, 1 \leq x \leq 240 \tag{5.4}$$

$$BF_{yx}^{(j)} = \begin{cases} BF_{yx}^{(j)} \; ; \; bf_{yx}^{(j,i)} < 200 \; or \; bf_{yx}^{(j,i)} > 5000 \\ BF_{yx}^{(j)} \; ; 200 < bf_{yx}^{(j,i)} < 5000 \; , BF_{yx}^{(j)} \leq bf_{yx}^{(j,i)} \\ bf_{yx}^{(j)} \; ; \; 200 < bf_{yx}^{(j,i)} < 5000 \; , BF_{yx}^{(j)} > bf_{yx}^{(j,i)} \end{cases} \tag{5.5}$$

5.5.2 Background Subtraction

A background subtracted depth image for the j^{th} sensor ($BS^{(j)}$) was obtained by pixel-wise comparison with the corresponding sensor's background frame ($BF^{(j)}$) [89]. For a depth frame from the j^{th} sensor ($DF^{(j)}$), pixel values less than the background frame's pixel value, and greater than the minimum value (200 mm), were considered the same value in the $BS^{(j)}$ frame. Other pixel values were assigned the maximum value (5000 mm), as presented in (5.6), for a pixel in y^{th} row and x^{th} column. For further processing, $BS^{(j)}$ was linearly scaled down to [0, 255] from [0, 5000]. From the scaled-down image ($SBS^{(j)}$), a Binary Background Subtracted

image ($BBS^{(j)}$) was constructed based on (5.7).

A connected component filter [102] with 1000 pixels connected area cut-off was applied to the $BBS^{(j)}$ image and output was a Binary Filtered Background Ground Subtracted image ($BFBS^{(j)}$). The $BS^{(j)}$ image was modified based on the $BFBS^{(j)}$ (5.8). A sample BFBS frames from all sensors shown in the Figure 5.3. White pixels in the BFBS frames were foreground and black pixels were background. Depth data was not captured in the small gaps among foreground.

5.5.3 Point Cloud Construction

3D point cloud points were constructed from each sensor's background-subtracted depth images and then transformed into the first sensor's coordinate system as in our previous work [8]. This "combined point cloud" was transformed into RCS using the transformation matrix $'T_{R\leftarrow 1}'$ (5.3). For a point $Pt_1(x, y, z) = [x_1, y_1, z_1]$ in the first sensor's coordinate system, the corresponding point in the RCS, $Pt_R(x, y, z) = [x_R, y_R, z_R]$ was calculated using (5.9).

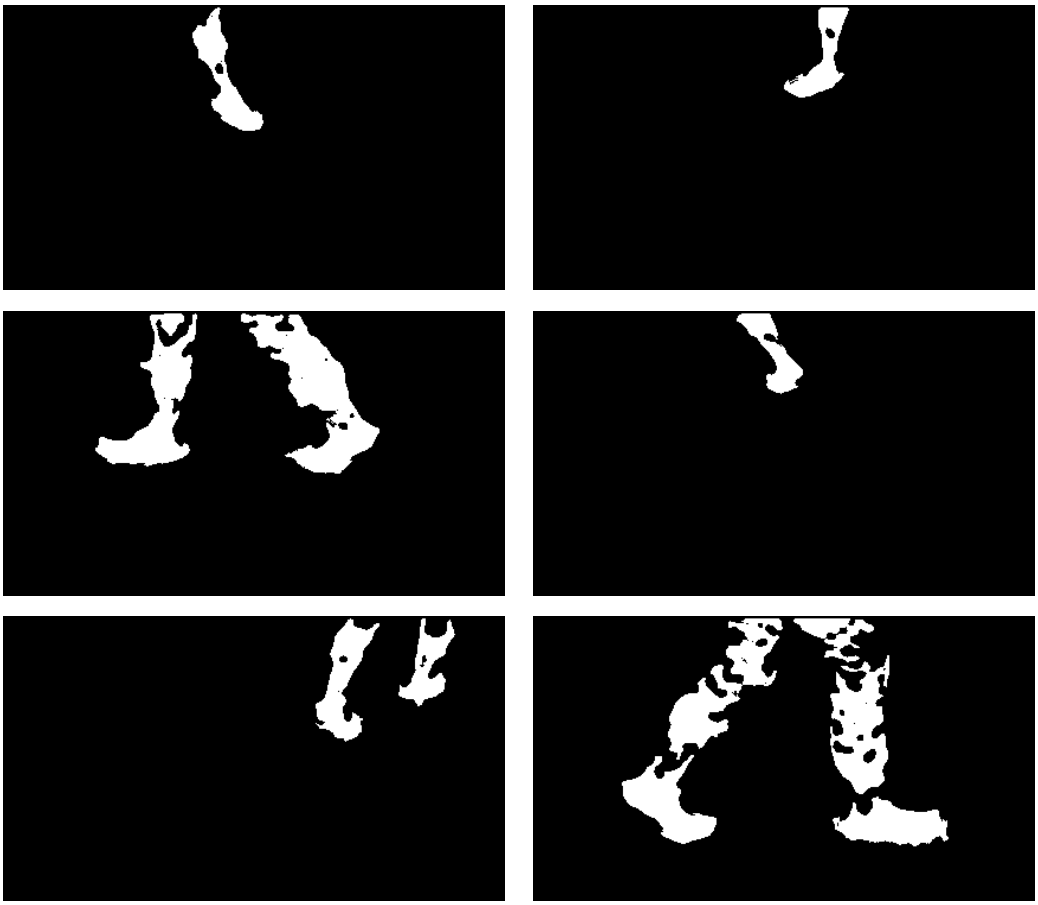


Figure 5.3 Background subtracted binary images from six sensors

$$BS^{(j)}_{yx\ y\in\{1,2,\dots,424\},x\in\{1,2,\dots,240\}} = \begin{cases} DF^{(j)}_{yx}; & DF^{(j)}_{yx} < BF^{(j)}_{yx}, DF^{(j)}_{yx} > 200 \\ 5000; & Other\ wise \end{cases} \quad \textbf{(5.6)}$$

$$BBS^{(j)}_{yx\ y\in\{1,2,\dots,424\},x\in\{1,2,\dots,240\}} = \begin{cases} 0; & SBS^{(j)}_{yx} \geq 250 \\ 255; & SBS^{(j)}_{yx} < 250 \end{cases} \quad \textbf{(5.7)}$$

$$BS^{(j)}_{yx\ y\in\{1,2,\dots,424\},x\in\{1,2,\dots,240\}} = \begin{cases} 0; & BFBS^{(j)}_{yx} = 0 \\ BS^{(j)}_{yx}; & BFBS^{(j)}_{yx} = 255 \end{cases} \quad \textbf{(5.8)}$$

$$\begin{bmatrix} x_R \\ y_R \\ z_R \\ 1 \end{bmatrix} = T_{R\leftarrow 1} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} \quad \textbf{(5.9)}$$

5.5.4 Filtering

The combined point cloud was filtered using a statistical outlier filter [95], smoothed with a moving least-squares technique [96], and then down-sampled with a voxel grid filter [97]. OpenCV libraries [103] were used for 2D image processing and PCL [104] for 3D point cloud processing.

For every 3D point in a point cloud, 100 neighbor points were analyzed to find outliers. Mean and standard deviation of distances of the closest 100 points from each point of interest were found. Points farther than one standard deviation from the point of interest were considered outliers and removed.

Point cloud points were smoothed by fitting a second-order polynomial equation to points within 30mm of each point of interest in the point cloud. The point cloud was divided into 5mm × 5mm × 5mm voxels (3D boxes) and then down-sampled by replacing points in a voxel with the centroid of these points. This method of down-sampling retained the point cloud surface and reduced computation time for point cloud processing.

5.6 Leg Segmentation

Since this research is concerned with tracking a walking person's foot, point cloud points less than 70cm from the floor were selected, since the foot and shank are always present in this region. This lower leg point cloud was divided into left and right leg points clouds. To segment a current point cloud frame, Euclidean clustering, average leg dimensions (calculated from the point cloud data), and past point cloud frames were used.

5.6.1 Euclidean Clustering

Point cloud points were divided into clusters based on the Euclidean distances between points [105]. The clustering tolerance was 50mm, which implies that the points within 50mm radial distance from an interested point in the point cloud were clustered together.

Each cluster was verified using the number of points and cluster volume (i.e., volume of bounding box around the cluster). Point clouds with two clusters, each with a minimum of 1000 points and cluster volume greater than 75 percent of average leg volume were considered to contain data from two legs, and each cluster was considered an individual leg (Figure 5.4). Point clouds with a single cluster greater than 1000 points and volume between 0.75 to 1.25 times the average volume was considered single leg data.

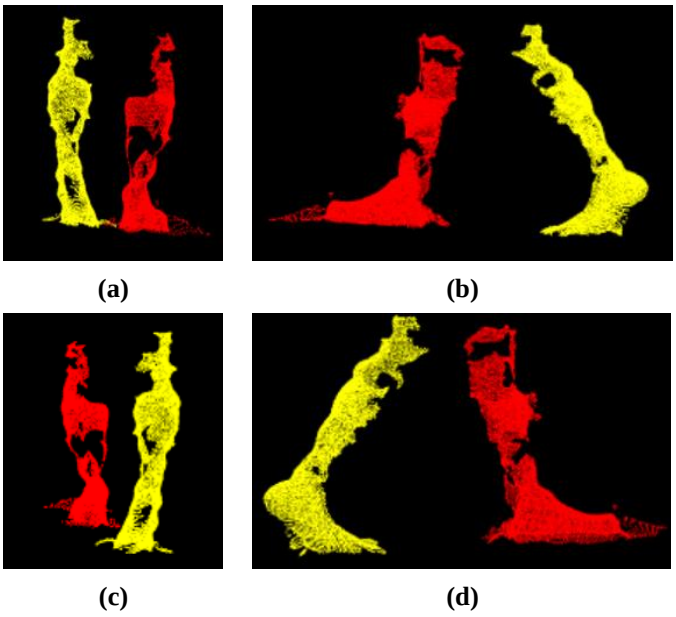


Figure 5.4 (a) Front, (b) left, (c) back, and (d) right views of left (red) and right (yellow) leg points segmented using Euclidean clustering

During mid-swing, when both legs are close together, noise between the legs caused the points to group into a single cluster. In these cases, the two legs data were identified as a single cluster (Figure 5.5), with cluster volume greater than two times the average leg volume. Therefore, a different approach was used to segment the legs ("Moving Points Segmentation").

Point cloud frames that were not in one of these three categories (Two valid legs, single leg, two legs as a single cluster) were ignored.

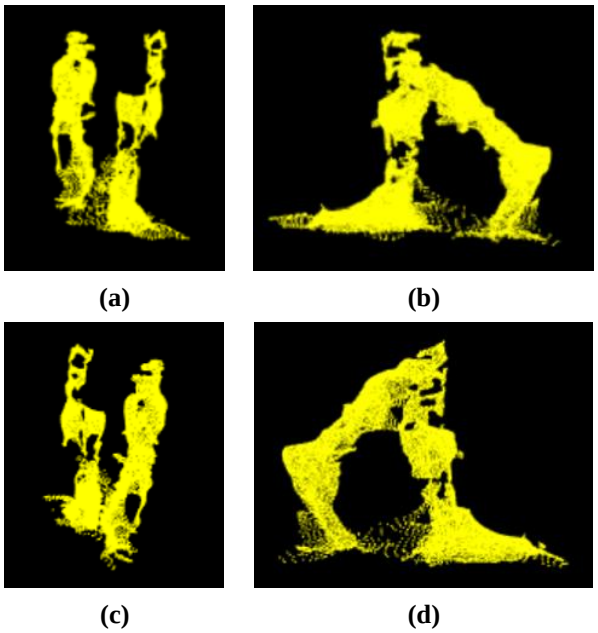


Figure 5.5 (a) Front, (b) left, (c) back, and (d) right views of a single clustered point cloud with two legs

5.6.2 Leg Dimensions

Leg dimensions were calculated from a closest oriented bounding box (OBB, Table 5-1) around the leg point cloud. Frames with two separate Euclidean clusters (two legs) and each leg with more than 1000 points were considered for calculating average leg dimensions, from 40 valid leg point clouds (Table 5-2). Dimensions (l, w, h) were calculated using Algorithm I (Table 5-1) and then sorted before calculating the average of center 20 elements for each dimension.

Algorithm I oriented the bounding box around the point cloud data (Table 5-1). This was achieved by calculating the point cloud's centroid, to locate the bounding box, and then determining box rotations to orient the box around the point cloud.

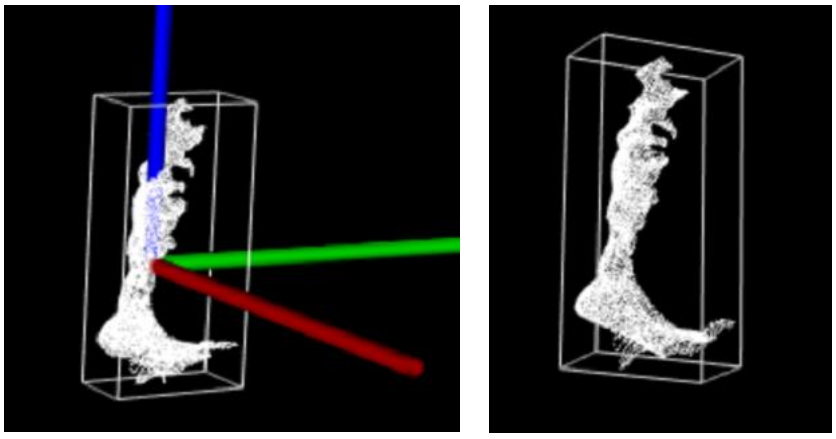
Table 5-1 Algorithm I: Oriented bounding box around point cloud data

Step	Methodology
1	Calculate the point cloud's ($'PC'$) centroid (C) and normalized covariance matrix (COV_N). For COV_N, each covariance matrix element is divided by the number of points in the point cloud. The covariance matrix, also known as dispersion matrix, contains information on dispersion direction (Eigen vectors) and dispersion magnitude (Eigen values).
2	Calculate Eigen vectors ($[v_1 \quad v_2 \quad v_3]$) for COV_N. Each vector is an OBB axis, where v_3 is replaced with a cross product of v_1, v_2 such that axes obey the right-hand rule.
3	Find an initial 4×4 transformation matrix (T_{OBB}^{init}) formed using centroid and Eigen vectors as $T_{OBB}^{init} = \begin{bmatrix} v_1 & v_2 & v_3 & C \\ 0 & 0 & 0 & 1 \end{bmatrix}$
4	Find point cloud $'PC_o'$ with centroid at origin by transforming PC. The relation between point p_s in PC, its corresponding point p_o in PC_o is $\begin{bmatrix} p_s \\ 1 \end{bmatrix} = T_{OBB}^{init} \begin{bmatrix} p_o \\ 1 \end{bmatrix}$. A transformed PC_o is shown in Figure 5.6(a).
5	Find the minimum (min_x, min_y, min_z) and maximum (max_x, max_y, max_z) of each x, y, z dimension. Obtain OBB dimensions from the absolute difference between minimum and maximum values in each dimension.
6	Calculate OBB position (POS_{OBB}) $\begin{bmatrix} POS_{OBB} \\ 1 \end{bmatrix} = T_{OBB}^{init} \begin{bmatrix} \frac{min_x + max_x}{2} \\ \frac{min_y + max_y}{2} \\ \frac{min_z + max_z}{2} \\ 1 \end{bmatrix}$ The OBB rotation matrix is $R_{OBB}^{init} = [v_1 \quad v_2 \quad v_3]_{3\times3}$. With these dimensions, position, and rotation, an OBB around the leg point cloud can be drawn (Figure 5.6(b)).
7	Since rotations are not unique, PC_o could be aligned in different ways (Figure 5.7). The following steps ensure that foot AP is along the $x - axis$, ML is along the $y - axis$, and the bottom of foot is in negative $z - axis$. This aligned point cloud at origin is represented as PC_o^{align}.
8	The four closest corners to floor are considered OBB bottom corners. The rest are top corners. Any bottom corner can be considered the reference point. Distances from the reference point to the remaining three bottom corners are calculated. Minimum distance is OBB width (w), maximum distance is diagonal, and middle distance is OBB length (l). The minimum distance from the reference corner to the top corners is the OBB height (h).
9	From a reference point, calculate a unit vector in the length direction ($\hat{l}$) and a unit vector in the height direction ($\hat{h}$).
10	Calculate the rotation matrix (R_{OBB}) between a coordinate system with point $O(0,0,0)$ as the origin, unit vectors $\hat{x}(1,0,0)$, and $\hat{z}(0,0,1)$ and the other coordinate system with corresponding values $POS_{OBB}, \hat{l}$, and $\hat{h}$. This R_{OBB} is the new rotation matrix for the OBB.

Algorithm II identified valid frames for measuring average foot and leg dimensions. Valid frames had the complete foot aligned parallel to the bottom OBB plane (Table 5-2).

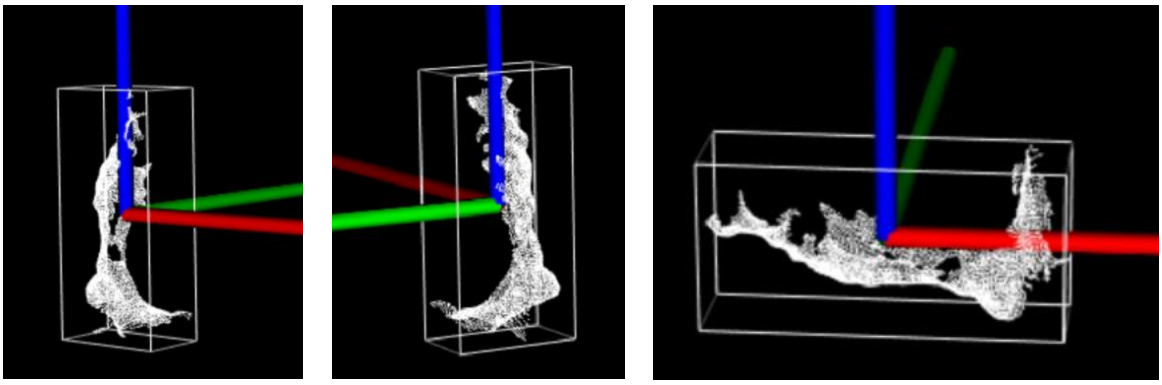
Table 5-2 Algorithm II: Valid frames to measure dimensions

Step	Methodology
1	Calculate $POS_{OBB}, R_{OBB}, l, w, and\ h$ for a leg point cloud (PC_l), using Algorithm I (Table 5-1).
2	For every point (p_l) in PC_l, a corresponding point (p_o) in the point cloud PC_o^{align} is calculated using $\begin{bmatrix} p_o \\ 1 \end{bmatrix} = \begin{bmatrix} R_{OBB} & POS_{OBB} \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} p_l \\ 1 \end{bmatrix}$
3	A box with centroid at the origin, height (h) in the z-axis, width (w) in the y-axis, and length (l) in the x-axis bounds the PC_o^{align}.
4	Divide this box into 12 horizontal slices along the z-axis, starting from the bottom, with 25 mm height, length (l), and width (w).
5	For each slice, calculate minimum (min_x, min_y, min_z) and maximum (max_x, max_y, max_z) for each x, y, z dimension using points inside the slice. Calculate the slice volume as $V_{slice} = (max_x - min_x) \times (max_y - min_y) \times 25$.
6	A leg point cloud is considered valid for calculating dimensions if the maximum V_{slice} was one of bottom two slices and value greater than 80 percent of $l \times w \times 25$.



(a) **(b)**

Figure 5.6 (a) Point cloud $'PC_o'$ with centroid at origin, (b) OBB around $'PC'$ (Red: $x - axis$, Green: $y - axis$, and Blue: $z - axis$)



(a) **(b)** **(c)**

Figure 5.7 Leg PC_o with foot AP along (a) $x-axis$ (red), (b) $y-axis$ (green), (c) $z-axis$ (blue)

5.6.3 Moving Points Segmentation

For a current point cloud frame (PC_t) with two legs identified as a single cluster, a reference point cloud (PC_{ref}) from the past frames was found using Algorithm III (Table 5-3). For a current point cloud frame, Algorithm III identified a reference frame from previous frames. Point cloud points were projected onto the floor plane and their centroids were calculated. The reference frame's centroid was separated by a minimum distance of 100mm from the current frame's centroid.

Table 5-3 Algorithm III: Past reference point cloud frame

Step	Methodology
1	Project point cloud data (PC_t) of N^{th} frame onto the floor plane and calculate centroid on the floor (C_N).
2	Reference frame at $R = N - 5$ and $R > 0$.
3	If the number of points in the R^{th} frame's point cloud is less than 1000, go to step 5.
4	Project point cloud data at R^{th} frame onto the floor plane and calculate centroid on the floor (C_R).
5	R^{th} frame is the reference point cloud (PC_{ref}). If the distance between C_N and C_R is greater than 100 mm, else go to the next step.
6	For $R = R - 5$; if $R > 0$ go to step 3, else ignore the N^{th} frame.

From every point in the reference point cloud (PC_{ref}), points within 20mm in PC_t were categorized as the non-moving leg point cloud $'PC_1'$ (repeated points were ignored). All other points in PC_t were moving points and categorized as the other leg's point cloud $'PC_2'$. Each point cloud had at least 30 percent of total points in PC_t and the statistical outlier removal filter was applied (Section 5.5.4). Euclidean clustering (Section 5.6.1) was applied to PC_1 and PC_2, the biggest cluster from each point cloud was considered (Figure 5.8).

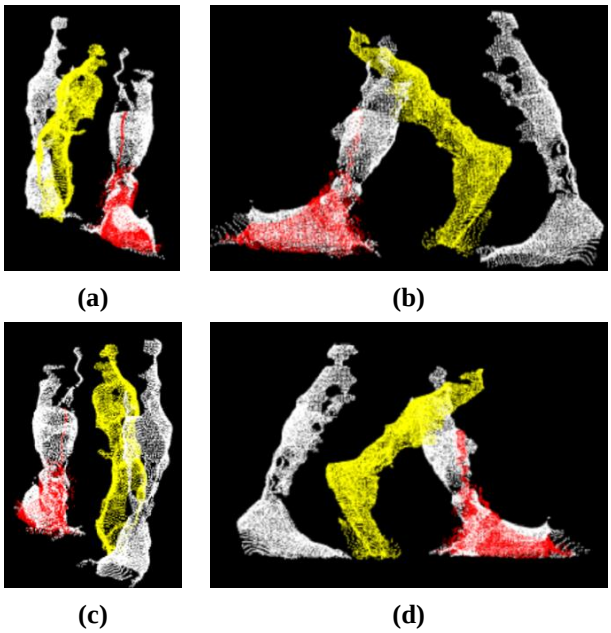


Figure 5.8 (a) Front, (b) left, (c) back, and (d) right views of a reference point cloud frame (white) and current point cloud frame segmented into non-moving leg points (red) and moving leg (yellow)

5.7 Foot Tracking

Foot tracking was achieved from point cloud data by fitting a box with average foot dimensions around each foot, in each frame. The foot's bottom plane was calculated and used to define bounding box rotation and position. The foot's heel and toe points were based on the walking direction.

5.7.1 Foot Dimensions

Using Algorithm II (Table 5-2), for a valid frame, the volume of points in 12 slices were calculated. The volumes of these 12 slices were median filtered with filter size=3 and filter stride length=1 (first and last elements were untouched).

The cut-off slice (i.e., slice defining top of foot) was defined by identifying the slice with the maximum volume (V_{slice}^{max}) and then scanning upwards to find the slice with volume less than 60 percent of V_{slice}^{max}. The points below this cut-off slice define the foot. An OBB was calculated around these points (Table 5-1, Algorithm I) and OBB dimensions were foot length (f_l), foot width (f_w), and foot height (f_h). These dimensions were found for forty frames, values of each dimension were sorted and the center 20 elements were averaged.

5.7.2 Foot Oriented Bounding Box

The foot's bottom plane was found using Algorithm IV (Table 5-4). Points above this plane within the distance f_h were considered to belong to foot point cloud (PC_{foot}) and points between 0.1 times f_h and 0.9 times f_h were segmented as the center foot's point cloud (PC_{foot}^{center}).

Algorithm IV set the foot bottom plane (Table 5-4). In this algorithm, foot corners points were defined and the best plane with these corner points was calculated.

Table 5-4 Algorithm IV: Foot bottom plane

Step	Methodology
1	Apply Algorithm I to a leg point cloud (PC_L) and align the outputs at origin (PC_o^{align}), OBB position (POS_{OBB}), rotation (R_{OBB}), length (l), width (w), and height (h).
2	From an OBB bottom corner around PC_o^{align}, five sub-boxes with length ($l/4$), width (w), and height ($h/10$) are taken. The two outer bottom corner points of first sub-box from the bottom, with a minimum of 50 points, are transformed using POS_{OBB} and R_{OBB}. These transformed points (p_c^{OBB}) are the foot OBB corners.
3	Repeat step 2 for the foot's other side.
4	From a point p_c^{OBB}, the centroid of thirty closest points in PC_L is considered the foot bottom point (f_b). Repeat to find all the four bottom points on the foot.
5	Using these four f_b points, four plane equations are obtained by taking different combinations of three points and distance from the fourth point to the plane is calculated. The plane with the smallest distance from the fourth point is the "Best Plane".
6	Points within 5mm above the best plane are fitted to a RANSAC plane model [106] and the output considered as foot's bottom plane.

PC_{foot}^{center} points were projected onto the foot's bottom plane and then Algorithm I was partially applied (Table 5-1, until step 6) since it was not possible to find four bottom corners for the projected points on the plane using the method described. OBB rotation around the foot ($R_{OBB}^{foot} = R_{OBB}^{init}$) was obtained with dimensions ($x_{OBB}^{foot}, y_{OBB}^{foot}, z_{OBB}^{foot}$). Then OBB dimensions mapped with average foot dimensions as the minimum of $x_{OBB}^{foot}, y_{OBB}^{foot}, z_{OBB}^{foot}$ were replaced with f_h, maximum with f_l, and the remaining dimension with f_w.

The foot's OBB position (POS_{OBB}^{foot}) was the centroid of PC_{foot}. Position (POS_{OBB}^{foot}), rotation (R_{OBB}^{foot}), and dimensions ($x_{OBB}^{foot}, y_{OBB}^{foot}, z_{OBB}^{foot}$) were used to locate a box around the foot.

5.7.3 Heel and Toe Segmentation

The point cloud data was transformed using a translation of 5000mm in x-axis and y-axis such that the walking pathway was always in the positive xy-plane. This reduced the complexity of further processing and understanding.

Left and right leg segmentation was based on the walking direction, calculated using OBB centroid trajectory. When walking towards the origin along a pathway parallel to the y-axis, the leg closer to the y-axis was the right leg and the other leg was labelled as left. Opposite leg classification was applied when walking away from the origin.

For each foot OBB, the center point (p_{OBB}^{toe}) of front two bottom corners and the center point (p_{OBB}^{heel}) of back two bottom corners were calculated using Algorithm V (Table 5-5) These points were considered as the toe and heel, respectively (Figure 5.9).

5.8 Validation

The foot tracking algorithm was validated by comparing gold standard Vicon system output with the marker-less smart hallway system. Volunteer walking trials were captured simultaneously with both the systems and post-processing filters were applied. This section describes the data collection protocol and post-processing processes.

Algorithm V segmented heel and toe for each foot (Table 5-5). The centroid of two back bottom corners of foot OBB was the heel and centroid of two front bottom corners was the toe.

Table 5-5 Algorithm V: Heel and toe segmentation

Step	Methodology
1	Calculate OBB centroid $(c_x^{OBB}, c_y^{OBB}, c_z^{OBB})$.
2	Identify OBB four front corners and four back corners. front corner $\begin{cases} \text{y} - \text{axis value} < c_y^{OBB}; \text{walking towards} \\ \text{y} - \text{axis value} > c_y^{OBB}; \text{walking away} \end{cases}$ back corner $\begin{cases} \text{y} - \text{axis value} > c_y^{OBB}; \text{walking towards} \\ \text{y} - \text{axis value} < c_y^{OBB}; \text{walking away} \end{cases}$
3	Calculate centroid of front corners $(fc_x^{OBB}, fc_y^{OBB}, fc_z^{OBB})$ and back corners $(bc_x^{OBB}, bc_y^{OBB}, bc_z^{OBB})$.
4	From front corners, front left and front right corners are: front left $\begin{cases} \text{x} - \text{axis value} > fc_x^{OBB}; \text{walking towards} \\ \text{x} - \text{axis value} < fc_x^{OBB}; \text{walking away} \end{cases}$ front right $\begin{cases} \text{x} - \text{axis value} < fc_x^{OBB}; \text{walking towards} \\ \text{x} - \text{axis value} > fc_x^{OBB}; \text{walking away} \end{cases}$
5	From back corners, back left and back right corners are: back left $\begin{cases} \text{x} - \text{axis value} > bc_x^{OBB}; \text{walking towards} \\ \text{x} - \text{axis value} < bc_x^{OBB}; \text{walking away} \end{cases}$ back right $\begin{cases} \text{x} - \text{axis value} < bc_x^{OBB}; \text{walking towards} \\ \text{x} - \text{axis value} > bc_x^{OBB}; \text{walking away} \end{cases}$
6	For front, the front left bottom corner has a lower z-axis value and the other is the front left top corner. The same logic is applied to front right corners, back left corners, and back right corners.
7	Calculate foot length, width, and height and validated with average foot dimensions (f_l, f_w, f_h). foot length $= \begin{cases} \lVert \text{front left bottom} - \text{back left bottom} \rVert \\ \lVert \text{front left top} - \text{back left bottom} \rVert \end{cases}$ foot width $= \begin{cases} \lVert \text{front left bottom} - \text{front right bottom} \rVert \\ \lVert \text{front left top} - \text{front right top} \rVert \end{cases}$ foot height $= \begin{cases} \lVert \text{front right bottom} - \text{front right top} \rVert \\ \lVert \text{back left bottom} - \text{back left top} \rVert \end{cases}$
8	OBB toe point is the center of front bottom two points (p_{OBB}^{toe}) and heel point is the center of back bottom two points (p_{OBB}^{heel}).

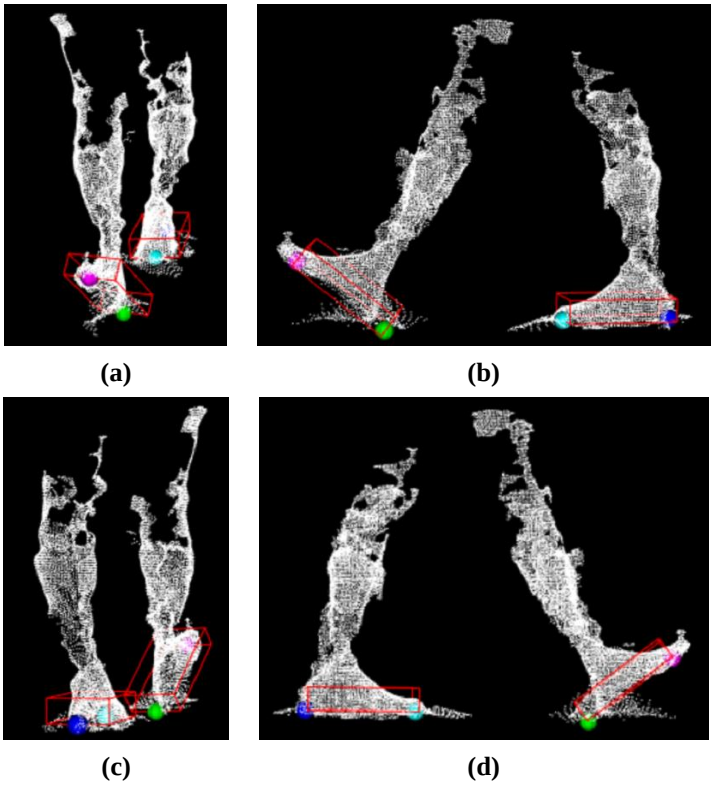


Figure 5.9 (a) Front, (b) left, (c) back, and (d) right views of a point cloud frame with heel and toe segmentation (right heel – green, right toe – magenta, left heel – blue, left toe – cyan)

5.8.1 Protocol

Twenty-two able-bodied volunteers were recruited from students and staff at the University of Ottawa. After informed consent, reflective markers were attached to the participant's lower body (Figure 5.10) and then the participant walked 12 times along a walkway with a 1.5m capture zone. Data were captured simultaneously with a 13 cameras Vicon system at 100Hz [107] and the new marker-less system at approximately 60Hz. This protocol was approved by the Research Ethics Board at the University of Ottawa [108].

5.8.2 Post-processing

3D positions of left toe, left heel, right toe, and right heel markers were reconstructed using Vicon Nexus software [109]. Gaps in the trial data were filled using cubic spline interpolation and then filtered using a 4th order dual-pass Butterworth lowpass filter with 20Hz cut-off frequency.

(a) **(b)**

Figure 5.10 (a) Front and (b) back views of a participant with reflective markers

Marker-less point cloud data were constructed from the depth images, then 3D locations of toes and heels were tracked. Left toe, left heel, right toe, and right heel were processed independently. Data outliers were statistically filtered, with values two standard deviations or more from the mean removed. Based on time stamp information, trajectory gaps were filled using cubic spline interpolation.

Since the capture time between frames was inconsistent, cubic spline interpolation used to re-sample the data to 60Hz. This re-sampled data was low pass filtered using 4th order dual-pass Butterworth filter with a cut-off frequency of 12Hz. Using cubic spline interpolation, the low pass filtered data was then re-sampled gain to the originally captured time stamps.

5.9 Stride Parameters

This section describes the stride parameters used with both Vicon and marker less systems. The stride parameters were calculated by finding the foot events from the segmented heel and toe points obtained from the foot-tracking algorithm.

Foot events such as foot strike (FS) and foot-off (FO) were identified to calculate the stride parameters. The vertical dimension (z-axis) in the foot tracking data was used to identify FS and FO frames [110].

5.9.1 Foot Events

5.9.1.1 Vicon

Peak vertical values in swing phase were detected for heel (Figure 5.11) and toe (Figure 5.12) markers. These peaks were based on the zero cross over from positive to negative in the vertical velocity, then a peak value greater than 75 percent of maximum vertical value condition was applied.

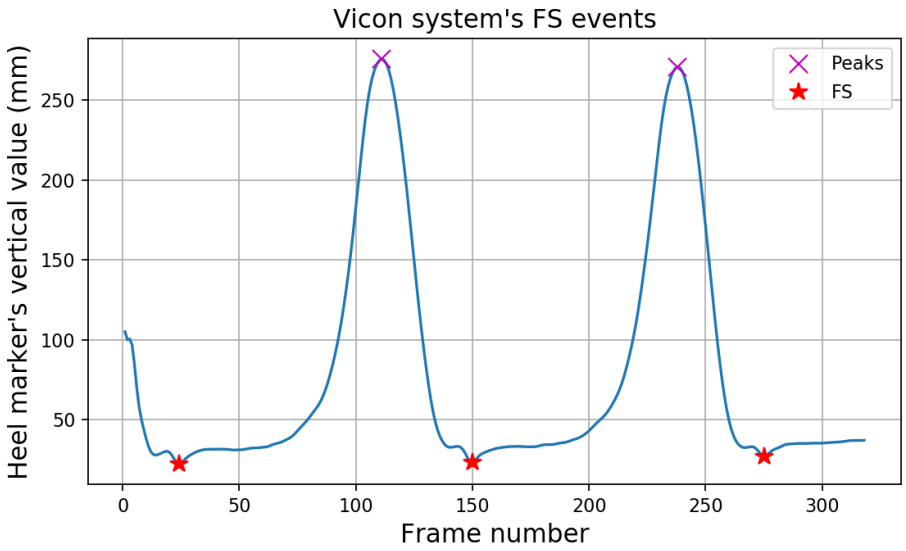


Figure 5.11 FS events from the Vicon heel marker's vertical values

Between two peaks, FS and FO should only occur once. Zero crossovers from negative to positive in the vertical velocity were concave shaped dips in the vertical displacement graph.

These concave dips within the bottom 20 percent of vertical range were identified. The minimum dip between the two peaks in heel data was the FS frame (Figure 5.11). The minimum dip for toe data was the FO frame (Figure 5.12).

Additional conditions were applied to the minimum concave dips before the first peak and after the last peak. The minimum concave dip before the first peak with a distance (in frames) less than 50 percent of frame length between the first two heel peaks was ignored, and greater than 50 percent in the toe data was ignored. Similarly, the number of frames between the last peak and the minimum concave dip after the last peak must be less than 50 percent of frame length between the last two peaks for heel data and greater than 50 percent for toe data.

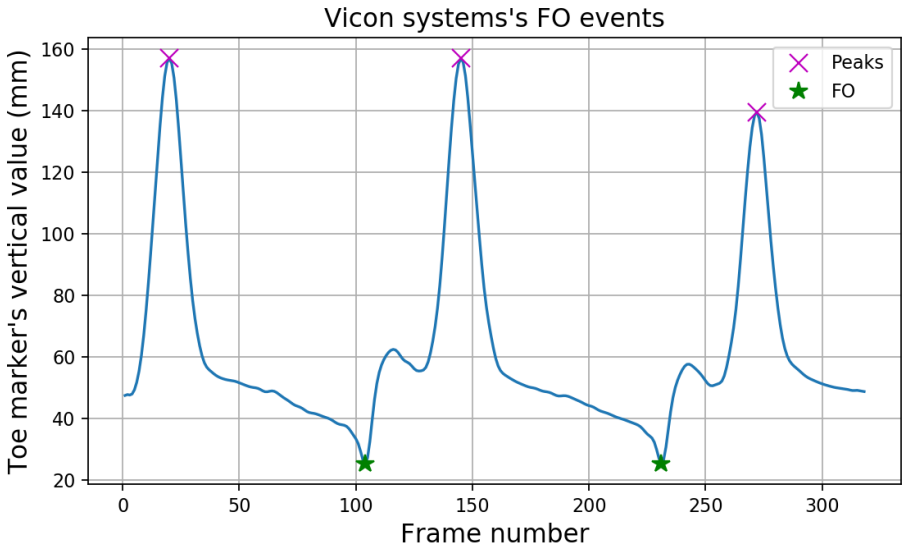


Figure 5.12 FO events from the Vicon toe marker's vertical values

5.9.1.2 Marker-less

Vertical data from the Marker-less system was not as smooth as the Vicon data. Foot event frames were initially estimated using AP (y-axis) data, then finalized based on the vertical data.

The frame where foot reached a stationary state in AP heel position was considered the FS frame (Figure 5.13). This FS frame was adjusted since vertical heel movements may occur after AP movements halt; therefore, the closest concave dip within the next five frames in vertical data was considered FS (Figure 5.14). The initial frame was used if no concave dip occurred.

Similarly, the FO frame was initially defined as the frame where AP toe position changed (Figure 5.15). The previous five frames of vertical toe position was then examined for a concave dip, with FO adjusted to the concave dip frame (Figure 5.16). This method is detailed in Algorithm VI (Table 5-6)

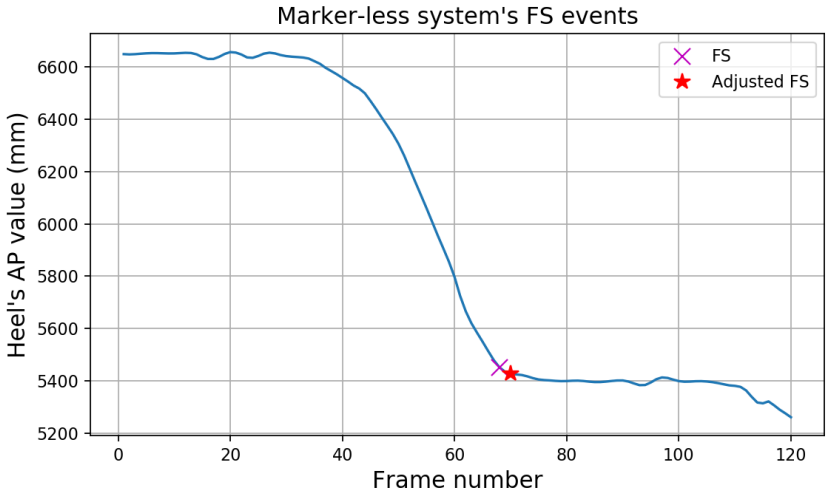


Figure 5.13 FS events in marker-less system's AP heel data

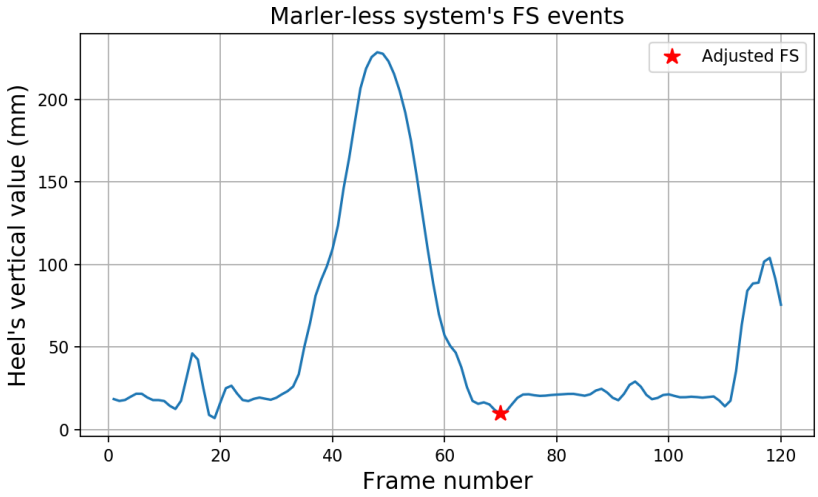


Figure 5.14 FS events in marker-less system's vertical heel data

Algorithm VI identified FS and FO events using AP and vertical data (Table 5-6). Foot movement in the AP position data was used to find FS and FO events, then the FS and FO events were adjusted using the closest concave dips in vertical data.

Table 5-6 Algorithm VI: Foot events identification in the marker-less data

Step	Methodology
1	Create a binary version of arrays for AP heel data (B_{AP}^{heel}) and AP toe data (B_{AP}^{toe}). For AP data (D) with N frames, the i^{th} frame in binary version (B) is calculated as $$B(i) = \begin{cases} 1, \lvert D(i) - D(i+1) \rvert > 20 \; and \; i < N \\ 1, \lvert D(i) - D(i-1) \rvert > 20 \; and \; i = N \\ 0, else \end{cases}$$
2	Median filter binary arrays with filter size = 3 and filter stride = 1.
3	In the sequence of binary heel frames, the transition frame from high to low with at least three high-valued frames and two low-valued frames is the FS frame.
4	The closest zero cross-over frame from negative to positive in the heel's vertical velocity within the next five frames after the initial FS frame was used to adjust the FS frame.
5	In the sequence of binary toe frames, the transition frame from low to high with at least three low-valued frames and two high-valued frames is the FO frame.
6	The closest zero cross over frame from negative to positive in the toe's vertical velocity within the before five frames of initial FO frame is used to adjust the FO frame.

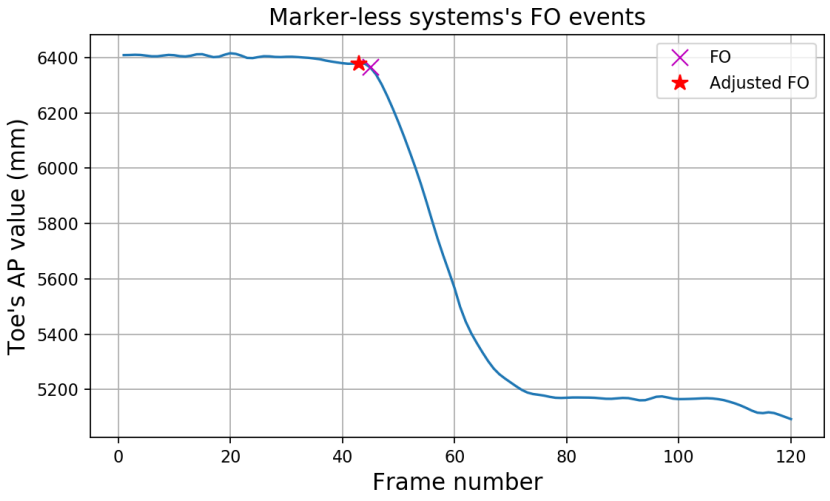


Figure 5.15 FO events in marker-less system's AP toe data

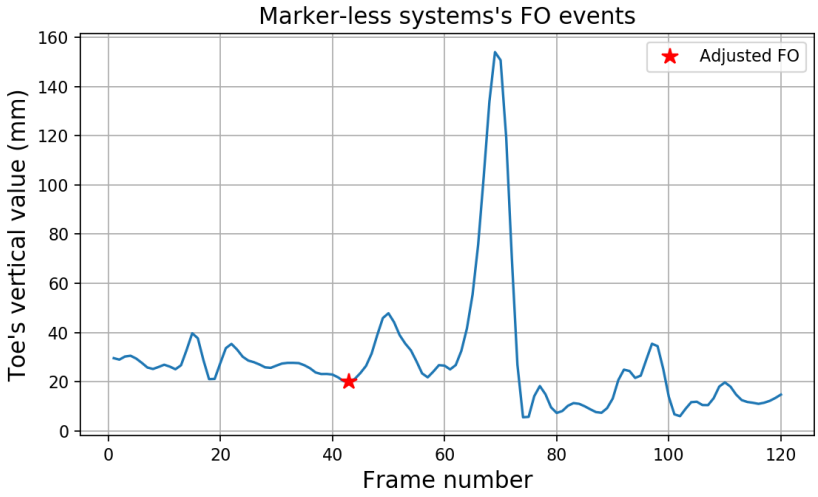


Figure 5.16 FO events in marker-less system's vertical toe data

Left Foot Strike (LFS), Left Foot-Off (LFO), Right Foot Strike (RFS), and Right Foot-off (RFO) were validated according to the normal gait cycle event sequence (i.e., after LFO, the expected next event is LFS and then RFO). If multiple LFS events are identified before RFO, the closest to the RFO event was considered. If no LFS event was identified between LFO and RFO, then the events were ignored.

5.9.2 Parameters

The stride parameters in this research were from one gait cycle (Table 5-7). Primary parameters were directly obtained from the tracking data and the derived parameters were calculated from the primary parameters.

Foot events from Vicon and marker-less systems were synced based on foot position of first common foot event in the marker-less system.

Stride parameters from the both systems were compared and analyzed. For $'n'$ samples, with i^{th} sample represented as x_i, the mean (μ) and standard deviation (σ) were calculated using (5.10) and (5.11), respectively. For a parameter, with value $'v'$ from the Vicon system and value $'m'$ from marker-less system, the sample error (e) was calculated using (5.12).

Table 5-7 Stride parameters in a normal gait cycle

Parameter	Definition
Left stride length	Distance covered by the left heel between two consecutive LFS along the sagittal plane (line of progression, $y - axis$)
Right stride length	Sagittal distance covered by the right heel between two consecutive RFS
Left stride time	Time between two consecutive LFS
Right stride time	Time between two consecutive RFS
Left stride speed [d]	Left stride length / Left stride time
Right stride speed [d]	Right stride length / Right stride time
Left step length	Sagittal distance between right and left heels at LFS
Left step width	Frontal distance ($x - axis$) between right and left heels at LFS
Left step time	Time between RFS and LFS
Right step length	Sagittal distance between left and right heels at RFS
Right step width	Frontal distance between left and right heels at RFS
Right step time	Time between LFS and RFS
Left cadence [d]	Left steps per minute
Right cadence [d]	Right steps per minute
Left stance time	Time between LFS and LFO
Left swing time	Time between LFO and LFS
Left stance to swing ratio [d]	Left stance time / Left swing time
Right stance to swing ratio [d]	Right stance time / Right swing time
Left double support time	Time between LFS and RFO
Right double support time	Time between RFS and LFO
Left foot angle	Angle of left foot midline with sagittal plane when the left foot is entirely in contact with the floor.
Right foot angle	Angle of right foot midline with sagittal when the right foot is entirely in contact with the floor.
Left foot max. velocity	Maximum left foot velocity during the left swing phase (LFO to LFS)
Right foot max. velocity	Maximum right foot velocity during the right swing phase (RFO to RFS)
Left foot clearance	Minimum vertical distance of left foot during mid-swing
Right foot clearance	Minimum vertical distance of right foot during its mid-swing

[d] Derived stride parameters

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{5.10}$$

$$\sigma = \sqrt[2]{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}} \tag{5.11}$$

$$e = v - m \tag{5.12}$$

For each primary stride parameter, μ and σ of error values were calculated. Values farther than two σ from the μ were categorized as outliers and removed from the analysis. Primary stride parameter inliers were used to calculate derived stride parameters.

For a stride parameter with N samples (inliers), i^{th} Vicon's sample $'v_i'$, i^{th}marker-less system's sample $'m_i'$, the mean error (e_μ), error's standard deviation (e_σ), absolute mean error ($e_{|\mu|}$), minimum error (e_{min}), maximum error (e_{max}), Pearson coefficient (r), and the percentage of inliers ($I_{\%}$) were calculated (Table 5-8).

$$e_\mu = \frac{1}{N}\sum_{i=1}^{N} v_i - m_i \tag{5.13}$$

$$e_\sigma = \sqrt[2]{\frac{\sum_{i=1}^{n}(v_i - m_i - e_\mu)^2}{n}} \tag{5.14}$$

$$e_{|\mu|} = \frac{1}{N}\sum_{i=1}^{N} |v_i - m_i| \tag{5.15}$$

$$I_{\%} = \frac{inliers\ count}{inliers\ count + outliers\ count} \times 100 \tag{5.16}$$

5.9.3 Results

From Table 5-8, this novel marker-less system showed minimal mean absolute errors for the majority of stride parameters, compared with the Vicon system. For all parameters, inliers were greater than 90% (rounded-off). Most outliers were due to false detection of foot events because of insufficient capturing volume and noisy data.

The foot tracking algorithm was based on the fixed size OBB and foot bottom plane. This foot orientation definition counteracted AP noise to some extent. Average absolute errors were greater in ML dependent stride parameters such as step width and foot angle. The mean and absolute error values for step length and step time were within the MDC for older people (step length MDC_{95} = 47mm, step time MDC_{95} = 47mm). However, step width was slightly greater than MDC (left step width mean error = 24.15mm, left step width absolute error = 29.86mm, right step width mean error = 27.42mm, right step width absolute mean error = 32.27mm, step width for older people MDC_{95} = 42ms) [71]. The best Pearson correlation coefficient was for stride speed ($r = 0.98$) and lower values were obtained for left foot angle ($r = 0.08$) and left foot clearance ($r = 0.18$).

To the best of our knowledge, this work is the first-time foot clearance has been measured with marker-less depth sensors. The maximum absolute mean error of 1.25cm was observed for right foot clearance. With a marker-less system frame rate at approximately 60fps, all temporal stride parameters were accurate within 30ms absolute mean error. Errors in spatial stride parameters were due to “floor-plane to foot plantar surface” noise generated in the depth images during foot contact phases.

Table 5-8 Stride parameters comparison between the marker-less system and the Vicon system

Parameter (units)		Vicon ($\mu \pm \sigma$)	Marker-less ($\mu \pm \sigma$)	Error ($e_\mu \pm e_\sigma$)	e_μ^{abs}	e_{min}	e_{max}	r	Inliers count	$I_{\%}$
Left	Stride speed (m/s)	1.06 ± 0.11	1.05 ± 0.11	0.012 ± 0.02	0.018	-0.03	0.07	0.98	-	-
	Stride length (m)	1.29 ± 0.09	1.28 ± 0.09	0.012 ± 0.03	0.023	-0.05	0.12	0.95	90	97.8
	Stride time (s)	1.22 ± 0.09	1.23 ± 0.1	-0.002 ± 0.04	0.027	-0.1	0.11	0.93	87	94.5
	Cadence (steps/min)	100.95 ± 8.5	100.79 ± 9.29	0.159 ± 5.35	4.043	-22.22	15.48	0.82	-	-
	Step length (cm)	66.32 ± 5.14	66.07 ± 5.52	0.247 ± 2.45	1.825	-13.46	8.93	0.9	207	92
	Step width (cm)	9.65 ± 4.06	7.24 ± 4.4	2.415 ± 2.64	2.986	-5.28	8.74	0.81	213	94.6
	Step time (s)	0.6 ± 0.05	0.6 ± 0.06	-0.002 ± 0.03	0.024	-0.1	0.1	0.85	221	98.2
	Stance time (s)	0.78 ± 0.07	0.79 ± 0.08	-0.008 ± 0.03	0.024	-0.1	0.07	0.92	185	97.8
	Swing time (s)	0.42 ± 0.04	0.43 ± 0.04	-0.003 ± 0.03	0.026	-0.08	0.08	0.6	173	96.6
	Stance to swing ratio (NA)	1.86 ± 0.14	1.9 ± 0.22	-0.031 ± 0.21	0.172	-0.67	0.38	0.41	-	-
	Double support time (s)	0.17 ± 0.03	0.18 ± 0.04	-0.004 ± 0.03	0.027	-0.11	0.1	0.53	248	98.4
	Foot angle (°)	8.34 ± 5.14	1.12 ± 0.63	7.209 ± 5.13	7.256	-1.22	19.14	0.08	176	93.6
	Foot max. velocity(m/s)	3.71 ± 0.31	3.88 ± 0.37	-0.163 ± 0.28	0.247	-1.01	0.45	0.66	171	95.5
	Foot clearance (cm)	2.6 ± 0.62	1.99 ± 1.36	0.605 ± 1.39	1.224	-4.19	3.07	0.18	113	93.4
Right	Stride speed (m/s)	1.1 ± 0.12	1.09 ± 0.12	0.006 ± 0.02	0.017	0.06	-0.04	0.98	-	-
	Stride length (m)	1.29 ± 0.1	1.29 ± 0.1	0.001 ± 0.02	0.017	0.06	-0.04	0.97	79	96.3
	Stride time (s)	1.19 ± 0.09	1.19 ± 0.09	-0.007 ± 0.03	0.025	0.08	-0.07	0.95	77	93.9
	Cadence (steps/min)	99.44 ± 7.55	99.28 ± 7.95	0.157 ± 4.68	3.678	12.7	-12.72	0.82	-	-
	Step length (cm)	64.8 ± 4.96	64.42 ± 5.41	0.3767 ± 2.19	1.722	8.25	-8.82	0.91	177	89.8
	Step width (cm)	9.54 ± 3.76	6.8 ± 4.3	2.742 ± 2.75	3.227	8.86	-5.19	0.77	185	93.9
	Step time (s)	0.61 ± 0.05	0.61 ± 0.05	-0.001 ± 0.03	0.023	0.1	-0.08	0.82	193	97.9
	Stance time. (s)	0.77 ± 0.07	0.78 ± 0.08	-0.009 ± 0.03	0.025	0.08	-0.1	0.92	206	98.5
	Swing time. (s)	0.43 ± 0.03	0.42 ± 0.04	0.003 ± 0.03	0.023	0.08	-0.06	0.69	141	94.6
	Stance to swing time (NA)	1.82 ± 0.13	1.85 ± 0.18	-0.027 ± 0.15	0.127	0.38	-0.36	0.56	-	-
	Double support time (s)	0.17 ± 0.03	0.17 ± 0.05	-0.002 ± 0.04	0.027	0.1	-0.16	0.65	245	99.6
	Foot angle (°)	9.05 ± 5.71	1.14 ± 0.58	7.915 ± 5.62	7.935	21.21	-0.92	0.2	201	96.2
	Foot max. velocity (m/s)	3.67 ± 0.28	3.83 ± 0.36	-0.155 ± 0.25	0.204	0.32	-1.07	0.73	139	93.3
	Foot clearance (cm)	2.68 ± 0.62	2.1 ± 1.44	0.572 ± 1.35	1.251	3.26	-3.35	0.36	102	96.2

5.10 Conclusion

In this research, we proposed a smart hallway using depth sensors for foot tracking and stride parameter analysis. With six temporally and spatially synchronized Intel RealSense D415 depth sensors, depth data were successfully background-subtracted and merged to form a walking human's point cloud time series. The point cloud was then effectively segmented into left and right foot point clouds. A bounding box was fitted around the foot in each leg's point cloud data. The bounding box around the foot in each frame enabled foot tracking, and stride parameter calculation. Stride parameters obtained from this newly developed marker-less system were comparable with gold standard Vicon system output.

The marker-less system had promising results with accurate temporal stride parameters and minimal errors in spatial stride parameters. Step width accuracy needs to improve and poor foot angle accuracy was observed due to noise around the foot as it approaches the floor plane. Since foot clearance error was greater than 1cm, and foot clearance varying between 2 and 3.2 cm, this error would need to be reduced to provide usable results for clinical decision-making.

Unlike the machine learning based skeleton tracking systems, our proposed system's foot tracking never moves outside the foot and also captures data at approximately 60fps. This system could monitor a large number of people for long hours with no preparation time (no sensors attached to the body), without any discomfort, and without expert intervention.

6 Thesis Conclusions and Future Work

This thesis work aimed to design and evaluate innovative marker-less foot tracking system with six Intel RealSense D415 depth sensors for use in institutional hallways. The hallway setup was based on a novel 3D animation technique developed to quantify the percentage of visibility of the human foot during gait. An important contribution was developing a novel foot tracking algorithm by fitting a bounding box to the foot's point cloud, and then validating it by comparing the walking stride parameters from a cohort of able-bodied participants with results from a gold standard Vicon system. This foot tracking system captured data close to 60fps. Spatial stride parameter errors were due to medial-lateral noise generated from the depth sensors around the foot. The following objectives and results were defined and obtained in this thesis:

Objective1: Find an appropriate depth sensor for motion capturing in an institutional hallway.

After literature and market surveys, three depth-sensing technologies were evaluated: Microsoft Kinect V2 (Time-of-flight), Orbbec Astra Pro (Structured Light), and Intel RealSense D415 (Stereoscopic). The Intel RealSense D415 depth sensor was suitable for this Smart Hallway application since it showed no interference when multiple sensors were used together and also captured high-resolution depth images of 848 × 480 at 60fps.

Objective2: Design the physical setup with multiple depth sensors in a hallway scenario and synchronize all the depth sensors.

A 3D animation of human walking through an institutional hallway was designed, and different setups with virtual depth sensors were simulated. At each instance, four sensors from the diagonal corners of a person captured the foot surface. Based on the simulation results, a hallway scenario with six sensors setup was created. The six sensors were frame by frame time-synchronized using a server-client approach, and a chessboard based calibration was used for spatial synchronization.

Objective3: Develop an algorithm to track human foot in the 3D point cloud data.

A novel foot tracking algorithm was developed which combined 3D point cloud data obtained from six Intel RealSense D415 sensors. The left and right legs were segmented in the point cloud data, and then for each leg, foot tracking was achieved by fitting a box to the foot, with the base

of box on the bottom plane of the foot.

Objective4: Validate the foot tracking algorithm by calculating and comparing the stride parameters with a gold standard Vicon system.

The novel foot tracking algorithm developed in this thesis was applied to the walking data collected from the twenty-two volunteers. Stride parameters were compared between the proposed marker-less system and marker-based Vicon system. The mean differences between systems were less than 10ms for temporal stride parameters. Errors were observed in step width and foot angle due to noise in the medial-lateral direction.

6.1 Future Work

6.1.1 Robustness

The current foot tracking algorithm was tested with a regular walking style, with participants walking naturally at a comfortable speed (mean walking speed of 1.08 m/s). Different walking speeds ranging from slow walking (0.5 m/s) to fast walking (1.6 m/s) could be evaluated [72]. The robustness of this novel foot tracking algorithm could be validated by testing on different walking styles such as fore-foot, flat foot, toe inwards, and toe outwards walking, representing a range of pathological walking styles.

6.1.2 Full Body Gait Analysis

The current physical setup was designed to capture a person's lower leg because of the trade-off between depth sensors range and accuracy. The capture volume could be extended to capture the whole body by using more depth sensors, enabling full-body gait analysis.

6.1.3 GPU and Parallel Computing

The source code for time-synchronized data capturing (depth and color images) was multi-threaded, and construction of combined point clouds from the six camera depth images at each instance was implemented in a serial fashion using a CPU. Construction and processing of point clouds with thousands of 3D points required high computing time. Parallel computing and GPUs with capability to process massive data sets will help achieve near real-time full body analysis.

6.1.4 Hardware

Time synchronization was achieved by using a server-client software approach over Ethernet. Recently, Intel released hardware sync for the RealSense D415 [111]. Hardware synchronization of multiple sensors with physical cables saves computation time and computation power. One board computer with Power over Ethernet (POE) could process data from at least two depth sensors and post-process the images locally, reducing overall system computation time and complexity.

6.1.5 Calibration Tool

The spatial calibration could be made efficient by auto-calibrating all the sensors based on the consistent features in a hallway such as wall planes, floor plane, edges, etc. The calibration method could be made user-friendly by providing feedback to the user about the quality of the calibration, or indication when the calibration is off and recalibration is needed.

6.1.6 Advanced Point Cloud Noise Filtering

In the current foot tracking algorithm, errors were due to noise around the foot, especially during foot strike and foot off phases. Point cloud filters such as statistical outlier removal, moving least squares smoothing, and voxel grid filters were used. However, these traditional filters were not able to fully remove the randomly generated noise. Advanced noise filtering algorithms based on human body shape estimation or machine learning based filters could achieve clean point cloud data and more accurate results.

6.1.7 Textured Point Clouds

Color textured point cloud data has the possibility for integrating many features with gait analysis. A Smart Hallway can be built with facial recognition for patient identification, full-body gait analysis for monitoring mobility changes, fall risk estimation, and monitoring progression of nervous system disorders that affects mobility. From facial analysis, respiratory rate, heart rate, emotion detection, smoking effects, and fat percentage could be monitored. With the marker-less Smart Hallway approach, patients or long-term-care residents could be diagnosed regularly without any discomfort and also without external supervision. This approach could potentially be revolutionary in future long-term care, health care, and rehabilitation facilities.

References

[1] M. Pistacchi *et al.*, “Gait Analysis and Clinical Correlations in Early Parkinson’s Disease,” *Funct. Neurol.*, vol. 32, no. 1, pp. 28–34, 2017.

[2] L. M. Waite, D. A. Grayson, O. Piguet, H. Creasey, H. P. Bennett, and G. A. Broe, “Gait Slowing as a Predictor of Incident Dementia: 6-Year Longitudinal Data from the Sydney Older Persons Study,” *J. Neurol. Sci.*, vol. 229–230, pp. 89–93, 2005.

[3] D. Kindregan, L. Gallagher, and J. Gormley, “Gait Deviations in Children with Autism Spectrum Disorders: A Review,” *Autism Res. Treat.*, vol. 2015, pp. 1–8, 2015.

[4] M. G. Benedetti, R. Piperno, L. Simoncini, P. Bonato, A. Tonini, and S. Giannini, “Gait Abnormalities in Minimally Impaired Multiple Sclerosis Patients,” *Mult. Scler. J.*, vol. 5, no. 5, pp. 363–368, 1999.

[5] R. Baker, “Gait Analysis Methods in Rehabilitation,” *J. Neuroeng. Rehabil.*, vol. 3, pp. 1–10, 2006.

[6] R. S. Goncalves, T. Hamilton, and H. I. Krebs, “MIT-Skywalker: On the Use of a Markerless System,” *IEEE Int. Conf. Rehabil. Robot.*, vol. 2017, pp. 205–210, Jul. 2017.

[7] V. Gutta, E. D. Lemaire, N. Baddour, and P. Fallavollita, “A Comparison of Depth Sensors for 3D Object Surface Reconstruction,” in *CMBES, Vol.42*, 2019.

[8] V. Gutta, N. Baddour, P. Fallavollita, and E. D. Lemaire, “Multiple Depth Sensor Setup and Synchronization for Marker-less 3D Human Foot Tracking in a Hallway,” in *IEEE International Workshop on Software Engineering for Healthcare*, 2019, pp. 77–80.

[9] A. Muro-de-la-Herran, B. García-Zapirain, and A. Méndez-Zorrilla, “Gait Analysis Methods: An Overview of Wearable and Non-wearable Systems, Highlighting Clinical Applications,” *Sensors (Switzerland)*, vol. 14, no. 2, pp. 3362–3394, 2014.

[10] J. A. Cohen *et al.*, “Development of a Multiple Sclerosis Functional Composite as a Clinical Trial Outcome Measure,” *Brian*, vol. 122, pp. 871–882, 1999.

[11] J. C. Hobart, A. Riazi, D. L. Lamping, R. Fitzpatrick, and A. J. Thompson, “Measuring the Impact of MS on Walking Ability: The 12-item MS Walking Scale (MSWS-12),” *Neurology*, vol. 60, no. 1, pp. 31–36, 2003.

[12] M. E. Tinetti, “Performance-Oriented Assessment of Mobility Problems in Elderly Patients,” *J. Am. Geriatr. Soc.*, vol. 34, no. 2, pp. 119–126, 1986.

[13] B. Mathias, S.; Nayak, U.S.; Isaacs, “Balance in elderly patients: the ‘get-up and go’ test.,” *Arch. Phys. Med. Rehabil.*, vol. 67, no. 6, pp. 387–389, 1986.

[14] L. Wolfson, R. Whipple, P. Amerman, and J. N. Tobin, “Gait Assessment in the Elderly: A Gait Abnormality Rating Scale and its Relation to Falls,” *Journals Gerontol.*, vol. 45, no. 1, pp. 12–19, 1990.

[15] A. V. Fried, J. Cwikel, H. Ring, and D. Galinsky, “ELGAM - Extra-Laboratory Gait Assessment Method: Identification of Risk Factors for Falls among the Elderly at Home,” *Disabil. Rehabil.*, vol. 12, no. 4, pp. 161–164, 1990.

[16] M. E. Taylor, K. Delbaere, A. S. Mikolaizak, S. R. Lord, and J. C. T. Close, “Gait Parameter Risk Factors for Falls under Simple and Dual Task Conditions in Cognitively Impaired Older People,” *Gait Posture*, vol. 37, no. 1, pp. 126–130, 2013.

[17] A. Schinkel-Ivy, J. S. Wong, and A. Mansfield, “Balance Confidence Is Related to Features of Balance and Gait in Individuals with Chronic Stroke,” *J. Stroke Cerebrovasc. Dis.*, vol. 26, no. 2, pp. 237–245, 2017.

[18] O. Beauchet, “Gait Variability at Fast-pace Walking Speed : A Biomarker Of Mild Cognitive Impairment ?,” vol. 17, no. 3, pp. 3–4, 2013.

[19] “Gait cycle.” [Online]. Available: https://clinicalgate.com/assessment-of-gait/. [Accessed: 25-Feb-2020].

[20] “Vicon.” [Online]. Available: www.vicon.com/motion-capture/life-sciences. [Accessed: 22-Apr-2019].

[21] “OptiTrack.” [Online]. Available: www.optitrack.com/motion-capture-movement-sciences/. [Accessed: 26-Apr-2019].

[22] T. Cloete and C. Scheffer, “Benchmarking of a Full-body Inertial Motion Capture System for Clinical Gait Analysis,” in *30th Annual International IEEE EMBS Conference*, 2008, pp. 4579–4582.

[23] X. Robert-Lachaine, H. Mecheri, C. Larue, and A. Plamondon, “Validation of Inertial Measurement Units with an Optoelectronic System for Whole-body Motion Analysis,” *Med. Biol. Eng. Comput.*, vol. 55, no. 4, pp. 609–619, 2017.

[24] K. Kobayashi, L. Gransberg, E. Knutsson, and P. Nolen, “A New System for Three-dimensional Gait Recording using Electromagnetic Tracking,” *Gait Posture*, pp. 63–75, 1997.

[25] “Gypsy.” [Online]. Available: www.metamotion.com/gypsy/gypsy-motion-capture-system-workflow.htm. [Accessed: 22-Apr-2019].

[26] Y. Qi, C. Boon Soh, E. Gunawan, K. S. Low, and R. Thomas, “Lower Extremity Joint Angle Tracking with Wireless Ultrasonic Sensors during a Squat Exercise,” *Sensors (Switzerland)*, vol. 15, no. 5, pp. 9610–9627, 2015.

[27] M. R. Patterson *et al.*, “Does External Walking Environment affect Gait Patterns?,” *2014 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC 2014*, pp. 2981–2984, 2014.

[28] S. J. Preece, L. P. J. Kenney, M. J. Major, T. Dias, E. Lay, and B. T. Fernandes, “Automatic Identification of Gait Events using an Instrumented Sock,” *J. Neuroeng. Rehabil.*, vol. 8, no. 32, pp. 1–10, 2011.

[29] “GAITRite.” [Online]. Available: www.gaitrite.com/. [Accessed: 25-Apr-2019].

[30] “Noraxon.” [Online]. Available: www.noraxon.com/our-products/medilogic-insoles/. [Accessed: 25-Apr-2019].

[31] “Gait Lab - Case Study.” [Online]. Available: www.vicon.com/case-studies/life-sciences/the-gait-lab-of-the-future. [Accessed: 25-Apr-2019].

[32] P. Zanuttigh, C. D. Mutto, L. Minto, G. Marin, F. Dominio, and G. M. Cortelazzo, *Time-of-flight and Structured Light Depth Cameras: Technology and Applications*. 2016.

[33] “Kinect for Xbox 360.” [Online]. Available: www.wikipedia.org/wiki/Kinect#Kinect_for_Xbox_360_(2010). [Accessed: 21-May-2019].

[34] “Kinect Xbox 360 vs Xbox One.” [Online]. Available: www.zugara.com/how-does-the-kinect-2-compare-to-the-kinect-1. [Accessed: 21-May-2019].

[35] “Asus Xtion Pro Live.” [Online]. Available: www.xtionprolive.com/asus-xtion-pro-live. [Accessed: 21-May-2019].

[36] “Kinect for Xbox One.” [Online]. Available: www.wikipedia.org/wiki/Xbox_One. [Accessed: 21-May-2019].

[37] “Intel RealSense Wiki.” [Online]. Available: www.wikipedia.org/wiki/Intel_RealSense. [Accessed: 21-May-2019].

[38] “Intel RealSense F200.” [Online]. Available: https://ark.intel.com/content/www/us/en/ark/products/92255/intel-realsense-camera-f200.html. [Accessed: 21-May-2019].

[39] “Intel RealSense R200.” [Online]. Available: https://ark.intel.com/content/www/us/en/ark/products/92256/intel-realsense-camera-r200.html. [Accessed: 29-Jun-2019].

[40] “ZED.” [Online]. Available: www.stereolabs.com/zed/. [Accessed: 29-Jun-2019].

[41] “Orbbec Astra Pro.” [Online]. Available: www.orbbec3d.com/product-astra-pro/. [Accessed: 02-Jul-2019].

[42] “Orbbec Presee.” [Online]. Available: www.orbbec3d.com/product-persee/. [Accessed: 02-Jul-2019].

[43] “Intel RealSense SR300.” [Online]. Available: https://ark.intel.com/content/www/us/en/ark/products/92329/intel-realsense-camera-sr300.html. [Accessed: 01-Jul-2019].

[44] “Intel RealSense Camera D415.” [Online]. Available: www.intelrealsense.com/depth-camera-d415/. [Accessed: 02-Jul-2019].

[45] “Duo3D.” [Online]. Available: https://duo3d.com/product/duo-minilx-lv1. [Accessed: 21-May-2019].

[46] “e-con systems.” [Online]. Available: www.e-consystems.com/3D-USB-stereo-camera.asp. [Accessed: 02-Jul-2019].

[47] “Occipital.” [Online]. Available: www.structure.io. [Accessed: 02-Jul-2019].

[48] “Azure Kinect DK.” [Online]. Available: www.azure.microsoft.com/en-ca/services/kinect-dk/. [Accessed: 02-Jul-2019].

[49] “Intel RealSense D435i.” [Online]. Available: www.intelrealsense.com/depth-camera-d435i/. [Accessed: 02-Jul-2019].

[50] “Orbbec.” [Online]. Available: www.orbbec3d.com. [Accessed: 02-Jul-2019].

[51] S. K. Natarajan, X. Wang, M. Spranger, and A. Gräser, “Reha@Home - A Vision Based Markerless GAIT Analysis System for Rehabilitation at Home,” *Biomed. Eng. (NY).*, no. BioMed, pp. 32–41, 2017.

[52] K. E. Y. Words, “A Study on the Correlation between the Foot Features and Gait,” pp. 210–213, 2017.

[53] A. D. C and A. Coroiu, “Interchangeability of Kinect and Orbbec Sensors for Gesture Recognition,” pp. 309–315, 2018.

[54] F. Uccheddu and L. Governi, “A Preliminary 3D Depth Camera-Based System to Assist Home Physiotherapy Rehabilitation,” in *Design Tools and Methods in Industrial Engineering*, 2019, pp. 696–706.

[55] F. L. Siena, B. Byrom, P. Watts, P. Breedon, and P. Watts, “Utilising the Intel RealSense Camera for Measuring Health Outcomes in Clinical Research,” *J. Med. Syst.*, vol. 42, no. 53, pp. 1–10, 2018.

[56] A. Baldominos, Y. Saez, and C. García, “An Approach to Physical Rehabilitation using State-of-the-art Virtual Reality and Motion Tracking Technologies,” *Procedia - Procedia Comput. Sci.*, vol. 64, pp. 10–16, 2015.

[57] Y. Kailun, W. Kaiwei, C. Hao, and B. Jian, “Reducing the Minimum Range of a RGB-depth Sensor to Aid Navigation in Visually Impaired Individuals,” *Appl. Opt.*, vol. 57, no. 11, 2018.

[58] K. Yang, K. Wang, W. Hu, and J. Bai, “Expanding the Detection of Traversable Area with RealSense for the Visually Impaired based in Stereo,” *Sensors (Basel).*, vol. 16, no. 11, pp. 1–20, 2016.

[59] N. L. Dao, Y. Zhang, J. Zheng, and J. Cai, “Kinect-based Non-intrusive Human Gait Analysis and Visualization,” *2015 IEEE 17th Int. Work. Multimed. Signal Process. MMSP 2015*, pp. 1–6, 2015.

[60] A. Procházka, O. Vyšata, M. Vališ, O. Ťupa, M. Schätz, and V. Mařík, “Use of the Image and Depth sensors of the Microsoft Kinect for the Detection of Gait Disorders,” *Neural Comput. Appl.*, vol. 26, no. 7, pp. 1621–1629, 2015.

[61] R. Taban, A. Parsa, and H. Moradi, “Tip-toe Walking Detection using CPG Parameters from Skeleton Data Gathered by Kinect,” in *International Conference on Ubiquitous Computing and Ambient Intelligence*, 2017, vol. 10586 LNCS, pp. 287–298.

[62] J. Zhao, F. E. Bunn, J. M. Perron, E. Shen, and R. S. Allison, “Gait Assessment using the Kinect RGB-D Sensor,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2015-Novem, pp. 6679–6683, 2015.

[63] P. Newland, J. M. Wagner, A. Salter, F. P. Thomas, M. Skubic, and M. Rantz, “Exploring the Feasibility and Acceptability of Sensor Monitoring of Gait and Falls in the Homes of Persons with Multiple Sclerosis,” *Gait Posture*, vol. 49, pp. 277–282, 2016.

[64] J. Latorre, R. Llorens, A. Borrego, M. Alcaniz, C. Colomer, and E. Noe, “A Low-cost Kinect (TM) for Windows (R) v2-based Gait Analysis System - A Feasibility Study with Healthy Subjects,” in *2017 International Conference on Virtual Rehabilitation (Icvr)*, 2017, pp. 2–3.

[65] D. Geerse, B. Coolen, D. Kolijn, and M. Roerdink, “Validation of Foot Placement Locations from Ankle Data of a Kinect v2 Sensor,” *Sensors (Switzerland)*, vol. 17, no. 10, pp. 6–9, 2017.

[66] J. W. Qiu *et al.*, “Continuous Human Location and Posture Tracking by Multiple Depth sensors,” *Proc. - 2014 IEEE Int. Conf. Internet Things, iThings 2014, 2014 IEEE Int. Conf. Green Comput. Commun. GreenCom 2014 2014 IEEE Int. Conf. Cyber-Physical-Social Comput. CPS 20*, no. iThings, pp. 155–160, 2014.

[67] R. A. Clark, K. J. Bower, B. F. Mentiplay, K. Paterson, and Y.-H. Pua, “Concurrent Validity of the Microsoft Kinect for Assessment of Spatiotemporal Gait Variables.,” *J. Biomech.*, vol. 46, no. 15, pp. 2722–2725, Oct. 2013.

[68] M. Ghobadi and E. T. Esfahani, “A Robust Method to Improve Estimation Accuracy of Walking Gait Kinematics by Considering Geometrical Constraints on Kinect Data,” *Proc. ASME Des. Eng. Tech. Conf.*, vol. 3, pp. 1–6, 2016.

[69] A. Dubois and F. Charpillet, “A Gait Analysis Method Based on a Depth Camera for Fall Prevention,” *2014 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC 2014*, pp. 4515–4518, 2014.

[70] E. E. Stone and M. Skubic, “Passive In-home Measurement of Stride-to-Stride Gait Variability Comparing Vision and Kinect Sensing,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 6491–6494, 2011.

[71] L. Wroblewska *et al.*, “The Test–Retest Reliability and Minimal Detectable Change of Spatial and Temporal Gait Variability during Usual Over-Ground Walking for Younger and Older Adults,” *Gait Posture*, vol. 33, no. 8, pp. 839–841, 2016.

[72] R. W. Bohannon and A. Williams Andrews, “Normal Walking Speed: A Descriptive Meta-analysis,” *Physiotherapy*, vol. 97, no. 3, pp. 182–189, 2011.

[73] D. J. Geerse, B. H. Coolen, and M. Roerdink, “Kinematic Validation of a Multi-Kinect v2 Instrumented 10-meter Walkway for Quantitative Gait Assessments,” *PLoS One*, vol. 10, no. 10, pp. 1–15, 2015.

[74] Y. Nizam, M. N. Mohd, and M. Mahadi Abdul Jamil, “Biomechanical Application: Exploitation of Kinect Sensor for Gait Analysis,” *ARPN J. Eng. Appl. Sci.*, vol. 12, no. 10, pp. 3183–3188, 2017.

[75] R. Z. L. Hu, A. Hartfiel, J. Tung, A. Fakih, J. Hoey, and P. Poupart, “3D Pose Tracking of Walker Users’ Lower Limb with a Structured-light Camera on a Moving Platform,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 29–36, 2011.

[76] G. Baldewijns, G. Verheyden, B. Vanrumste, and T. Croonenborghs, “Validation of the Kinect for Gait Analysis using the GAITRite Walkway,” *2014 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC 2014*, pp. 5920–5923, 2014.

[77] H. A. Nguyen and J. Meunier, “Gait Analysis from Video: Camcorders vs. Kinect,” in *11th International Conference on Image Analysis and Recognition*, 2014, pp. 66–72.

[78] G. Paolini *et al.*, “Validation of a Method for Real time Foot Position and Orientation Tracking with Microsoft Kinect Technology for use in Virtual Reality and Treadmill based Gait Training Programs,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 5, pp. 997–1002, 2014.

[79] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster, “Full Body Gait Analysis with Kinect,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 1964–1967, 2012.

[80] M. R. Kharazi *et al.*, “Validity of Microsoft KinectTM for Measuring Gait Parameters,” *2015 22nd Iran. Conf. Biomed. Eng. ICBME 2015*, no. November, pp. 375–379, 2016.

[81] B. Müller, W. Ilg, M. A. Giese, and N. Ludolph, “Validation of Enhanced Kinect Sensor based Motion Capturing for Gait Assessment,” *PLoS One*, vol. 12, no. 4, pp. 14–16, 2017.

[82] W. Zijlstra and A. L. Hof, “Assessment of Spatio-temporal Gait Parameters from Trunk Accelerations during Human Walking,” *Gait Posture*, vol. 18, no. 2, pp. 1–10, 2003.

[83] F. Hebenstreit, A. Leibold, S. Krinner, G. Welsch, M. Lochmann, and B. M. Eskofier, “Effect of Walking Speed on Gait Sub Phase Durations,” *Hum. Mov. Sci.*, vol. 43, pp. 118–124, 2015.

[84] L. Yang, B. Yang, H. Dong, and A. El Saddik, “3-D Markerless Tracking of Human Gait by Geometric Trilateration of Multiple Kinects,” *IEEE Syst. J.*, pp. 1–11, 2016.

[85] H. Sarbolandi, D. Lefloch, and A. Kolb, “Kinect Range Sensing: Structured-light versus Time-of-Flight Kinect,” *Comput. Vis. Image Underst.*, vol. 139, pp. 1–20, 2015.

[86] “Orbbec Astra Pro.” [Online]. Available: www.orbbec3d.com/products/. [Accessed: 16-Jan-2019].

[87] “Intel RealSense D400 Series.” [Online]. Available: https://realsense.intel.com/stereo-depth-vision-basics. [Accessed: 16-Jan-2019].

[88] “RANSAC Sphere - PCL.” [Online]. Available: http://docs.pointclouds.org/1.8.1/group__sample__consensus.html. [Accessed: 30-Jan-2019].

[89] E. Stone and M. Skubic, “Evaluation of an Inexpensive Depth Camera for Passive In-Home Fall Risk Assessment,” *Proc. 5th Int. ICST Conf. Pervasive Comput. Technol. Healthc.*, pp. 71–77, 2011.

[90] J. Colley, H. Zeeman, and E. Kendall, “‘Everything Happens in the Hallways’: Exploring User Activity in the Corridors at Two Rehabilitation Units,” *Heal. Environ. Res. Des. J.*, vol. 11, no. 2, pp. 163–176, 2018.

[91] “Blender Software,” *2.79b*, 2017. [Online]. Available: www.blender.org/. [Accessed: 30-Sep-2019].

[92] “Intel RealSense D4xx Tuning.” [Online]. Available: https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/BKMs_Tuning_RealSense_D4xx_Cam.pdf. [Accessed: 30-Jan-2019].

[93] “Intel RealSense D4xx Post processing.” [Online]. Available: https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-Depth-PostProcess.pdf. [Accessed: 30-Jan-2019].

[94] “Stereo Calibration OpenCV.” [Online]. Available: https://docs.opencv.org/3.4/d9/d0c/group__calib3d.html#ga91018d80e2a93ade37539f01e6f07de5. [Accessed: 30-Jan-2019].

[95] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz, “Towards 3D Point Cloud Based Object Maps for Household Environments,” *Rob. Auton. Syst.*, vol. 56, no. 11, pp. 927–941, 2008.

[96] M. Alexa, J. Behr, D. Cohen-Or, S. Fleishman, D. Levin, and C. T. Silva, “Computing and Rendering Point Set Surfaces,” *IEEE Trans. Vis. Comput. Graph.*, vol. 9, no. 1, pp. 3–15, 2003.

[97] “Voxel Grid Filter.” [Online]. Available: http://docs.pointclouds.org/1.8.1/classpcl_1_1_voxel_grid.html. [Accessed: 30-Jan-2019].

[98] J. P. Ferreira, M. M. Crisostomo, and A. P. Coimbra, “Human Gait Acquisition and Characterization,” *IEEE Trans. Instrum. Meas.*, vol. 58, no. 9, pp. 2979–2988, 2009.

[99] G. Panahandeh, N. Mohammadiha, A. Leijon, and P. Handel, “Continuous Hidden Markov Model for Pedestrian Activity Classification and Gait Analysis,” *IEEE Trans. Instrum. Meas.*, vol. 62, no. 5, pp. 1073–1083, 2013.

[100] Ö. Bebek *et al.*, “Personal Navigation via High-Resolution Gait-Corrected Inertial Measurement Units,” *IEEE Trans. Instrum. Meas.*, vol. 59, no. 11, pp. 3018–3027, 2010.

[101] Z. Q. Zhang and X. Meng, “Use of an Inertial/Magnetic Sensor Module for Pedestrian Tracking during Normal Walking,” *IEEE Trans. Instrum. Meas.*, vol. 64, no. 3, pp. 766–783, 2015.

[102] C. Grana, D. Borghesani, and R. Cucchiara, “Decision Trees for Fast Thinning Algorithms,” *Proc. - Int. Conf. Pattern Recognit.*, vol. 19, no. 6, pp. 2836–2839, 2010.

[103] “OpenCV.” [Online]. Available: www.opencv.org. [Accessed: 02-Oct-2019].

[104] “PCL.” [Online]. Available: www.pointclouds.org. [Accessed: 02-Oct-2019].

[105] R. B. Rusu, “Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments,” Techniche Universitat Munchen, 2009.

[106] M. A. Fischler and R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[107] “Vicon.” [Online]. Available: www.vicon.com. [Accessed: 17-Oct-2019].

[108] “Research Ethics Board.” [Online]. Available: www.research.uottawa.ca/ethics/. [Accessed: 17-Oct-2019].

[109] “Nexus.” [Online]. Available: https://www.vicon.com/software/nexus/. [Accessed: 17-Oct-2019].

[110] J. Sinclair, C. Edmundson, D. Brooks, and S. Hobbs, “Evaluation of Kinematic Methods of Identifying Gait Events during Running,” *Int. J. Sport. …*, vol. 05, no. 03, pp. 188–192, 2011.

[111] A. Grunnet-jepsen *et al.*, “Using the RealSense D4xx Depth Sensors in Multi-Camera Configurations,” Santa Monica, CA, USA, 2018.

Appendix

29/10/2018

Université d'Ottawa
Bureau d'éthique et d'intégrité de la recherche

University of Ottawa
Office of Research Ethics and Integrity

CERTIFICAT D'APPROBATION ÉTHIQUE | CERTIFICATE OF ETHICS APPROVAL

Numéro du dossier / Ethics File Number	H-08-18-860
Titre du projet / Project Title	Human stride analysis using marker-less 3D depth sensors
Type de projet / Project Type	Thèse de maîtrise / Master's thesis
Statut du projet / Project Status	Approuvé / Approved
Date d'approbation (jj/mm/aaaa) / Approval Date (dd/mm/yyyy)	29/10/2018
Date d'expiration (jj/mm/aaaa) / Expiry Date (dd/mm/yyyy)	28/10/2019

Équipe de recherche / Research Team

Chercheur / Researcher	Affiliation	Role
Vinod GUTTA	École de science informatique et de génie électrique / School of Electrical Engineering and Computer Science	Chercheur Principal / Principal Investigator
Pascal FALLAVOLLITA	École interdisciplinaire des sciences de la santé / Interdisciplinary School of Health Sciences	Superviseur / Supervisor
Natalie BADDOUR	Département de génie mécanique / Department of Mechanical Engineering	Co-superviseur / Co-supervisor
Edward LEMAIRE		Co-superviseur / Co-supervisor

Conditions spéciales ou commentaires / Special conditions or comments

550, rue Cumberland, pièce 154
Ottawa (Ontario) K1N 6N5 Canada

550 Cumberland Street, Room 154
Ottawa, Ontario K1N 6N5 Canada

613-562-5387 • 613-562-5338 • ethique@uOttawa.ca / ethics@uOttawa.ca
www.recherche.uottawa.ca/deontologie | www.recherche.uottawa.ca/ethics

29/10/2018

Université d'Ottawa
Bureau d'éthique et d'intégrité de la recherche

University of Ottawa
Office of Research Ethics and Integrity

Le Comité d'éthique de la recherche (CÉR) de l'Université d'Ottawa, opérant conformément à l'*Énoncé de politique des Trois conseils* (2014) et toutes autres lois et tous règlements applicables, a examiné et approuvé la demande d'éthique du projet de recherche ci-nommé.

L'approbation est valide pour la durée indiquée plus haut et est sujette aux conditions énumérées dans la section intitulée "Conditions Spéciales ou Commentaires". Le formulaire « Renouvellement ou Fermeture de Projet » doit être complété quatre semaines avant la date d'échéance indiquée ci-haut afin de demander un renouvellement de cette approbation éthique ou afin de fermer le dossier.

Toutes modifications apportées au projet doivent être approuvées par le CÉR avant leur mise en place, sauf si le participant doit être retiré en raison d'un danger immédiat ou s'il s'agit d'un changement ayant trait à des éléments administratifs ou logistiques du projet. Les chercheurs doivent aviser le CÉR dans les plus brefs délais de tout changement pouvant augmenter le niveau de risque aux participants ou pouvant affecter considérablement le déroulement du projet, rapporter tout évènement imprévu ou indésirable et soumettre toute nouvelle information pouvant nuire à la conduite du projet ou à la sécurité des participants.

The University of Ottawa Research Ethics Board, which operates in accordance with the *Tri-Council Policy Statement* (2014) and other applicable laws and regulations, has examined and approved the ethics application for the above-named research project.

Ethics approval is valid for the period indicated above and is subject to the conditions listed in the section entitled "Special Conditions or Comments". The "Renewal/Project Closure" form must be completedfour weeks before the above-referenced expiry date to request a renewal of this ethics approval or closure of the file.

Any changes made to the project must be approved by the REB before being implemented, except when necessary to remove participants from immediate endangerment or when the modification(s) only pertain to administrative or logistical components of the project. Investigators must also promptly alert the REB of any changes that increase the risk to participant(s), any changes that considerably affect the conduct of the project, all unanticipated and harmful events that occur, and new information that may negatively affect the conduct of the project or the safety of the participant(s).

Kim THOMPSON
Responsable d'éthique en recherche / Protocol Officer
Pour/For **Daniel LAGAREC** Président(e) du/ Chair of the **Comité d'éthique de la recherche en sciences sociales et humanités / Social Sciences and Humanities Research Ethics Board**

550, rue Cumberland, pièce 154
Ottawa (Ontario) K1N 6N5 Canada

550 Cumberland Street, Room 154
Ottawa, Ontario K1N 6N5 Canada

613-562-5387 • 613-562-5338 • ethique@uOttawa.ca / ethics@uOttawa.ca
www.recherche.uottawa.ca/deontologie | www.recherche.uottawa.ca/ethics