SSI-DT79165206

Doctoral Thesis

Development and Accuracy Analysis of a Human Motion Capturing Method Using a Single IR Camera

(単眼 IR カメラを用いた人間のモーションキャプチャー手法の開発と精度解析)

Alireza Bilesan

September, 2021

Division of Systems Science and Informatics
Graduate School of Information Science and Technology
Hokkaido University

Doctoral Thesis
submitted to Graduate School of Information Science and Technology,
Hokkaido University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy.

Alireza Bilesan

Thesis Committee: Prof. Atsushi Konno
Prof. Takayuki Tanaka
Prof. Hajime Igarashi

Development and Accuracy Analysis of a Human Motion Capturing Method Using a Single IR Camera*

Alireza Bilesan

Abstract

Motion capture systems are used to measure the kinematic features of motion in numerous fields of research. Despite the high accuracy of commercial systems, these systems are costly and used in limited conditions. Kinect has been proposed as a low-cost markerless motion capture sensor, and its accuracy has been assessed compared to previous motion capture systems like Vicon and OptiTrack. Kinect comes with a software development kit (SDK) that helps determine the human skeleton structure from the point cloud of the body using machine learning algorithms. Since the Kinect skeleton model detected by the SDK is different from the skeleton model estimated by the traditional motion capture systems, the incompatibilities in determining the human joint angles in different motion capture systems can cause imprecision in the accuracy assessment results reported in previous researches. To achieve a proper accuracy evaluation of the Kinect as a motion capture camera, we applied the inverse kinematics techniques in both skeleton models represented by Kinect and a traditional system (Vicon) to estimate lower body joint angles of a human during gait. The results indicated acceptable accuracy for Kinect in tracking knee and hip flex-extension angles using Vicon data as the true value. However, Kinect showed major errors in capturing delicate motions like ankle and pelvic joint angles. We developed a marker-based motion capture system using the Kinect IR camera to overcome these inaccuracies in human motion capture using Kinect skeleton. We introduced joint use of the Kinect

*Doctoral Thesis, Division of Systems Science and Informatics, Graduate School of Information Science and Technology, Hokkaido University, SSI-DT79165206, September 24, 2021.

IR camera as a pinhole camera model and its depth data in order to determine the 3-dimensional coordinates of retroreflective markers placed on the leg of a humanoid robot. This method helped us estimate the ankle joint angle of the robot accurately, which paved the way for a breakthrough in capturing other delicate joint angles.

Due to the limited number of detected joints by the Kinect algorithm and the inaccuracy in estimating the 3D position of the joints' centers, Kinect could not achieve the same level of accuracy as traditional systems. Joint use of the Kinect skeleton algorithm and Kinect-based marker tracking method was developed to track the 3D coordinates of multiple landmarks on human subjects using a single Kinect. By applying the joint constraints and inverse kinematics techniques on the acquired landmarks' 3D positions, the motion's kinematic parameters of five human subjects and a humanoid robot were calculated during gait trails. The humanoid robot test was used to evaluate the accuracy of the proposed method and a traditional system (OptiTrack) compared to the robot data used as true value. The results of the robot test assured the high-level accuracy of the OptiTrack system. Furthermore, the advantage of applying joint constraints on the captured data by the Kinect was demonstrated. Finally, the full accuracy assessment of the proposed method was done in capturing lower-body joint angles of five healthy subjects during ten gait trials for each subject. The OptiTrack data were used as ground-truth while the accuracy of the proposed Kinect-base system was compared to the Kinect skeleton model used in previous researches and an IMU-based system (Perception Neuron). The absolute agreement and consistency between each optical system and the robot data in the robot test and between each motion capture system and OptiTrack data in the human gait test were determined using intraclass correlations coefficients (ICC3). The reproducibility between systems was evaluated using Lin's concordance correlation coefficient (CCC). The correlation coefficients with 95% confidence intervals (95%CI) were interpreted substantial for both OptiTrack and the proposed method (ICC>0.75 and CCC>0.95) in the humanoid test. The results of the human gait experiments demonstrated the advantage of the proposed method (ICC>0.75 and RMSE=1.1460°) over the Kinect skeleton model (ICC<0.4 and RMSE=6.5843°) and Perception Neuron (ICC<0.4

and RMSE=4.2396°).

Keywords: motion capture systems, accuracy assessment, Kinect, humanoid robot, gait analysis, joint angles

Contents

1. Introduction **1**

1.1 Background 1

1.1.1 Accuracy assessment of the Kinect 3

1.1.2 Improving the accuracy of the Kinect 3

1.2 Previous Works 3

1.2.1 Kinect v2 accuracy assessment 4

1.2.2 Improving Kinect motion capture accuracy 4

1.3 Research Purpose and Contribution 5

1.3.1 Research purpose 5

1.3.2 Contribution 6

1.4 Outline of the Thesis 7

2. Markerless human motion tracking using Kinect skeleton model **9**

2.1 Introduction 9

2.2 Setup of the experiment 12

2.3 Joint angles' definition 13

2.4 Pelvic and hip joint angles estimation using inverse kinematics . . . 15

2.5 Experiment's results 18

2.6 Summary 21

3. Marker-based motion tracking using Kinect IR camera **25**

3.1 Introduction 25

3.2 Tracking retroreflective markers using Kinect IR camera 26

3.2.1 Detecting the markers using IR image data 26

3.2.2 Calculating the 3D positions of the markers 28

3.3 Accuracy assessment of the proposed marker tracking system 30

3.3.1 Static wand test 30

3.3.2 Dynamic humanoid test 32
3.4 Summary 33

4. Developed motion capturing using Kinect skeleton and depth sensor 39
4.1 Introduction 39
4.2 Lower body motion capturing using Kinect skeleton and depth sensor 41
4.2.1 Concept 41
4.2.2 Tracking human motions 42
4.2.3 Tracking humanoid robot motions 44
4.2.4 Kinematic analysis 46
4.2.5 OptiTrack model 47
4.3 Accuracy assessment experiments 47
4.3.1 Humanoid robot motion capturing 47
4.3.2 Human motion capturing 51
4.4 Summary 56

5. Conclusion 71
5.1 Conclusion 71
5.1.1 Summary 71
5.1.2 Future work 73

Acknowledgements 75

References 77

List of Publications 84

Appendix A Statistical analysis 85

List of Figures

1.1 Human motion capture using motion capture cameras. 2

2.1 Human motion capture using Kinect v2 and Vicon. The Kinect accuracy is evaluated by comparing the results to the Vicon system as the gold standard. The Kinect world coordinate is located on its IR camera, and the Vicon world coordinate is set on the floor. The plug-in-gait marker set is utilized in the Vicon system to capture human motion. 10

2.2 Kinect v2 skeleton model with 25 estimated joints. 12

2.3 Experimental setup, (A) a Kinect sensor is placed in three different positions (0 degree, 45 degree, and 90 degree) to capture a walking motion. Concurrently, six Vicon motion capture cameras record the same motion. (B) walking path is located at a distance of 1 m to 3 m from the Kinect. 13

2.4 Definition of the pelvic and the femur coordinate systems in the Kinect model. The hip joint angles are determined as the Euler rotation angles between the pelvic and femur coordinate systems. The pelvic angles are identified as the Euler rotation angles between the permuted Kinect coordinate and the pelvic coordinate frame. The spatial angles between the two vectors crossed at the knee and the ankle joint centers are considered as the knee and ankle joint angles, respectively. 16

2.5 Definition of the pelvic and femur coordinate systems in the Vicon model. The hip joint angles are determined as the Euler rotation angles between the pelvic and femur coordinate systems. The pelvic angles are identified as the Euler rotation angles between the Vicon world coordinate and the pelvic coordinate systems. The black solid line circles indicate the markers located on the lower body and the dashed circles indicate the estimated joint centers in the Vicon software. 17
2.6 Comparison of the captured joint angles using Kinect and Vicon, (A) rotation about $x-$axis (ψ), (B) rotation about $y-$axis (θ), (C) rotation about $z-$axis (φ), and (D) spatial angles. Black dotted lines: Kinect data, red solid lines: Vicon data. 22
2.7 RMSE of Kinect data compared to Vicon data, (A) pelvic angles measurement error related to the Kinect position, (B) lower body joint angles measurement error related to the Kinect position, (C) pelvic angles measurement error related to the motion speed, and (D) lower body joint angles measurement error related to the motion speed. 23

3.1 Kinect v2 sensor. 27
3.2 A reflective marker detected by Kinect IR camera. 27
3.3 Sample calibration image captured by Kinect IR camera showing the chessboard. 28
3.4 A wand with two markers placed at 240 mm interval. The wand was also placed at seven specific distances from the Kinect. 32
3.5 Reflective markers placed on the HRP-2 knee, ankle, and foot segments. The 3D positions of the markers were used to determine the ankle joint angle. 34
3.6 Layout of the Kinect, HRP-2, and four Optitrack cameras. 35
3.7 Ankle joint angle captured by the Kinect and Optitrack in a sinusoidal flexion/extension test. Humanoid data was used as ground-truth. 36

4.1 Concept of the proposed marker-based motion capture using Kinect v2. 42
4.2 A human gait is captured by Kinect, Perception Neuron, and OptiTrack. 58
4.3 Plug-in-Gait lower body marker set with redundant markers (left), and detected virtual markers in specific search domains defined using the Kinect skeleton (right). 59
4.4 A virtual marker with a radius of 3 pixels is estimated in the midpoint of two adjacent reflective markers in a Kinect IR image (this technique was used in the human walking test using redundant markers). 59
4.5 Search domain v_1 to detect the landmark RASI. 60
4.6 Plug-in-Gait marker set placed on the main and lower part of an HRP-2 humanoid robot (up), and estimated markers in specific search domains defined using the human skeleton (down). 61
4.7 An imaginary marker area (blob) with a radius of 3 pixels is estimated around the center of a reflective marker in a Kinect IR image (this technique was used to track the markers placed on the humanoid robot in the robot walking test). 62
4.8 Layout of the gait experiment. A Kinect and six OptiTrack cameras were used to capture the gait motion of a humanoid robot (in the robot walking test) and five human subjects (in the human walking test). 63
4.9 Captured joint angles of the humanoid during a walking test. 64
4.10 Definition of the lower body joint angles using the segments' coordinate systems. 65
4.11 Average results of the lower body joint angles of ten trials of the five human subjects. 66
4.12 Average results of the ankle joint angles of ten trials of the five human subjects. 67
4.13 Results of the lower body joint angles of a single trial of a random subject. 68
4.14 Results of the ankle joint angles of a single trial of a random subject. 69

List of Tables

2.1 Mean correlation, standard deviation (SD), and RMSE error between the motion capture systems data in the moderate speed motion test. Kinect is positioned in front of subject. 20

2.2 Error comparison between our method and previous studies. 20

3.1 The Kinect IR camera's estimated intrinsic and distortion parameters. 29

3.2 Mean errors of distance measurement using the Kinect in the static wand test. 31

3.3 Mean correlation, standard deviation (SD), and RMSE error of the motion capture systems in the dynamic humanoid test. 35

4.1 Pearson correlation (r) between the motion capture systems and humanoid data. r is interpreted as perfect linear correlation ($r = 1$), no linear correlation ($r = 0$), and strong negative correlation ($r = -1$). 48

4.2 Lin's concordance correlation (CCC) between the motion capture systems and humanoid data. CCC is interpreted as perfect (CCC>0.99), substantial (0.95−0.99), moderate (0.90−0.95), and poor (CCC<0.90). 48

4.3 Absolute agreement (ICC) between the motion capture systems and humanoid data. ICC is interpreted as excellent (ICC>0.75), good (0.6−0.75), fair (0.4−0.6), and poor (ICC<0.4). 49

4.4 Consistency (ICC) between the motion capture systems and humanoid data. ICC is interpreted as excellent (ICC>0.75), good (0.6−0.75), fair (0.4−0.6), and poor (ICC<0.4). 49

4.5 RMSE between the motion capture systems and humanoid data. . . 50

4.6 Pearson correlation (r) between the low-cost motion capture systems and OptiTrack data. r is interpreted as perfect linear correlation ($r = 1$), no linear correlation ($r = 0$), and strong negative correlation ($r = -1$). 52
4.7 Lin's concordance correlation (CCC) between the low-cost motion capture systems and OptiTrack data. CCC is interpreted as perfect (CCC>0.99), substantial (0.95−0.99), moderate (0.90−0.95), and poor (CCC<0.90). 53
4.8 Absolute agreement (ICC) between the low-cost motion capture systems and OptiTrack data. ICC is interpreted as excellent (ICC>0.75), good (0.6−0.75), fair (0.4−0.6), and poor (ICC<0.4). 54
4.9 Consistency (ICC) between the low-cost motion capture systems and OptiTrack data. ICC is interpreted as excellent (ICC>0.75), good (0.6−0.75), fair (0.4−0.6), and poor (ICC<0.4). 55
4.10 RMSE between the low-cost motion capture systems and OptiTrack data. 55

Chapter 1. Introduction

1.1 Background

Human motion data are used in a wide range of applications, including motion analysis, disability recognition, and rehabilitation in biomechanics [1]. Furthermore, these data are conveyed to graphical models and robots to generate human-like motions [2, 3, 4], and to teach new skills to robots as in imitation learning [5]. A common approach to obtain human motion data is by using marker-based motion capture systems, which are highly accurate, such as OptiTrack (NaturalPoint Inc., USA), a research-grade motion capture system [6]. However, these systems are expensive, require a highly trained operator, and their use is limited to laboratory settings. By improvement in camera and sensor technologies, markerless motion capture systems emerged as publicly available, cost-efficient motion capture cameras. Kanko et al. [7] developed a markerless motion capture system using multiple 2D cameras and a deep learning algorithm to estimate the 3D skeleton of humans and assessed their system's accuracy in gait experiments (commercialized as Theia3D). Microsoft Kinect, which was developed as an accessory for the Xbox video game console, attracted the research community in order to capture human motion data without using markers. Kinect version 2 (Kinect v2) was released in 2014 as an RGB-D camera with improved RGB and infrared (IR) camera resolutions compared to Kinect version 1 (Kinect v1). Moreover, the Kinect v2 came with a software development kit (Kinect SDK), which could create a 3D skeletal model of the human body with 25 estimated joint centers. The Kinect SDK utilizes a machine-learning algorithm to recognize different parts of the human body. Each part of the body is estimated with a connecting line as a bone, and the points which connect two adjacent bones are estimated as the joint centers. The achieved human skeleton using the Kinect SDK is called Kinect skeleton model

(Fig. 4.1a) [8]. Thanks to these technical improvements, Kinect v2 was introduced to research society as a low-cost motion capture sensor [9, 10].

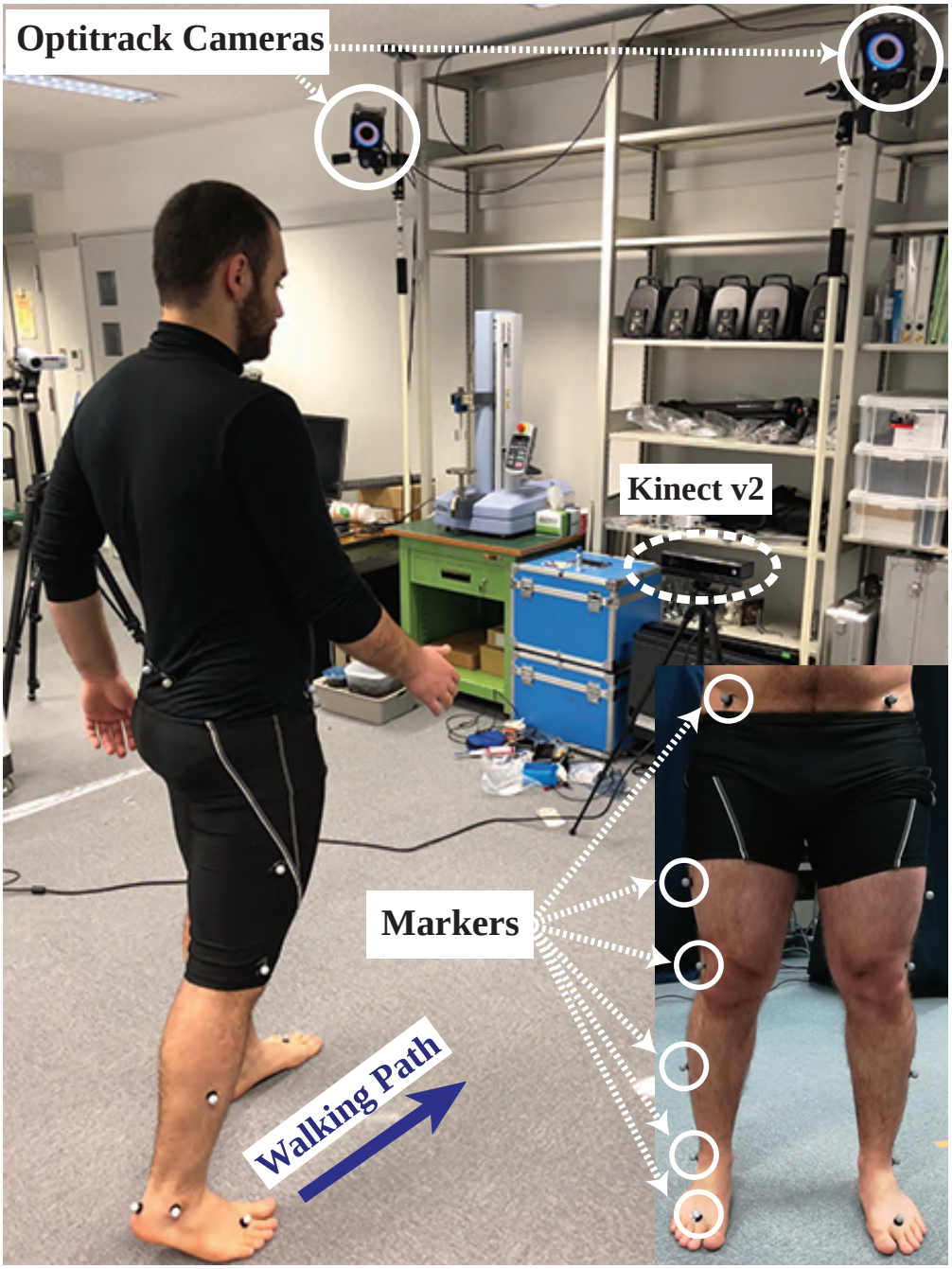


Figure 1.1: Human motion capture using motion capture cameras.

In order to use Kinect as a motion capture camera, the accuracy of such a system needs to be evaluated compared to traditional motion capture systems, which their high accuracy has been proved (see Fig. 1.1). Due to the limitation

in the identification of specific joint centers, and inaccuracy in estimating the position of the joint centers, further development in motion capturing using Kinect is needed in order to improve the accuracy.

1.1.1 Accuracy assessment of the Kinect

Most of previous studies evaluate the accuracy of Kinect in capturing joint angles in the anatomical plane which is beneficial in biomechanical applications. They use the Kinect skeleton model data to estimate the human joint angles and compare the results to traditional motion capture systems, which uses different methods to calculate these joint angles. The first goal of this thesis is to achieve a better assessment of the Kinect in detecting kinematic parameters of a human in a motion. Evaluating the accuracy of Kinect in capturing spatial joint angles compared to traditional systems can pave the way for further uses of Kinect in imitation learning and motion recognition applications.

1.1.2 Improving the accuracy of the Kinect

The limited number of identified joints and the error in the position estimation of the joint centers could result in inaccuracies in the joint angles measurement using Kinect skeletal data. By using the Kinect IR sensor data and the skeleton model detected by the Kinect algorithm, further improvement in capturing the positions of multiple landmarks on human body using a single Kinect can help to improve the motion capturing capability of this sensor. Moreover, by implementing anatomical constraints on the data we could achieve a high-level accuracy of human motion capture using Kinect. Finally, the advantage of such a system needs to be assessed in capturing human and a humanoid robot in several dynamic tests.

1.2 Previous Works

The previous studies focused on evaluating the accuracy of the Kinect in capturing human motion, by comparing the results to the traditional motion capture

systems. Moreover, some researches tried to improve the motion capturing accuracy of Kinect by using several methods.

1.2.1 Kinect v2 accuracy assessment

Th accuracy of Kinect in estimating the joint positions of the lower limbs was shown to be lower than the accuracy of estimating the positions of the upper limbs' joins [11]. Additionally, using the Kinect skeletal data to track 3D kinematics of lower body in a dynamic test could result in significant errors compared to traditional systems [12]. Vilas-Boas et al. investigated the accuracy of Kinect v1 and v2 in capturing spatial joint angles of the whole body in a walking test. Their results showed a poor correlation between the Kinect and a traditional optical system for most joint angles, except for knee angles [13]. Other studies indicated fair accuracy for the Kinect skeletal model in capturing knee joint angles in the sagittal plane during gait. However, they reported imprecision in capturing hip and ankle joint angles and did not investigate joint angles in other anatomical planes [14, 15, 16]. A review on validity of the Kinect for gait assessment also attested poor performance of the Kinect skeleton tracking algorithm in estimating most kinematic variables. Although, Kinect could estimate some spatiotemporal factors such as step width and step length accurately detectable [17]. The Kinect accuracy has been evaluated using optoelectronic motion capture systems or IMU sensors (such as Perception Neuron (Noitom Ltd., China)) as the ground truth, ignoring the fact that these commercial systems have imprecisions. Similar to Kinect, the Perception Neuron system was developed for gaming applications, and its validation for human motion capture purposes is still under study.

1.2.2 Improving Kinect motion capture accuracy

The limited number of identified joints and the error in the position estimation of the joint centers could result in inaccuracies in the joint angles measurement using Kinect skeletal data. In order to improve the Kinect skeleton accuracy, multi-camera and data fusion systems came to attention. There are several benefits for human motion assessment using multiple depth cameras since it would reduce

occlusion, increase accuracy by fusing data, and extend the field of view [18, 19]. However, there are some limitations in using multi-camera systems like complexity in the setup, which is similar to commercial motion capture systems. Additionally, extra hardware requirements could add to the cost of the system, and the use of different depth cameras in an overlapped field of view would negatively affect the results. Contrary to the setup complexity, these systems could not always provide superior results compared to single-eye RGB-D camera systems [20]. Some of these limitations can be solved by merging the Kinect skeleton data with an IMU wearable motion capture data. This method would improve the accuracy of the human motion assessment; however, the complexity and costs of such systems would likely top multi-camera systems [21].

Tracking 3D coordinates of colored markers using Kinect RGB camera and depth sensor was performed to capture lower-body joint angles in a gait test [22]. This method showed good agreement with a traditional motion capture system. However, this method cannot be implemented in many environments since the color of the background objects would affect the results; hence, excessive considerations are necessary before each test.

1.3 Research Purpose and Contribution

1.3.1 Research purpose

Firstly, the accuracy of the Kinect v2 is evaluated in capturing the kinematic parameters of multiple human subjects during gait using the Kinect skeleton model provided by the Kinect SDK. Since the previous studies had reported poor results in lower limb results, we focused on capturing the lower body joint angles of the subjects. Most of these studies evaluate the Kinect accuracy in capturing joint angles in the anatomical plane, which is beneficial in biomechanical applications. This study aims to better assess the Kinect spatial joint angles detection, which can be valuable in imitation learning and motion recognition applications.

Secondly, in order to improve the motion capturing using a single Kinect camera, a marker-based motion capture system using a single Kinect v2 is introduced to overcome the previous limitations of using the Kinect skeleton model. Image

processing of the data collected by the Kinect sensor is used to track the ankle joint angle of a humanoid robot, and the accuracy of the proposed method and a commercial motion capture system is compared to the robot data as the definite ground-truth.

Lastly, the proposed marker-based human motion capturing using Kinect v2 is developed to capture the kinematic parameters of lower body during a dynamic motion. Primarily, the significant advantage of a commercial motion capture system over Kinect needs to be evaluated in estimating the 3D kinematics of a humanoid robot, where the robot data could be used as the ground truth. Also, the effectiveness of kinematic constraints on the Kinect data needs to be studied by analyzing the accuracy improvement of Kinect data after applying the constraints. Importantly, the accuracy of the proposed system needs to be evaluated in capturing kinematic variables of the lower body of various subjects in dynamic tests, where the commercial motion capture system can be used as the ground truth.

1.3.2 Contribution

The contributions of this thesis are as follows.

Contribution in the accuracy assessment of the Kinect

1. Using inverse kinematics to calculate the lower body joint angles of the Kinect skeleton model

Despite previous studies which used the anatomical planes to estimate the joint angles in the Kinect skeleton model, we used inverse kinematics in order to estimate the 3D joint angles of the lower body using the Kinect skeleton model. Therefore, we could reach a better evaluation of the Kinect accuracy in capturing human joint angles. The same method is implemented in commercial motion capture systems; hence, the comparison of the results of the two systems would present better assessment of the Kinect as a motion capture camera. Moreover, the relation of the position of the Kinect and the accuracy of this method was evaluated in this part to determine the most optimized position for the Kinect to achieve the best accuracy in human motion capture. Lastly, the relation of the Kinect accuracy and

the subject speed was evaluated to provide a recommendation for the movement speed in motion capturing using Kinect.

Contribution in the improvement of the motion capturing using a single Kinect

1. Marker-based motion capturing using Kinect IR sensor

In this thesis, we proposed a marker-tracking method using Kinect's RGB-D camera. By performing single-eye camera calibration, and utilization of the Kinect's image and depth data, a marker-based motion capture using single Kinect was developed. Tracking the position of retroreflective markers placed on a humanoid robot during a flexion/extension motion demonstrated the accuracy of the proposed technique compared to a commercial motion capture system.

2. Accuracy assessment of the proposed method in capturing the kinematic parameters of multiple human subjects and a humanoid robot

The advantage of the proposed motion capture system was assessed in estimating the joint angles of multiple human subjects and a humanoid robot during gait. By utilizing the Kinect skeleton model and IR camera to detect and track multiple retroreflective markers placed on a human or a humanoid robot, the joint angles of lower part of the tracking subject was estimated and compared to a ground-truth. The motion capturing systems' statistical analysis provides an accurate assessment of the proposed method compared to other traditional motion capture systems.

1.4 Outline of the Thesis

The thesis consists of five chapters.

Chapter 1 is an introduction of this thesis and the reason why the single-eye motion capturing using Kinect is selected and improved.

In Chapter 2, the accuracy of the skeleton model provided by the Kinect software is evaluated in capturing multiple human subjects. Despite previous researches, we use inverse kinematics to calculate the lower body joint angles of the subjects during gait, which is similar to the technique used in commercial systems. The results of a commercial motion capture system is used as ground-truth.

The accuracy assessment results indicates the need for further improvement in the accuracy of motion capturing using Kinect.

In Chapter 3, a marker-based motion capturing using single Kinect is presented in order to improve the motion capturing accuracy. Kinect IR camera as a single-hole camera is calibrated and the Kinect depth data were used to estimate the 3D coordinates of three retroreflective markers placed on a leg of a humanoid robot. The accuracy of the marker-based motion capture system is evaluated by comparing the Kinect data and a commercial system to the robot's data as ground-truth.

In Chapter 4, the motion capturing method using a single Kinect is expanded to detect several reflective markers place on different human and humanoid subjects. Capturing the 3D position of the markers placed on the humanoid robot using the proposed method and a commercial system is performed, the robot's joint angles were estimated using the inverse kinematics. In the second experiment, lower body joint angles of five human subjects were estimated using Kinect and other traditional motion capture systems in order to evaluate the accuracy of the proposed method. The statistical analysis are done in this section to prove the advantage of the proposed method over the other commercial systems.

Chapter 5 summarizes the thesis and suggests future works.

Chapter 2. Markerless human motion tracking using Kinect skeleton model

2.1 Introduction

In order to reconstruct and transfer human intentions into a robot, motion capture systems are employed. Many researchers have employed human motions in robot imitation learning and human-like motion generation [23][24][25]. For better imitation, human and humanoid walking patterns are analyzed to apply the human walking functions to the humanoid robots [2, 3]. Despite high accuracy, commercial motion capture systems are high-priced and difficult to use. The data collecting is restricted to special settings and conditions since several cameras are required to capture one motion. For example, operating multi-camera calibration is essential before each experiment. Commercial systems are mainly used in indoor environments due to the stated restrictions. Markerless motion capture systems were proposed to overcome the previous issues [26]. These motion analysis technologies enabled researchers to assess movement characteristics as more cost-effective and straightforward. Despite the benefits of these new systems, systematic limitations restrain their functionality. For example, wearable electromagnetic sensors are affected by gravity noise and signal drift [27]. Moreover, these sensors are still high-priced and require a skillful data analyzer to post-process the data.

Microsoft released Kinect version 1 (Kinect v1) as an accessory for the Xbox 360 video game platform in 2010, designed for gaming purposes. Still, it was also used as a markerless, affordable, and portable motion capture sensor by the

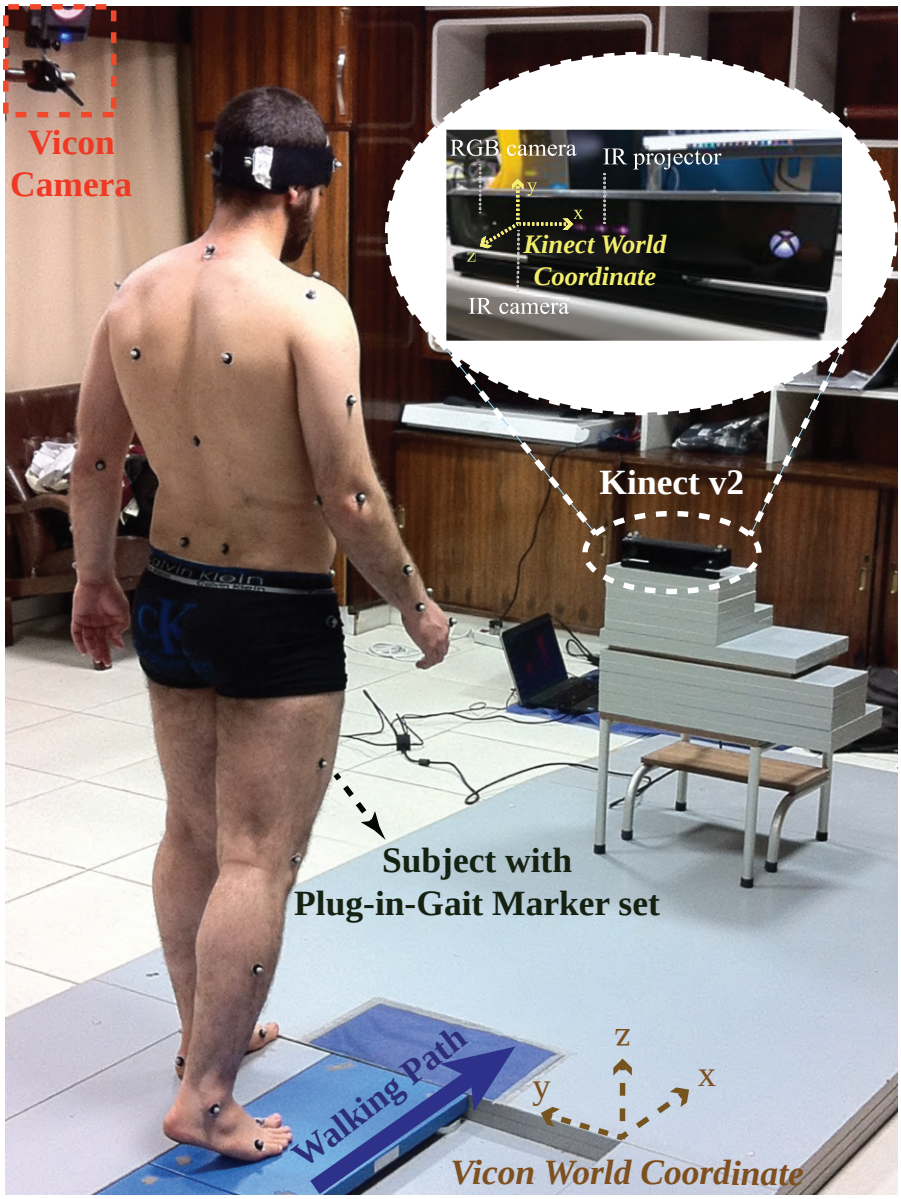


Figure 2.1: Human motion capture using Kinect v2 and Vicon. The Kinect accuracy is evaluated by comparing the results to the Vicon system as the gold standard. The Kinect world coordinate is located on its IR camera, and the Vicon world coordinate is set on the floor. The plug-in-gait marker set is utilized in the Vicon system to capture human motion.

research society. The Kinect v1 consists of one IR emitter, one IR camera, and one RGB camera, which acquires depth and color images of the scene. The depth of the scene is estimated employing speckle pattern technology. In 2014, Microsoft released the second version of the Kinect (Kinect v2) with enhanced RGB and IR camera resolution and a broader field of view (see Fig. 3.1). In Kinect v2, Microsoft used a different technology called time-of-flight (TOF) to measure the depth of the scene [28]. The speckle pattern of the Kinect v1 was shown to be remarkably affected by the sunlight. However, the TOF technology assisted the Kinect v2 in outdoor motion capture measurements. Still, the quality of the captured data strongly depends on the incidence angle of the rays in both versions of the Kinect when direct emission of infrared light is hitting the sensor [29].

With the help of the Kinect software development kit (SDK), the Kinect v2 provides a model of a three-dimensional skeleton with 25 joints of the human whose full body is placed within the field of view of the Kinect IR camera (see Fig. 2.2). The Kinect skeleton model has been used in various applications such as biomechanics [10], robotics [30], and computer vision [31]. Poor correlation between Kinect v1 skeleton data and commercial motion capture systems has been reported during lower extremity motion assessment in [16] and [32]. Due to the low technological specs of the Kinect v1, these deficient performances were not unforeseen. Furthermore, previous studies have shown that the Kinect is more capable of detecting spatiotemporal parameters compared to kinematic variables [17]. Thanks to the technological improvement of the Kinect v2, some researches have been able to track the sagittal plane joint angles of the human motion during functional movements [15][33][14].

Most of these studies evaluate the Kinect accuracy in capturing joint angles in the anatomical plane which is beneficial in biomechanical applications. The goal of this study is to achieve a better assessment of the Kinect spatial joint angles detection, which can be valuable in imitation learning and motion recognition applications. In addition, the relation of the Kinect accuracy with the subject position and movement speed is evaluated.

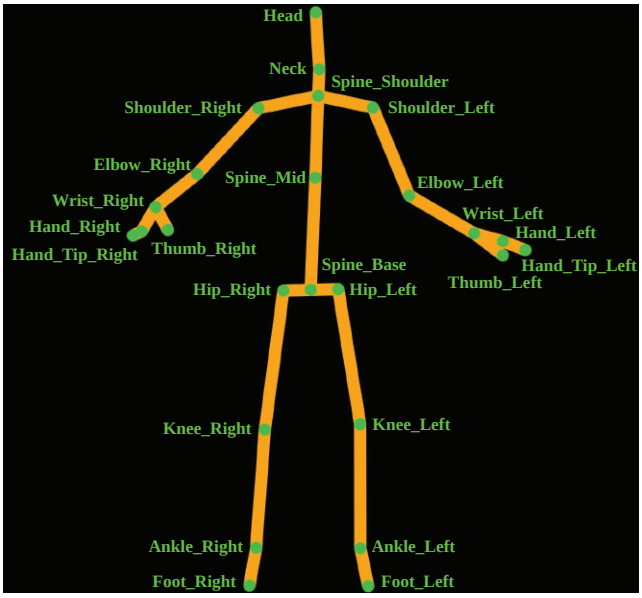


Figure 2.2: Kinect v2 skeleton model with 25 estimated joints.

2.2 Setup of the experiment

Five healthy adults (males, 24 years, height 175 ± 5 cm, weight 75 ± 10 kg) were selected to perform a normal gait motion at different speeds while motion capture systems were recording their movements. The subjects were asked to warm-up for five minutes before each test with different walking speeds (slow: $0.5 \pm 0.1\,\mathrm{m}/s$, moderate: $1.0 \pm 0.1\,\mathrm{m}/s$, and fast: $1.4 \pm 0.1\,\mathrm{m}/s$). Every trial was repeated 10 times for each subject, and the collected data were averaged before the statistical analysis. The motion capture systems were recording the human gait motion, simultaneously.

Kinematic data were captured simultaneously using a six-camera Vicon motion capture system (Vicon, Oxford, UK), sampled at 120 Hz using Vicon Nexus 2.1 software, and a Kinect v2 (Microsoft, Redmond, WA), sampled at 30 Hz using Kinect SDK 2.0. The Kinect was positioned in three different locations from the subject's walking path (0 degree: in front of the path, 45 degree: right side at a 45° angle to the path, and 90 degree: right side at a 90° angle to the path), (see Fig. 2.3A). Moreover, The walking path started from a distance of 3 m and ended at 1 m from the Kinect (see Fig. 2.3B). In order to capture full-body motion, the

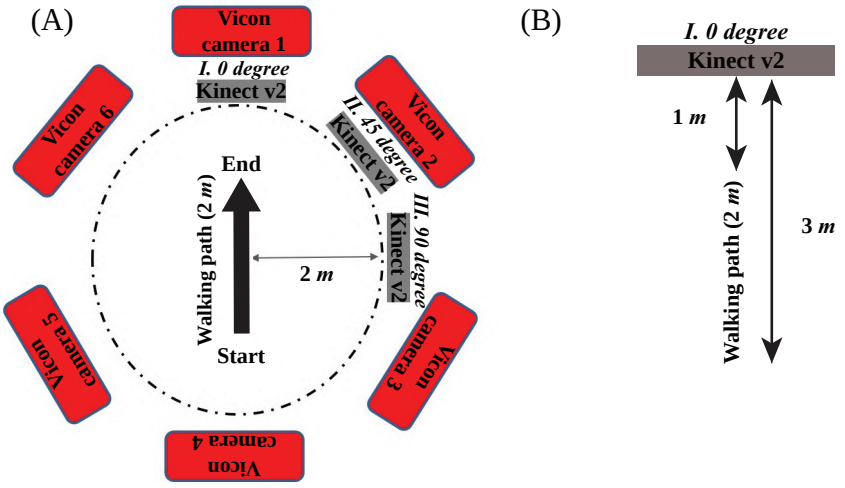


Figure 2.3: Experimental setup, (A) a Kinect sensor is placed in three different positions (0 degree, 45 degree, and 90 degree) to capture a walking motion. Concurrently, six Vicon motion capture cameras record the same motion. (B) walking path is located at a distance of 1 m to 3 m from the Kinect.

Kinect was located at the height of 0.75 m with 0.01° tilt angle (see Fig. 3.1).

The reflective markers were placed on each subject using the Vicon Plug-in-Gait marker set and modeled as previously described in [34]. The Subject was asked to perform a T-pose for 2 seconds to calibrate the two motion capture systems. Afterward, he started a normal walking motion through the designated path (see Fig. 3.1). The 3D model of the human body was created and captured in the Vicon Nexus software. Concurrently, the 3D skeleton model of the subject was generated, and the 3D positions of the body joint centers were recorded using the Kinect SDK.

2.3 Joint angles' definition

For the hip joint angles, the Vicon software uses the Plug-in-Gait kinematic calculations to determine the hip joint angles. The waist markers are used to determine an imaginary coordinate system located on the pelvis segment. An-

other coordinate system is defined on the femur segment using the thigh markers and the estimated hip and knee joint centers, which are calculated using the Vicon software. Eventually, the hip joint angles are distinguished using the Euler (Cardan) angles in rotation order of YXZ (see Fig. 2.5). The same method used in the Plug-in-Gait model was implemented to determine the hip joint angles in the Kinect model (see Fig. 2.2 and Fig. 2.4). The Hip_Right, $Spine\_Base$, and Hip_Left joint centers are used to define a plane on the pelvis segment in the Kinect skeleton model. The vector connecting the Hip_Right to the Hip_Left is assumed as y_{pelvic} and the normal vector of the pelvic plane is defined as z_{pelvic}. subsequently, x_{pelvic} is the cross product of the two previous vectors. The same method is used to determine the left femur plane using the Hip_Left, $Knee\_Left$, and the $Ankle\_Left$ joint centers in the Kinect model. The vector connecting the $Knee\_Left$ to the Hip_Left is defined as z_{femur}, the normal vector of the femur plane is y_{femur}, and the cross product of the two previous vectors is x_{femur}. Similarly, the right femur coordinate system is defined. It is worth remarking that the $y-$axis of the coordinate systems should be directed to the left side of the body and the $z-$axis should be directed upward (see Fig. 2.4).

The Euler rotation angles between the world coordinate and the pelvic coordinate frames are used to determine the pelvic angles in the Vicon model (see Fig. 2.5). The Vicon world coordinate system is set on the floor using a Vicon calibration wand at the beginning of the calibration procedure. Furthermore, the world coordinate system of the Kinect is located on its IR camera with the demonstrated coordinate axes shown in Fig. 3.1. To attain the Euler angles which represent the pelvic angles, it is required to permute the world coordinate system in the Kinect environment. The coordinate axes of the permuted Kinect coordinate system are parallel to the axes of the initial Kinect world coordinate. However, the x, y, and z axes of the permuted Kinect coordinate system have changed to the new axes shown in Fig. 2.4, where the $x-$axis is parallel to the walking path, the $y-$axis is toward the left side of the body, and the $z-$axis is directed upward. The Euler rotation angles between the permuted Kinect coordinate and the pelvic coordinate systems indicate the pelvic angles.

Eventually, the spatial angles between the two vectors passing through the lower body links and crossed at each joint centers were used to calculate the ankle

and the knee joint angles in the Kinect model (see Fig. 2.4). The same method was used in the Vicon model. Although these spatial angles were calculated in two different models with separate coordinate systems, we can still compare the Kinect data to the Vicon data to assess the Kinect accuracy in capturing these joint angles.

The coordinate systems definition of the body segments, in both Kinect and Vicon models, was compatible with the International Society of Biomechanics recommendations [35]. Moreover, the mathematical definitions of the joint angles in the Kinect and the Vicon model correspond as closely as possible to the existing clinical terminology.

2.4 Pelvic and hip joint angles estimation using inverse kinematics

To determine the hip joint angles, in both Kinect and Vicon models, the Euler (Cardan) rotation angles, which rotate the pelvic coordinate into the femur coordinate system, need to be calculated.

Equation (2.1) demonstrates the YXZ order Euler rotation matrix (E_{yxz}) which rotates the pelvic frame ($^{w}F_{pelvic}$) into the femur ($^{w}F_{femur}$) frame:

$$^{w}F_{femur} = E_{yxz} \cdot^{w} F_{pelvic} \ , \tag{2.1}$$

where E_{yxz} is:

$$E_{yxz} = \begin{bmatrix} c(\psi)c(\varphi)+s(\theta)s(\psi)s(\varphi) & -c(\psi)s(\varphi)+s(\theta)s(\psi)c(\varphi) & c(\theta)s(\psi) \\ c(\theta)s(\varphi) & c(\theta)c(\varphi) & -s(\theta) \\ -s(\psi)c(\varphi)+s(\theta)c(\psi)s(\varphi) & s(\psi)s(\varphi)+s(\theta)c(\psi)c(\varphi) & c(\theta)c(\psi) \end{bmatrix} ,$$

$^{w}F_{pelvic} = [x_{pelvic}|y_{pelvic}|z_{pelvic}]$ and $^{w}F_{femur} = [x_{femur}|y_{femur}|z_{femur}]$ are the projected femur and pelvic coordinate frames on the world frames, calculated using joint positions estimated by Kinect SDK. In E_{yxz}, $s(.)$ and $c(.)$ stand for $sin(.)$ and $cos(.)$, respectively. The rotation angles ψ, θ, and φ represents the rotations applied in $x-$, $y-$, and $z-$directions, respectively. Since $^{w}F_{pelvic}$ and $^{w}F_{femur}$ are orthogonal matrices, the (2.1) can be written as:

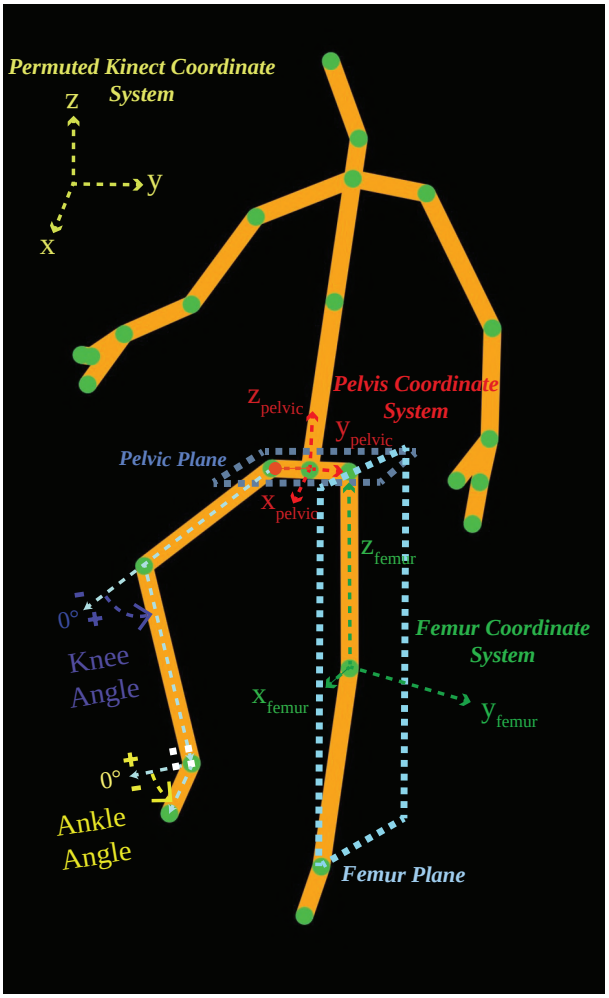


Figure 2.4: Definition of the pelvic and the femur coordinate systems in the Kinect model. The hip joint angles are determined as the Euler rotation angles between the pelvic and femur coordinate systems. The pelvic angles are identified as the Euler rotation angles between the permuted Kinect coordinate and the pelvic coordinate frame. The spatial angles between the two vectors crossed at the knee and the ankle joint centers are considered as the knee and ankle joint angles, respectively.

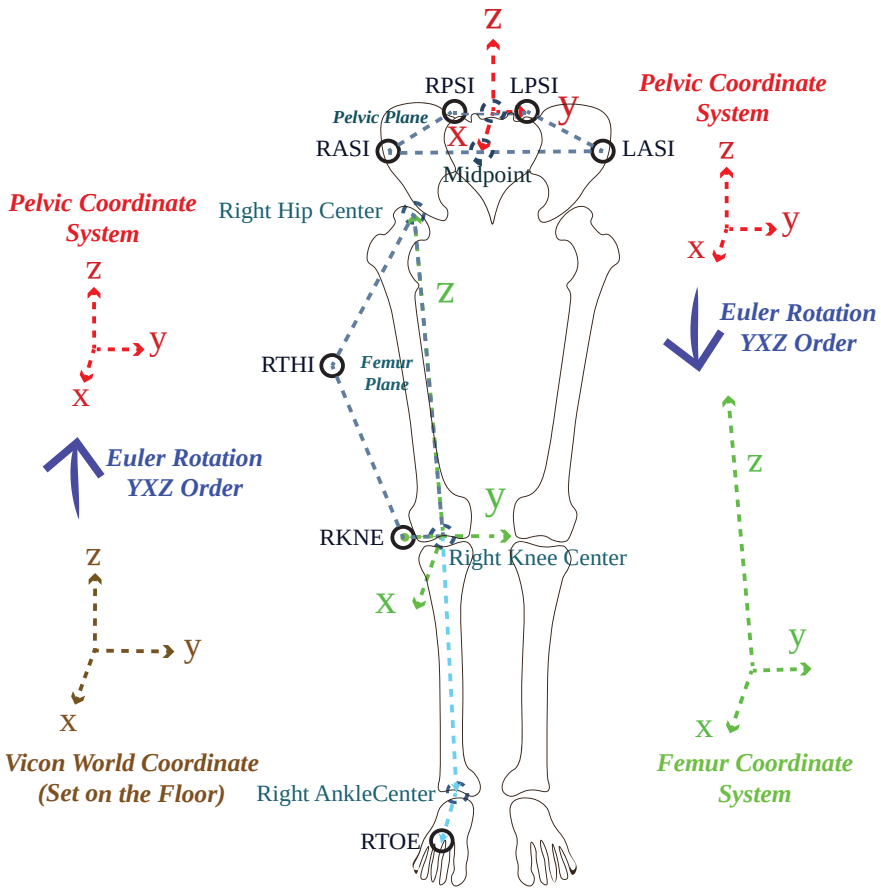


Figure 2.5: Definition of the pelvic and femur coordinate systems in the Vicon model. The hip joint angles are determined as the Euler rotation angles between the pelvic and femur coordinate systems. The pelvic angles are identified as the Euler rotation angles between the Vicon world coordinate and the pelvic coordinate systems. The black solid line circles indicate the markers located on the lower body and the dashed circles indicate the estimated joint centers in the Vicon software.

$$E_{yxz} = {}^{w}F_{femur} \cdot {}^{w}F_{pelvic}^{T} = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix}, \tag{2.2}$$

where ${}^{w}F_{pelvic}^{T}$ is the transpose matrix of ${}^{w}F_{pelvic}$, and e_{ij} represents the elements of matrices ${}^{w}F_{femur} \cdot {}^{w}F_{pelvic}^{T}$. Eventually, the Euler rotation angles are attained using (2.2):

$$\begin{aligned} \psi &= tan^{-1}(e_{13}/e_{33}) \\ \theta &= sin^{-1}(-e_{23}) \\ \varphi &= tan^{-1}(e_{21}/e_{22}). \end{aligned} \tag{2.3}$$

These Euler rotation angles indicate the hip joint angles. ψ represents the hip flexion/extension, θ represents the hip abduction/adduction, and φ represents the hip rotation angles.

The same method is used to calculate the pelvic angles in the Kinect and Vicon models. However, The pelvic angles are calculated using the Euler rotation angles between the world and pelvic coordinate frames. As previously stated, the initial world coordinate of the Kinect requires rotations to the permuted Kinect coordinate system:

$${}^{w}K_{p} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \end{bmatrix}, \tag{2.4}$$

where ${}^{w}K_{p}$ indicates the projection of the permuted Kinect coordinate frames on the initial world frames (see Fig. 2.4). The pelvic joint angles are calculated using (2.4) as the permuted world coordinate frame in the Kinect environment. However, this rotation is not required in the Vicon environment. In pelvis, ψ represents the pelvic tilt, θ represents the pelvic obliquity, and φ represents the pelvic rotation angles.

2.5 Experiment's results

To minimize any fluctuations in the Kinect data, a 5th-order low pass filter with a cut-off frequency of 4 Hz, was applied to the Kinect data [16]. The results

of the comparison between the two motion capture systems for a single trial, when the Kinect is located in front of the path and the subject walks slowly, are shown in Fig. 2.6. The red-line indicates the angles captured by Vicon cameras and the black dotted line indicates the Kinect captured data for a normal gait motion. Fig. 2.6 illustrates the capability of the Kinect in tracking the lower body joint angles. However, there are extreme differences between the Kinect and Vicon data in the hip rotation and ankle joint angles. Even though the range of the motions in the pelvic angles are notably low, these angles and their changes could be captured appropriately.

The average RMSE error of the captured joint angles, associated with the Kinect position and the walking speed, is shown in Fig. 2.7. Fig. 2.7(A) and Fig. 2.7(B) indicate the relation of the RMSE error with the Kinect position, while Fig. 2.7(C) and Fig. 2.7(D) indicate the effects of the walking speed on the Kinect accuracy. The best position for the Kinect is in front of the walking path. In this case, the Kinect can estimate the skeleton of the human more precisely. Furthermore, by increasing the movement speed, the RMSE increases in the ankle and pelvic angles. However, this effect is not reflected on the knee and hip joint angles.

The average value of the joint angles correlation (μ_r), its standard deviation (SD), and the root-mean-square-error (RMSE) derived from the Kinect and Vicon data are shown in table 2.1. The correlation coefficients and RMSE error of the pelvic rotation, hip ab/adduction, hip flex/extension, and knee flexion are within an acceptable range which indicates our method validity. However, the correlation coefficients of the pelvic obliquity, pelvic tilt, and hip rotation are considerably low, and further improvements are required in these parts. Moreover, the RMSE error in most joint angles are substantially small, and the Kinect data values are close to the Vicon data. The hip rotation angle is an internal rotation occurring over femur bone. The Kinect has extreme difficulties in estimating this joint angle. This can also be another research domain which requires further studies.

In table 2.2, the accuracy of the proposed method in capturing human gait using Kinect v2 is compared to similar Kinect validation studies. Kharazi et al. [15] and Xu et al. [36] used Kinect v2 and v1 to capture the lower body joint angles in the sagittal plane, respectively. Importantly, Guess et al. [12] calculated the relative Cardan rotation angles of the hip and knee using the joint orientations of the

Kinect SDK 2.0 model. They compared the Kinect data to the Vicon plug-in-gait model which is similar to our method. However, our method demonstrates better accuracy in capturing the hip and knee joint angles. The ankle joint angle captured in our method indicates the Kinect inability in tracking the foot segment. The results illustrate the feasibility of the proposed motion capture system in capturing knee and hip flexion/extension joint angles using the Kinect v2.

Table 2.1: Mean correlation, standard deviation (SD), and RMSE error between the motion capture systems data in the moderate speed motion test. Kinect is positioned in front of subject.

System	$\mu_r \pm$ SD	RMSE (degree)
Pelvic Obliquity	0.6641 ± 0.1204	2.288
Pelvic Tilt	0.2679 ± 0.1140	1.699
Pelvic Rotation	**0.9502 ± 0.0321**	**7.233**
Hip Ab/Adduction	**0.7562 ± 0.1739**	**3.240**
Hip Flex/Extension	**0.9779 ± 0.0113**	**3.914**
Hip Rotation	0.3592 ± 0.0663	53.996
Knee Flexion	**09834 ± 0.0314**	**3.247**
Ankle Dorsi/PlantarFlexion	0.7196 ± 0.1944	30.079

Table 2.2: Error comparison between our method and previous studies.

Method	Hip_RMSE (degree)	Knee_RMSE (degree)	Ankle_RMSE (degree)
Xu et al. [36]	11.9	29.0	-
Guess et al. [12]	12	11	-
Kharazi et al. [15]	5.9	6.3	23.3
Our Method	**3.9**	**3.2**	**30.0**

2.6 Summary

Capturing human joint angles using the Kinect sensor has attracted much attention with the applications in robotics, computer vision, and biomedical engineering. In this work, a markerless human motion tracking system was introduced using Kinect v2. The pelvic and hip joint angles were determined in the Kinect skeleton model using the inverse kinematics techniques. Moreover, for the knee and ankle joints, the spatial angles were determined. The same method was implemented in the skeleton model calculated in the Vicon software. Eventually, the Kinect accuracy was evaluated by comparing data obtained from the Kinect and Vicon. The mean correlation, standard deviation, and RMSE of the Kinect data were reported for the joint angles in table 2.1, and the accuracy of our motion capture system was compared to the previous researches in table 2.2. In order to reduce the RMSE error of the proposed motion capture system, the Kinect should be placed in front of the motion path. This indicates that the Kinect SDK have a better estimation of the human skeleton when the camera captures the whole front of the human body. Moreover, the subject would be better to perform motions with slower speed considering the Kinect low sampling time (30 Hz) (Fig. 2.7). The Kinect v2 is identified as a noisy sensor and its accuracy relies on its environmental situations. However, it managed to operate simultaneously with the Vicon cameras. The results illustrate the possibility of using Kinect in both human-body motion analysis and human-robot imitation tasks. However, further studies are required to improve the Kinect accuracy in estimating the ankle joint and hip rotation angles (table 2.1).

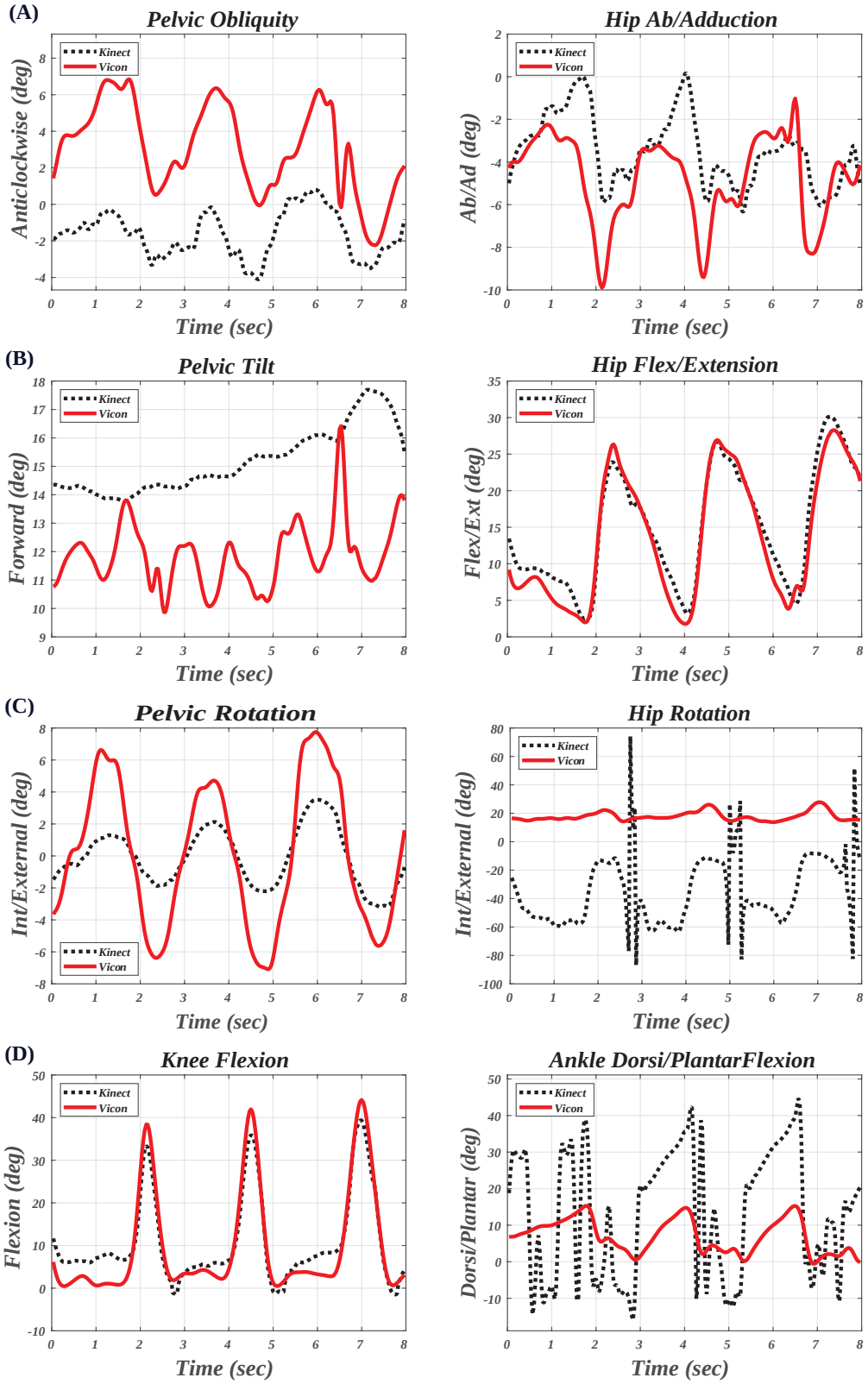


Figure 2.6: Comparison of the captured joint angles using Kinect and Vicon, (A) rotation about $x-$axis (ψ), (B) rotation about $y-$axis (θ), (C) rotation about $z-$axis (φ), and (D) spatial angles. Black dotted lines: Kinect data, red solid lines: Vicon data.

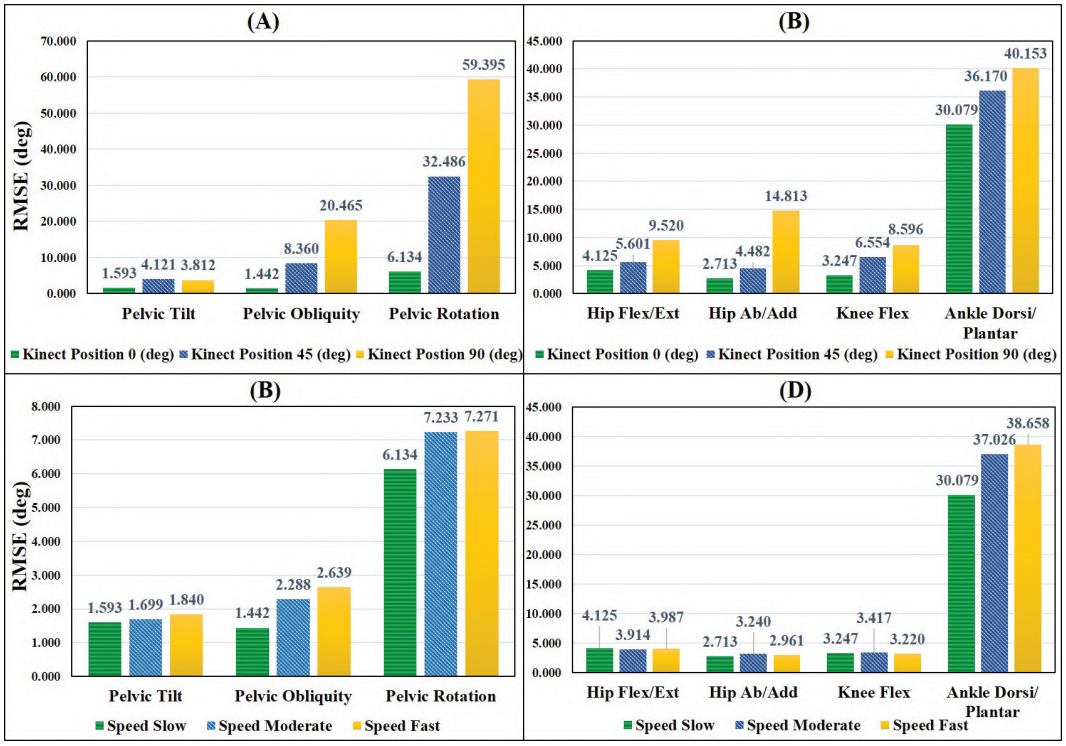


Figure 2.7: RMSE of Kinect data compared to Vicon data, (A) pelvic angles measurement error related to the Kinect position, (B) lower body joint angles measurement error related to the Kinect position, (C) pelvic angles measurement error related to the motion speed, and (D) lower body joint angles measurement error related to the motion speed.

Chapter 3. Marker-based motion tracking using Kinect IR camera

3.1 Introduction

In chapter 2, Kinect v2's accuracy in capturing lower body joint angles of a human subject was studied. Jebeli et al. [10] evaluated Kinect's accuracy in detecting the positions of the human's center of mass using the length of body links in the Kinect skeleton model. According to Jebeli's research, when measuring short parts of the body like the foot, Kinect suffers inaccuracy in recognizing the center of joints for that part. Kinect v1 and v2 accuracy in gait analysis was studied further in [16] and [15], where the commercial motion capture systems were used as the gold standard. The results indicated considerable errors in tracking ankle joint angles using both versions of Kinect. Despite providing valuable insights into Kinect as a capturing motion system, these previous researches do not provide information on the accuracy of the commercial systems. A landmark identification method used for hand kinematics as ground-truth measures and fabricated model of the upper body to evaluate the skeleton tracking capability of Kinect v1 in upper-body rehabilitation applications was done by [37] and [38], accordingly. However, the accuracy assessment of Kinect as a motion capture system with a ground-truth measure and how it compares to commercial motion capture systems to measure lower-extremity kinematics remains understudied. A digital inclinometer was employed to measure the angles of a ball-and-socket joint as ground-truth by [39]. The accuracy and repeatability of Kinect v1 in capturing static postures of the joint were qualified, and the performance was compared with a marker-based motion capture system. Due to the low sampling rate of Kinect compared

to commercial cameras, blurring or aliasing in data occurs during dynamic tests, such as walking or running.

This chapter introduces a new method to enable ankle joint angle measurement using Kinect v2 as a single motion capture camera in a dynamic test. Image processing of the data collected by the Kinect sensor was used to track the joint positions instead of using the Kinect skeleton data. In addition, the accuracy of the proposed method was quantified compared to a commercial motion capture system and a definite ground truth.

3.2 Tracking retroreflective markers using Kinect IR camera

3.2.1 Detecting the markers using IR image data

A Kinect v2 sensor was used to record a scene using its infrared (IR) camera (see Fig. 3.1). The IR image and the depth data were collected in real-time to be utilized later. The data above were employed to track the depth position of each marker in world coordination. For this purpose, each infrared frame was converted into a binary image, assuming that the intensity of the markers' reflection is beyond a certain threshold ($\alpha > 0.5$). In order to remove noises from the foreground, an erosion transformation, followed by a dilation was applied to the raw data. Connected pixels in the resulting binary image (blobs) were labeled based on the center of the nearest identified marker in the previous frame (see Fig. 4.7). The average value of depth in a circle with a radius of 3 pixels around each marker center was used to find the Z coordinate of each marker in meters. In contrast, X and Y components (in pixels) were assumed to be the same as the blob area center. There was no need for mapping the pixels of one onto another since both depth and infrared images are the Kinect IR sensor's output. Moreover, the time-of-flight technology used in this system estimated markers' depth, which was impossible by the Light Coding technology of the first version of Kinect.

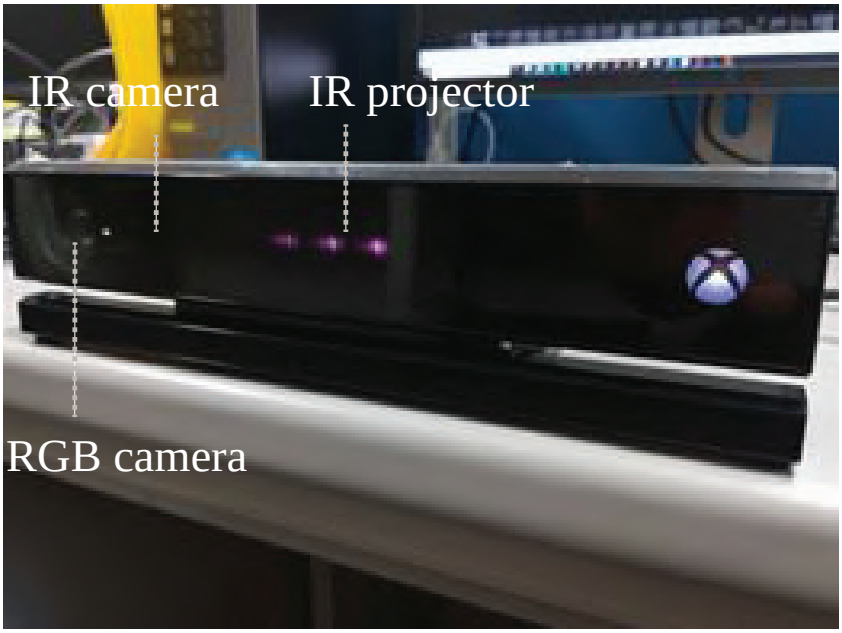


Figure 3.1: Kinect v2 sensor.

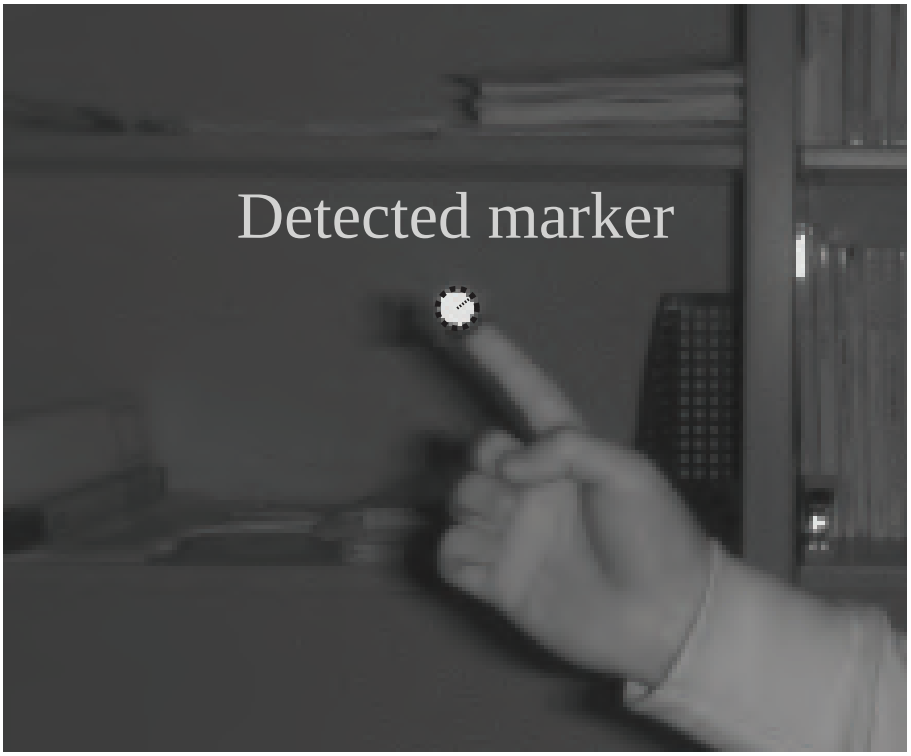


Figure 3.2: A reflective marker detected by Kinect IR camera.

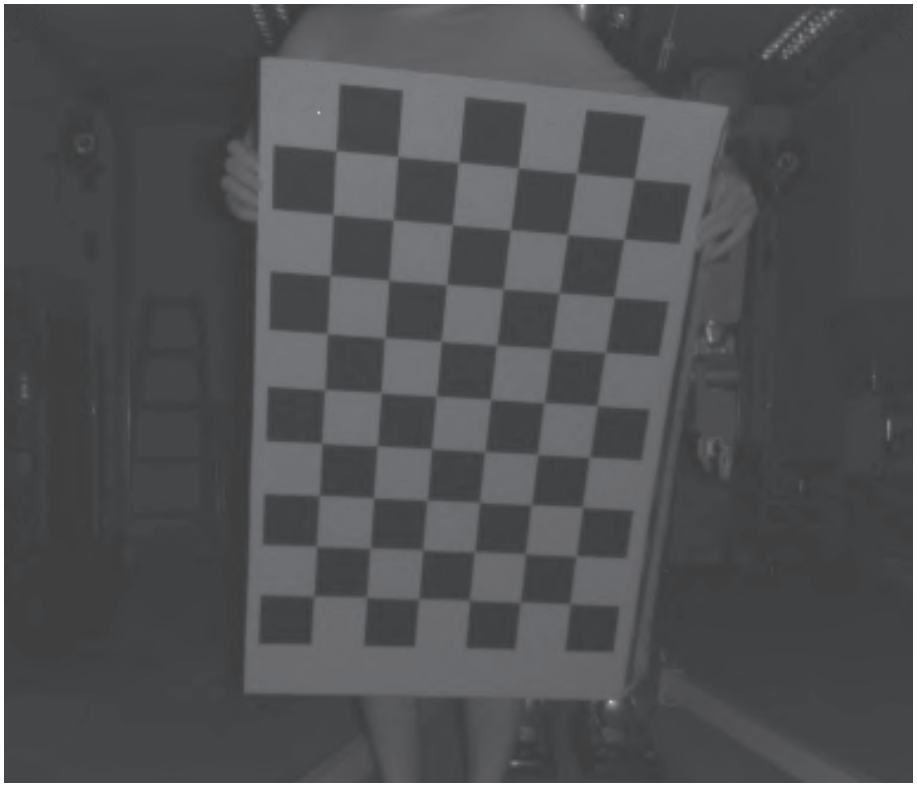


Figure 3.3: Sample calibration image captured by Kinect IR camera showing the chessboard.

3.2.2 Calculating the 3D positions of the markers

Kinect depth sensor was employed to capture the retroreflective markers' position data. Therefore, the markers' 2D position data were estimated in the Kinect IR camera image. By using the Kinect depth sensor, the corresponding depth (Z_c) parameter was determined. A camera calibration using an undistorted pinhole camera model is required to determine the intrinsic parameters and the lens distortion coefficients (the principal point, the focal lengths, the skew coefficient, the radial distortion, and the tangential distortion) [40], [41]. To calibrate the Kinect IR camera, a chessboard with 80 mm diameters squares was employed (Fig. 3.3). The camera parameters were calculated using MATLAB calibration toolbox [42].

The projection of a point from the image coordinates $p_i = [u_i, v_i]^T$ in pixels to the camera coordinates $P_c = [X_c, Y_c, Z_c]^T$ in meters is obtained through the following equations. First, the coordinates of the point were shifted by the coordinates

of the projection center shown in (3.1):

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} u_i \\ v_i \end{bmatrix} - \begin{bmatrix} u_{0i} \\ v_{0i} \end{bmatrix} , \tag{3.1}$$

where $p_{0i} = [u_{0i}, v_{0i}]^T$ is the principal point in pixels. Then, by using (3.2), distortion is corrected:

$$\begin{bmatrix} x_i^{corrected} \\ y_i^{corrected} \end{bmatrix} = \begin{array}{l} (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \begin{bmatrix} x_i \\ y_i \end{bmatrix} \\ + \begin{bmatrix} 2p_1 x_i y_i + p_2(r^2 + 2x_i^2) \\ p_1(r^2 + 2y_i^2) + 2p_2 x_i y_i) \end{bmatrix} \end{array} , \tag{3.2}$$

where f_x and f_y are the focal lengths in pixels and $r^2 = (x_i/f_x)^2 + (y_i/f_y)^2$. Moreover, k_1, k_2, and k_3 are the radial distortion coefficients and p_1, p_2, and p_3 are the tangential distortion coefficients of the lens. Finally, by using (3.3), the 2D parameters of the camera coordinates are calculated:

$$\begin{bmatrix} X_c \\ Y_c \end{bmatrix} = Z_c \cdot \begin{bmatrix} f_x & 0 \\ s & f_y \end{bmatrix}^{-1} \cdot \begin{bmatrix} x_i^{corrected} \\ y_i^{corrected} \end{bmatrix} , \tag{3.3}$$

where s is the skew coefficient. In table 3.1, The results of the estimated intrinsic and distortion parameters are demonstrated.

Table 3.1: The Kinect IR camera's estimated intrinsic and distortion parameters.

Intrinsic parameters	Values $(pixels)$	Distortion coefficients	Values
f_x	377.5511	k_1	0.2093
f_y	377.6204	k_2	-0.4019
u_{0i}	258.6627	k_3	-0.1426
v_{0i}	203.2391	p_1	0
-	-	p_2	0
-	-	p_3	0

3.3 Accuracy assessment of the proposed marker tracking system

3.3.1 Static wand test

Firstly, in a static wand test shown in Fig. 3.4, the validity of the proposed method was evaluated. A wand with two retroreflective markers placed on it at an exact interval of 240 mm was set in front of a Kinect v2 at seven specific distances. The Kinect was set at the height of 60 cm and its tilt angle was set to 0 degrees. The connecting line of the two markers was set parallel to the Kinect. In order to measure the markers' distance from the Kinect, a measuring tape was used to mark the seven positions of the wand (Fig. 3.4). A calibration procedure was necessary to relate the tape measurements to the Kinect depth data; therefore, the zero position of the Kinect depth sensor could be roughly estimated to place the zero point of the tape. Hence, the tape was placed on the ground and a marker was pasted on it at a specific position (e.g., at the amount of 200 cm). Afterward, the tape was moved forward and backward till the depth sensor could measure the same amount. In order to evaluate the accuracy of the motion capture system, the tape measurements and the markers' absolute interval were used as true values. The wand was first placed at an interval of 100 cm from the Kinect, and then we increased the distance to 400 cm with six stops at every 50 cm intervals. The test was repeated ten times, and the positions of the markers were measured by the tape and Kinect. Finally, a fifth-order low pass Butterworth filter with a cut-off frequency of 3 Hz was employed to filter the data and minimize any fluctuations in the Kinect data. Following presents the calculated error:

$$Error = Absolute_{data} - Kinect_{data} \ . \tag{3.4}$$

In table 3.2, the errors' mean value and standard deviation (SD) are presented. According to the results, The average error of the markers' distance from the sensor (μ_{depth}) and the average error of the markers' relative interval ($\mu_{relative}$) depend on the wand position. As the wand's distance from the Kinect increased, $\mu_{relative}$ reduced, but then it increased after the distance reached the amount of 200 cm. This pattern also occurred in the depth measurement, but the error started

to increase after the distance of 150 cm. These results matched with [43] which indicated by increasing the distance from the sensor, Kinect depth measurement accuracy decreases. Nevertheless, there is an optimum interval range (from 150 cm to 200 cm) in which the error is minimum. When the markers were placed closer than this range to the sensor, Kinect depth sensor error increased. This error could be caused by the experiment's situation where the Kinect is placed on a specific height (60 cm), and detecting the markers close to the ground could have been difficult for the sensor. In addition, the error in the relative interval ($\mu_{relative}$) might be caused by the IR camera calibration technique, too, since the focal lengths and the distortion coefficients measured by the toolbox are assumptions and these estimated intrinsic values include error. As the wand distance from the sensor increased, the $\mu_{relative}$ increased too which indicates that the calibration technique was not effective enough to reduce the potent lens distortion effects ultimately. A minus value for the error indicates that the intervals measured by the Kinect are larger than the true values measured by the tape and vise versa. The results represent the feasibility of the proposed method to measure markers' positions with an appropriate error. However, to achieve the optimum accuracy, it is recommended to place the markers within the range of 150 cm to 250 cm from the sensor (see table 3.2).

Table 3.2: Mean errors of distance measurement using the Kinect in the static wand test.

Distance from Kinect (cm)	$\mu_{depth} \pm$ SD (mm)	$\mu_{relative} \pm$ SD (mm)
100	6.3 ± 1.7	-4.8 ± 0.9
150	1.1 ± 2.1	-2.1 ± 0.7
200	3.6 ± 2.5	-0.4 ± 0.9
250	7.9 ± 2.7	2.5 ± 1.4
300	26.6 ± 8.3	2.9 ± 1.1
350	26.9 ± 7.0	3.5 ± 2.6
400	89.0 ± 17.2	6.7 ± 4.6

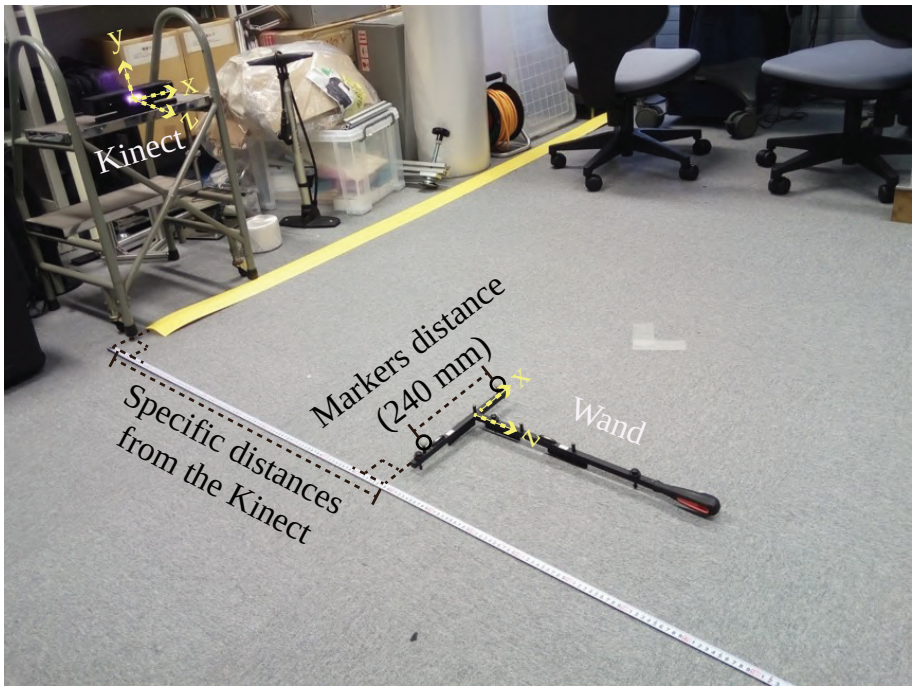


Figure 3.4: A wand with two markers placed at 240 mm interval. The wand was also placed at seven specific distances from the Kinect.

3.3.2 Dynamic humanoid test

Eventually, The accuracy of the proposed motion capture system was evaluated in a humanoid robot ankle flexion/extension test. The motion was captured by a Kinect v2 and a commercial motion capture system (Optitrack prime 41 cameras). The accuracy was assessed by using the humanoid joint data as the ground truth.

Our setup consisted of a Kinect v2 sensor and three reflective markers with 12.7 mm diameter placed on an HRP-2 humanoid robot lower-body links. The size and shape of the HRP-2 lower body make this robot a proper option to achieve a better human-like motion (see Fig. 4.6). Two markers were placed on the knee and foot segments of the humanoid. The third marker was set on the estimated pitch rotation center of the ankle joint. The Kinect was set in front of the robot at a height of 60 cm while four Optitrack cameras were placed around the robot to capture the same motion at the same time. The layout of the experiment is shown in Fig. 4.8. The humanoid was placed at the distance of 2 m from the Kinect, which is within the appropriate interval according to the static wand test

results. The robot performed a sinusoidal ankle flexion/extension motion from an initial position shown in Fig. 4.6. The only active degree of freedom of the robot was the ankle pitch motion, and all other joints were locked. Moreover, the robot was hanged for being enabled to perform the motion. The action started with a 30 degree flexion followed by a 60 degree extension and, finally, a 30 degree flexion to reach the initial position. The whole motion was performed in 8 seconds and was repeated ten times. By using the three markers' positions, the ankle joint angle captured by the Kinect and the Optitrack was determined. In order to define the 0 degree, the imaginary line perpendicular to the connecting line of the ankle and the knee markers was employed as the zero position. Consequently, the angle between this line and the connecting line of the foot and ankle markers was defined as the ankle joint angle. The initial ankle joint angle was measured by the Optitrack when the robot was not moving (-45.8 degree). The positive and negative directions shown in Fig. 4.6 were utilized to differentiate between the flexion and extension motions, respectively.

The average value of the ankle joint angle's correlation (μ_r), standard deviation (SD), and the root-mean-square error (RMSE) derived from the Kinect and the HRP-2 data, and also from the OptiTrack and the HRP-2 are presented in table 3.3. Plots of the ankle joint angle captured by the two motion capture systems are illustrated along with the robot's generated angles in Fig. 3.7. The Kinetc's average correlation coefficient and RMSE error indicate notable improvement in motion capturing using the proposed method compared to the previous researches[16], [15]. It is worth mentioning that the OptiTrack still has higher accuracy compared to the Kinect. The results illustrate the feasibility of developing a marker-based motion capture system using the Kinect depth sensor in order to overcome the Kinect skeleton's limits, such as estimating the ankle joint angles.

3.4 Summary

Capturing ankle joint angles using a Kinect sensor was known as a challenging task since the range of the motion is small and the shortness of the foot segment increases the sensor inaccuracy in gauging this delicate motion. In order to capture

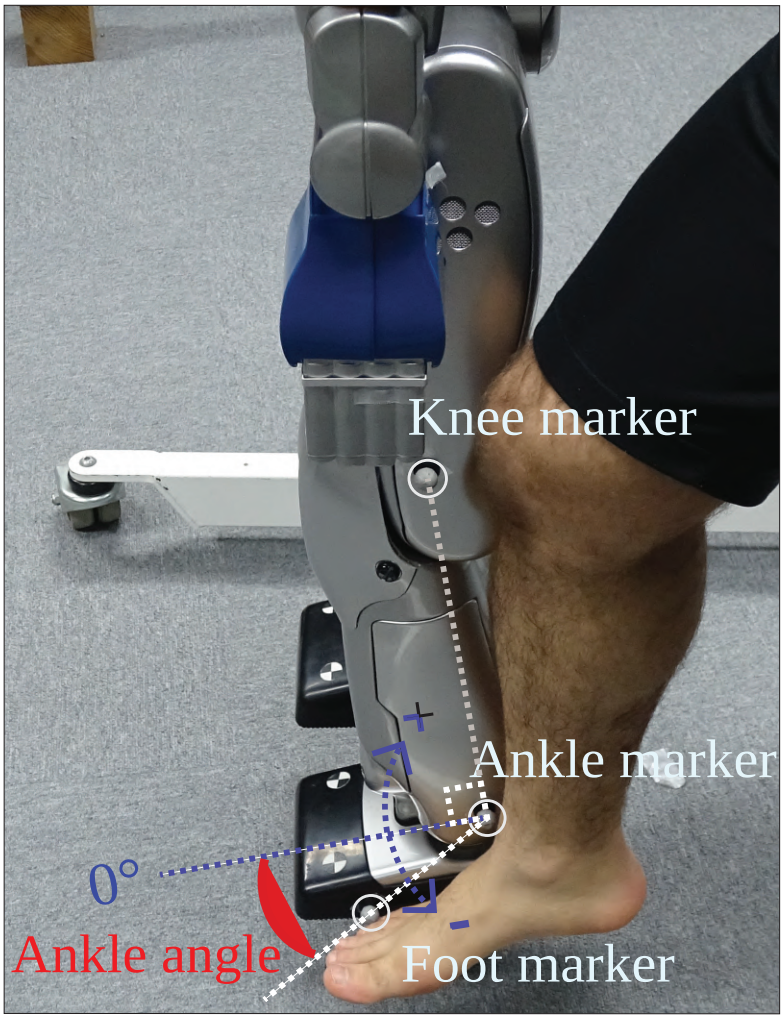


Figure 3.5: Reflective markers placed on the HRP-2 knee, ankle, and foot segments. The 3D positions of the markers were used to determine the ankle joint angle.

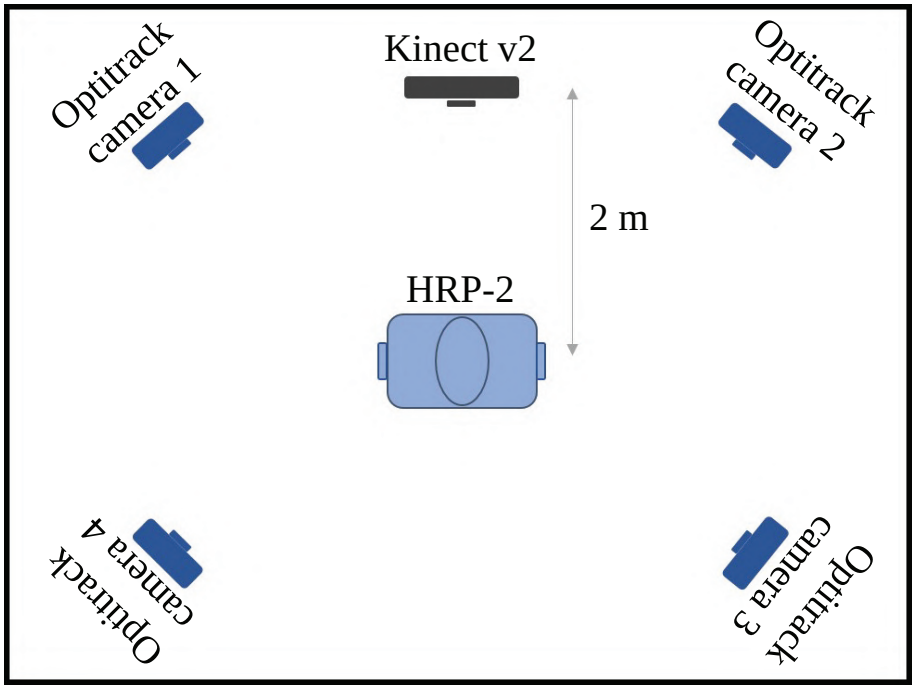


Figure 3.6: Layout of the Kinect, HRP-2, and four Optitrack cameras.

humanoid ankle joint angles, we introduced a novel marker-based motion tracking system with joint use of Kinect IR camera and its depth sensor data. By employing the humanoid data as ground truth, the accuracy of the novel motion capture system was evaluated. The mean correlation, standard deviation, and RMSE between the Kinect and the humanoid data were reported 0.8899, 0.0952, and 5.579, respectively. The results indicate remarkable improvement in capturing ankle joint angles using Kinect v2 by applying the proposed method. However,

Table 3.3: Mean correlation, standard deviation (SD), and RMSE error of the motion capture systems in the dynamic humanoid test.

System	$\mu_r \pm$ SD	RMSE (degree)
Jamali et al.	-	32.410
Kharazi et al.	0.0956 $\pm$ 0.2919	22.161
Our Method	**0.8899 $\pm$ 0.0952**	**5.579**
OptiTrack	0.9832 $\pm$ 0.0130	5.598

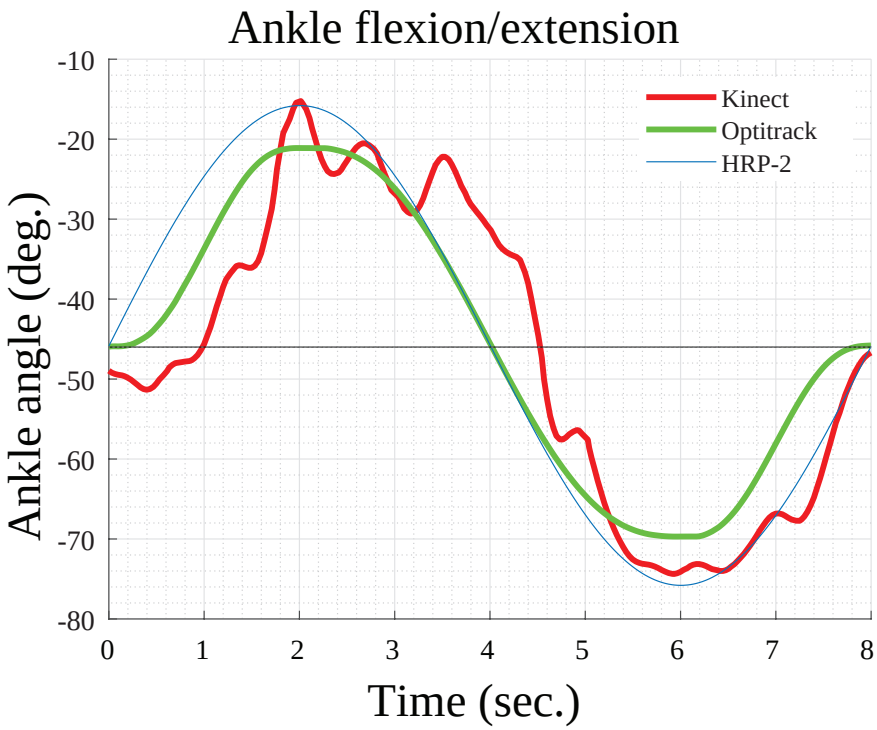


Figure 3.7: Ankle joint angle captured by the Kinect and Optitrack in a sinusoidal flexion/extension test. Humanoid data was used as ground-truth.

markers should be placed in a range between 150 cm and 200 cm from the Kinect IR camera in order to minimize the error.

Moreover, the calculated mean correlation, standard deviation, and RMSE of the OptiTrack motion capturing results gave 0.9832, 0.0130, and 5.598, supporting the superiority of this commercial motion capture system over the Kinect v2. Although Kinect v2 is identified as a noisy sensor and its accuracy relies on its environmental situations, it was able to estimate the markers' positions placed on the humanoid robot's surface (HRP-2). The results show the possibility of using the proposed method to capture dynamic motions in human motion analysis and robot telemanipulation tasks. Although, the Kinect data comes noisy and unstable, which becomes more significant in the forward dynamics-based simulation when the second-order differential of the joint trajectory is needed (Fig. 3.7. In order to overcome the instability of the Kinect data due to the noisiness, more research is required to improve the smoothness of the Kinect data so that the proposed method could be used in robotics applications.

Chapter 4. Developed motion capturing using Kinect skeleton and depth sensor

4.1 Introduction

A common approach to obtain human motion data is by using marker-based motion capture systems, which are highly accurate, such as OptiTrack (Natural-Point Inc., USA), a research-grade motion capture system [6]. However, these systems are expensive, require a highly trained operator, and their use is limited to laboratory settings. By improvement in camera and sensor technologies, markerless motion capture systems emerged as publicly available, cost-efficient motion capture cameras.

In chapter 2, Kinect v2 showed capability in capturing knee joint angles, imprecision exists in capturing hip and ankle joint angles. A review on validity of the Kinect for gait assessment also attested poor performance of the Kinect skeleton tracking algorithm in estimating most kinematic variables, and only some spatiotemporal factors such as step width and step length were accurately detectable [17]. In order to improve the Kinect skeleton accuracy, multi-camera and data fusion systems came to attention. There are several benefits for human motion assessment using multiple depth cameras since it would reduce occlusion, increase accuracy by fusing data, and extend the field of view [18, 19]. However, there are some limitations in using multi-camera systems like complexity in the setup, which is similar to commercial motion capture systems. Additionally, extra hardware requirements could add to the cost of the system, and the use of

different depth cameras in an overlapped field of view would negatively affect the results. Contrary to the setup complexity, these systems could not always provide superior results compared to single-eye RGB-D camera systems [20]. Some of these limitations can be solved by merging the Kinect skeleton data with an IMU wearable motion capture data. This method would improve the accuracy of the human motion assessment; however, the complexity and costs of such systems would likely top multi-camera systems [21].

In chapter 2, the limits of the Kinect skeleton in estimating the lower body was indicated. The proposed marker-based motion capture technique presented in chapter 3 indicated better accuracy in capturing the ankle joint angle of a humanoid robot in a flexion/extension test. In this chapter, we extend the development of the marker-based motion capturing using Kinect IR camera to track multiple landmarks of human subjects in order to estimate the lower body joint angles in dynamic tests. In order to determine the positions and orientations of the lower body segments more precisely, a global optimization method (GOM) [44] is applied to the 3D positions of the markers detected by the Kinect. Moreover, the GOM method also imposes joint constraints to the achieved Kinect model, which eliminates joint dislocations and gives better positions and orientation estimations for the model.

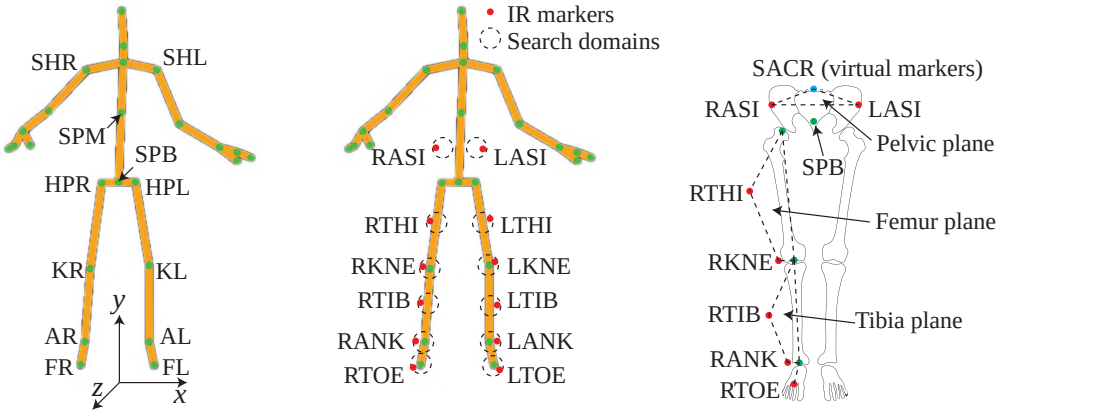
In order to evaluate a commercial motion capture system and the proposed method in estimating the 3D kinematics of a subject, a humanoid robot was used where the robot data could be used as the ground truth. The effectiveness of the kinematic constraints on the Kinect data needs to be studied by analyzing the accuracy improvement of Kinect data after applying the constraints. Importantly, the accuracy of the suggested system needs to be assessed in capturing kinematic variables of various subjects during gait, where the commercial motion capture system can be used as the ground truth.

4.2 Lower body motion capturing using Kinect skeleton and depth sensor

4.2.1 Concept

Fig. 4.1 shows the concept of the proposed marker-based motion capture using Kinect v2. Firstly, the joint centers (green dots) were detected using the Kinect skeleton model (Fig. 4.1a). The Kinect skeleton is obtained from a built-in algorithm that enables Kinect to detect human joint centers' positions. Secondly, search domains were defined to detect the IR markers (red dots) placed on the subject using the joint centers' positions (Fig. 4.1b). Thanks to the specific search domains, multiple IR markers could be detected using a single Kinect. Moreover, this method increases the algorithm's speed to search and recognize the high contrast pixels on the IR images to detect the virtual markers. The 3D positions of each marker were estimated using the Kinect depth sensor (chapter 3). The utilized technique in chapter 3 indicated the advantage of tracking IR markers using Kinect depth sensor; therefore, we extended this method to track multiple IR markers placed on the human subject to capture the kinematic parameters of the lower body. The name of each marker is chosen the same as the Plug-in-Gait landmarks (Fig. 4.1b). Since the Kinect camera could not detect the markers placed on the backside of the subject, SACR's (Sacral) position (blue dot) was estimated using the SPB (Spine_Base) joint's position (Fig. 4.1c). The detection of the SACR was necessary in order to create the Pelvic plane, using SACR, RASI (Right_Anterior_Superior_Iliac), and LASI (Left_Anterior_Superior_Iliac). Finally, a new skeleton model was attained using the detected markers similar to the International Society Biomechanics (ISB) model (Fig. 4.1c) [35]. The joint centers' positions of the new skeleton (ISB) model were calculated from the Plug-in-Gait landmarks using Kadaba et al.'s method [45].

Calculating joint angles using inverse kinematics on the Kinect skeleton model presented in chapter 2 provided better accuracy in capturing human joint angles. However, The Kinect skeleton suffers from technical errors in estimating the join centers' positions [11]. In this paper, we replaced the Kinect skeleton with the calculated ISB model to improve the estimation of the human skeleton using a single



(a) Kinect skeleton model

(b) Search domains to find IR markers

(c) International Society of Biomechanics (ISB) recommendation model

Figure 4.1: Concept of the proposed marker-based motion capture using Kinect v2.

Kinect camera. By applying the inverse kinematics on the attained ISB skeleton model, lower body joint angles were calculated in a dynamic test (Fig. 4.10). The advantage of the proposed marker-based motion capture was shown in human gait tests.

4.2.2 Tracking human motions

In the human gait test (Fig. 4.2), Plug-in-Gait lower body marker set was employed to analyze the gait kinematics of the subjects. Since the Time-Of-Flight sensor of the Kinect cannot estimate the depth of reflective markers directly [46], two adjacent markers were placed near each other. The midpoint of the adjacent markers was assumed as a virtual marker, and the Kinect depth sensor will estimate the virtual marker's depth. This technique would help to improve the reliability of the markers' depth estimation using the Kinect IR camera [47]. Two adjacent markers were placed around each front lower body landmark of the Plug-in-Gait, in a way that the midpoint of each adjacent markers would indicate the Plug-in-Gait landmarks (Fig. 4.3). A Kinect v2 was employed to record the IR images of the walking scene using its IR camera. A reliable threshold ($\alpha > 0.5$, where $1 > \alpha > 0$ denotes the brightness of each pixel in the Kinect IR image, as $\alpha = 0$ the darkest and $\alpha = 1$ the brightest pixel) was applied to each IR image to identify high-contrast mark-

ers. The two adjacent markers were employed to estimate a circle with a radius of 3 pixels in an area centered in the midpoint of the two markers. The new imaginary blob was assumed as a virtual marker in the Kinect IR camera (Fig. 4.4), and its depth data was recorded and used to calculate the 3D coordinates of the Plug-in-Gait landmarks by applying the pinhole camera calibration (chapter 3). Tracking virtual markers using Kinect IR camera was presented in [47] and expected to improve the reliability of the recorded depth data compared to the direct tracking of the retroreflective markers (chapter 3). Since the Kinect camera could not detect the landmarks located on the human's backside, only one marker was placed on each of these landmarks (RPSI (Right_Posterior_Superior_Iliac), LPSI (Left_Posterior_Superior_Iliac), RHEE (Right_Heel), and LHEE (Left_Heel) in Fig. 4.2), and these markers were used to detect the human skeleton in the OptiTrack model. In total, 28 retroreflective markers (24 markers on the front and 4 markers on the back) were placed on each human subject.

In order to track multiple markers simultaneously and online, the Kinect skeleton algorithm was used to define specific search domains in the Kinect IR images, in which every two adjacent markers were detected to estimate the virtual marker. On the right side of the pelvis, a circle search domain with a radius of 20 pixels and centered in $(x_{HPR},\ \ (y_{HPR} + y_{SPM})/2)$ was determined to detect the RASI landmark in the Kinect IR image (v_1), where SPM (Spine_Mid) and HPR (Hip_Right) indicate the joint centers in the Kinect skeleton model (see Fig. 4.5). For the right thigh markers, a circle with a radius of 20 pixels and centered in $(x_{KR},\ \ (y_{KR} + y_{HPR})/2)$ was determined to detect RTHI (Right_Thigh) (v_2). For the right knee markers, a circle with a radius of 15 pixels and centered in $(x_{KR},\ \ y_{KR})$ was determined to detect RKNE (Right_Knee) (v_3). For the right tibia markers, a circle with a radius of 15 pixels and centered in $(x_{AR},\ \ (y_{AR} + y_{KR})/2)$ was determined to detect RTIB (Right_Tibia) (v_4). For the right ankle markers, a circle with a radius of 10 pixels and centered in $(x_{AR},\ \ y_{AR})$ was determined to detect RANK (Right_Ankle) (v_5). For the right foot markers a lower-half-circle with a radius of 15 pixels and centered in $(x_{FR},\ \ y_{FR})$ was determined to detect the RTOE (Right_Toe) (v_6), where $x_{:}$ and $y_{:}$ indicate the 2D position of the joint centers in the Kinect IR images which are determined by Kinect skeleton (KR (Knee_Right), AR (Ankle_Right), and FR (Foot_Right)) (Fig. 4.3). The

same method can be applied to detect the markers placed on the left side of the human (v_7–v_{12}). This technique would increase the markers tracking algorithm's speed since it searches for the markers in specific domains instead of the whole image.

The RPSI and LPSI markers could not be detected by Kinect camera which was placed in front of the subject (Fig. 4.2); hence, to create the human skeleton the SACR landmark (Fig. 4.2) was assumed as $s = (x_{SPB}, \quad y_{SPB} + D, \quad z_{SPB} - D)$ in the Kinect skeleton model (Fig. 4.3), where D was empirically estimated 9 cm. This assumption was based on an approximate distance measurement between the imaginary SPB (Fig. 4.3) and SACR (Fig. 4.2) landmarks of all five subjects.

The new skeleton model (ISB model in Fig. 4.1c) can be achieved from the markers' positions (v_1–v_{12}) using Kadaba et al.'s method in calculating join centers' positions from Plug-in-Gait landmarks [45]. The s, v_1, and v_7 markers were used to define the 3D coordinate system located on the pelvis (chapter 2). Using the position of s and the lower body joint centers, 3D coordinate systems located on the pelvis, hip, and tibia were defined [45]. The lower body joint angles were calculated using the inverse kinematics (chapter 2). However, there was no marker assumption for the RHEE and LHEE markers (Fig. 4.2) in the Kinect model, and the foot coordinate system could not be achieved due to the lack of landmarks on foot. Therefore, the ankle joint angle was estimated by measuring the spatial angle consists of the foot marker, ankle, and knee joint centers.

4.2.3 Tracking humanoid robot motions

In the proposed motion capture technique, Kinect skeleton model is used to define specific search domains in the Kinect IR images to track multiple markers in real-time. However, since the Kinect skeleton algorithm is designed to detect joint centers of a human, it could not estimate a skeleton for the humanoid robot. Hence, to track the multiple markers placed on the robot simultaneously and online, the skeleton of the human standing beside the robot was projected on the robot body to define ten search domains (Fig. 4.6).

In the humanoid gait test (Fig. 4.6), ten retroreflective markers were placed on the covered surface of a humanoid robot HRP-2 [48] in order to analyze the

kinematics of the robot. Since the robot has a reflective surface, it was covered with black cloth to avoid any reflective noises. A Kinect v2 was employed to record the IR images of the walking scene. A reliable threshold ($\alpha > 0.5$) was applied to each IR image to identify high-contrast markers. An imaginary blob was defined as the marker area by estimating a circle with a radius of 3 pixels around each detected marker's center (Fig. 4.7). The average value of depth in each blob was recorded as the marker's depth, while the position of the center of the blob was assumed as X and Y components (in pixels). By applying pinhole camera calibration for the IR camera of the Kinect, 3D coordinates of the estimated markers were calculated (the method introduced in chapter 3). Since the Time-Of-Flight (TOF) sensor of the Kinect v2 is unable to estimate the depth of high reflective surfaces [46], we recorded the depth data of the pixels adjacent to the high contrast pixels, and determined the marker's depth by averaging the available depth data. We tried to use as few markers as possible attached to the robot body to avoid further complexity; hence, instead of using redundant markers, the reflective markers were tracked directly in the humanoid experiment (different from the tracking virtual marker technique used in the human walking test).

In order to detect the three markers placed on the upper body of the robot (u_1–u_3), three circle search domains were defined with a radius of 20 pixels and centered in $((x_{HTR} - L_1)\,,\ \ y_{SHR})$, $(x_{HTR},\ \ y_{SHL})$, and $((x_{HTR} - L_2)\,,\ \ y_{SPM})$, where $x_:$ and $y_:$ are the 2D position of the human joint centers in the Kinect IR images obtained from the Kinect skeleton model (see Figs 4.3 and 4.6 for the joints' names), and $L_1 = x_{SHL} - x_{SHR}$ and $L_2 = x_{SHL} - x_{SPM}$ are the distances used to reach the desired positions on the humanoid. To detect the markers placed on the left leg of the humanoid (u_4–u_8), five circle search domains were defined with a radius of 10 pixels and centered in $(x_{HTR},\ \ y_{HPR})$ for the hip, $(x_{HTR},\ \ (y_{KR} + y_{HPR})/2)$ for the thigh, $(x_{HTR},\ \ y_{KR})$ for the knee, $(x_{HTR},\ \ (y_{AR} + y_{KR})/2)$ for the tibia, $u_8 = (x_{HTR},\ \ y_{AR})$ for the ankle marker. The two markers placed on the left foot of the humanoid (u_9–u_{10}) were detected by employing two search domains with a radius of 8 pixels and centered in $(x_{HTR},\ \ y_{FR})$, $(x_{HTR},\ \ (y_{FR} - L_3))$, where $L_3 = x_{HPL} - x_{HPR}$ (HTR (Hand_Tip_Right), SHR (Shoulder_Right), SHL (Shoulder_Left), SPM (Spine_Mid), HPR (Hip_Right), HPL (Hip_Left), KR (Knee_Right), AR (Ankle_Right), and FR (Foot_Right)) (Fig. 4.6). This technique would increase

the speed of the markers tracking algorithm since it looks for the markers in specific domains instead of the whole image.

A rigid body with a 3D coordinate system (x_u, y_u, and normal vector z_u) was determined on the humanoid's upper body using the u_1, u_2, and u_3 markers (Fig. 4.6). Similarly, the upper leg coordinate system was determined using the u_4, u_5, and u_6 markers and the lower leg coordinate system was determined using the u_6, u_7, and u_8 markers. It is important to note that the u_6 marker was placed on the pitch rotation center of the robot's knee, empirically. Since there were only two markers placed on the foot segment, the foot's 3D coordinate system could not be achieved. However, the robot was programmed to walk in a way that the y–axis of the upper body and foot would always be parallel to each other; hence, the y–axis of the upper body's coordinate system (y_u) was used to determine the y–axis of the foot's 3D coordinate system (y_f) (Fig. 4.6). Finally, the humanoid's lower body joint angles could be calculated using the inverse kinematics in both Kinect and OptiTrack models.

4.2.4 Kinematic analysis

The lower body joint angles were calculated using the inverse kinematics techniques explained in chapter 2. A global optimization method (GOM) [44] was applied to the 3D positions of the markers detected by the Kinect in order to determine the positions and orientations of the lower body segments accurately. This method also imposes joint constraints, capable of eliminating joint dislocations and giving accurate model positions and orientation estimations. In order to apply the joint constraints, there should be at least three markers placed on each rigid body part; hence, the GOM could not be applied on the foot markers. The GOM is expected to improve the estimation of the pelvic, hip, and knee joint angles (MATLAB Genetic Algorithm).

The coordinate systems definition of the body segments was compatible with the International Society of Biomechanics recommendations [35]. The calculated joint angles for the ten trials were analyzed before the joint angles comparison and statistical analysis.

4.2.5 OptiTrack model

One marker from each couple markers in the human walking test was selected to create the human skeleton in the OptiTrack software (the marker closest to the Plug-in-Gait landmark). The OptiTrack skeleton model contains joint constraints, and the lower body joint angles were calculated automatically by the software. In the Kinect model, the 3D positions of the detected markers $v_{:}$ were employed to estimate the lower body joint centers using David et al. method [49]. It is worth mentioning that the SACR landmark is estimated in the middle of RPSI and LPSI markers in the OptiTrack software, which is used to create the pelvic coordinate system (Fig. 4.2).

In the humanoid robot test, the 3D positions of the markers $u_{:}$ were captured by the OptiTrack, and the robot joint angles were calculated using inverse kinematics.

4.3 Accuracy assessment experiments

4.3.1 Humanoid robot motion capturing

In order to evaluate the accuracy of the several motion capture methods, motion capturing experiments were performed using humanoid robot HRP-2 (Fig. 4.6). Assuming the joint angles measured by optical encoders installed at the joints of the HRP-2 as the ground truth, the joint angles calculated from the 3D coordinates of the markers are compared with the ground truth. Three motion capturing methods were simultaneously tested; 1) Kinect and markers with kinematic constraints (proposed method), 2) Kinect and markers without kinematic constraints, and 3) OptiTrack.

Similar to the Plug-in-Gait marker set, ten markers were placed on the HRP-2 in order to calculate the kinematics of the robot (Fig. 4.6). A Kinect v2 sensor was placed in front of the subject at the height of 75 cm in order to record a gait cycle of the robot in a range of 1.5-2.5 meter from the Kinect camera. (Fig. 4.8). This range was recommended by the results of experiments of chapter 3, in order to increase the accuracy of the marker tracking method using Kinect IR camera. The robot performed a walking motion (speed: 0.2 m/s), while the motion was recorded by the

Kinect (sampling rate: 30 Hz, approximately) and OptiTrack cameras (sampling rate: 120 Hz), simultaneously.

Table 4.1: Pearson correlation (r) between the motion capture systems and humanoid data. r is interpreted as perfect linear correlation ($r = 1$), no linear correlation ($r = 0$), and strong negative correlation ($r = -1$).

	Correlation r (95%CI) OptiTrack	Correlation r (95%CI) Kinect+Markers	Correlation r (95%CI) Kinect+Markers +Constraints
Hip Roll	0.9974	0.9685	0.9967
Hip Pitch	0.9997	0.9361	0.9765
Hip Yaw	0.4389	-0.1868	0.0985
Knee Pitch	0.9997	0.9053	0.9594
Ankle Roll	0.8921	0.9141	-
Ankle Pitch	0.9669	0.9594	-
Average (Hip Yaw excluded)*	**0.9712**	**0.9367**	**0.9775**

Table 4.2: Lin's concordance correlation (CCC) between the motion capture systems and humanoid data. CCC is interpreted as perfect (CCC>0.99), substantial (0.95−0.99), moderate (0.90−0.95), and poor (CCC<0.90).

	Correlation r_c (95%CI) OptiTrack	Correlation r_c (95%CI) Kinect+Markers	Correlation r_c (95%CI) Kinect+Markers +Constraints
Hip Roll	0.9906	0.8110	0.9628
Hip Pitch	0.9997	0.8823	0.9512
Hip Yaw	0.0526	-0.0363	0.0670
Knee Pitch	0.9990	0.8972	0.9554
Ankle Roll	0.8919	0.8926	-
Ankle Pitch	0.9616	0.9027	-
Average (Hip Yaw excluded)*	**0.9686**	**0.8772**	**0.9565**

Table 4.3: Absolute agreement (ICC) between the motion capture systems and humanoid data. ICC is interpreted as excellent (ICC>0.75), good (0.6−0.75), fair (0.4−0.6), and poor (ICC<0.4).

	Agreement ICC (95%CI) OptiTrack	Agreement ICC (95%CI) Kinect+Markers	Agreement ICC (95%CI) Kinect+Markers +Constraints
Hip Roll	0.9906	0.8117	0.9630
Hip Pitch	0.9997	0.8827	0.9514
Hip Yaw	0.0528	-0.0365	0.0673
Knee Pitch	0.9990	0.8976	0.9556
Ankle Roll	0.8923	0.8931	-
Ankle Pitch	0.9618	0.9030	-
Average (Hip Yaw excluded)*	**0.9687**	**0.8776**	**0.9567**

* Since Hip Yaw presents almost zero correlation.

Table 4.4: Consistency (ICC) between the motion capture systems and humanoid data. ICC is interpreted as excellent (ICC>0.75), good (0.6−0.75), fair (0.4−0.6), and poor (ICC<0.4).

	Consistency ICC (95%CI) OptiTrack	Consistency ICC (95%CI) Kinect+Markers	Consistency ICC (95%CI) Kinect+Markers +Constraints
Hip Roll	0.9908	0.8223	0.9650
Hip Pitch	0.9997	0.8829	0.9515
Hip Yaw	0.0543	-0.0365	0.0672
Knee Pitch	0.9995	0.9048	0.9581
Ankle Roll	0.8920	0.9120	-
Ankle Pitch	0.9622	0.9029	-
Average (Hip Yaw excluded)*	**0.9688**	**0.8850**	**0.9582**

The captured joint angles of the humanoid robot are shown in Fig. 4.9. The thin line indicates the angles captured by the Kinect after applying the global

Table 4.5: RMSE between the motion capture systems and humanoid data.

	RMSE (°) OptiTrack	RMSE (°) Kinect+Markers	RMSE (°) Kinect+Markers +Constraints
Hip Roll	0.5178	1.9916	0.9711
Hip Pitch	0.1557	2.5701	1.7248
Hip Yaw	1.3238	0.8619	0.2190
Knee Pitch	0.3068	3.1754	2.0677
Ankle Roll	1.9219	1.8610	-
Ankle Pitch	2.1662	2.8315	-
Average (Hip Yaw excluded)*	**1.0137**	**2.4859**	**1.5879**

* Since Hip Yaw presents almost zero correlation.

optimization method with joint constraints, the dotted line indicates the captured data by Kinect without the joint constraints, the dash-dotted line indicates the OptiTrack data, and the thick line indicates the true value of the HRP-2 joint angles (ground-truth). Figure 4.9 illustrates the capability of the proposed method in tracking the lower body joint angles of the humanoid robot. Moreover, it indicates the advantage of employing the global optimization method with joint constraints on the captured positions of the markers to reconstruct the poses of the lower body segments, which results in accurate joint angles estimation while using skin markers [44].

The statistical analysis results derived from the Kinect, OptiTrack, and the HRP-2 data are shown in Tables 4.1-4.5. Table 4.1 presents the Pearson correlation, Table 4.2 presents Lin's correlation, Table 4.3 presents the Absolute agreement, and Table 4.4 presents the Consistency between the motion capture and robot data (ground truth) (see Appendix for the details of correlation coefficients). The Root-Mean-Square-Error (RMSE) of the different methods are shown in Table 4.5. In each Table, the OptiTrack results indicate the best results (except the Hip Yaw) which attests the high performance of OptiTrack. The correlation coefficients and RMSE of the captured Hip Roll, Hip Pitch, Knee Pitch, Ankle Roll, and Ankle Pitch using Kinect remains in an acceptable range. Furthermore, accuracy of the captured joint angles improved by applying the joint constraints

to the Kinect data. In all systems, the correlation coefficients of Hip Yaw indicates no correlation between the robot and motion capture data. The robot data shows zero changes in the Hip Yaw; however, the motion capture systems detect a slight changes in the Hip Yaw angles (Fig. 4.9). This error could be caused by a slight rotational slip of the robot's standing foot on the carpet during the gait; therefore, although the robot does not perform any Hip Yaw motion, the motion capture systems record a small degree Hip Yaw rotation. An approximate zero correlation between the motion capture systems and robot data was detected for Hip Yaw angles. In order to have a better comparison between different motion capture techniques, the arithmetic average of the results for each system was calculated and demonstrated in Tables 4.1-4.5. Capturing the Hip Yaw was not successful for Kinect and OptiTrack due to the slight rotational slip of the robot's standing foot; hence, the averages were calculated excluding Hip Yaw in the humanoid experiment. The results indicate adequate accuracy for the Kinect method and the best performance for the OptiTrack.

In the humanoid robot test, the comparison between the OptiTrack and the ground truth (robot data) was implemented to assess the accuracy of OptiTrack system in order to use OptiTrack data as the approximate ground truth in the human test (section 5).

4.3.2 Human motion capturing

As shown in Table 4.5, RMSEs between the OptiTrack data and the ground truth are relatively small. Since, there are no ground truth in the human motion capture test, the motion data captured by the OptiTrack are considered as the approximated ground truth in this section. Four motion capturing methods were simultaneously tested; 1) Kinect and markers with kinematic constraints (proposed method), 2) Kinect and markers without kinematic constraints, 3) Kinect skeleton (SDK), and 4) Perception Neuron.

Twenty-eight retroreflective markers were placed on the lower extremity of a subject similar to the conventional Plug-in-Gait marker set. The four markers placed on the back side of the human cannot be detected by the Kinect. These markers were used to complete the human skeleton model in the OptiTrack motion

capture system (six prime 41 cameras) (Fig. 4.2). To compare the results of the proposed method to an IMU based motion capture system, a full body Perception Neuron was placed on the body segments of each subject to record the motion with sampling rate of 120 Hz.

A Kinect v2 sensor was placed in front of the subject at the height of 75 cm in order to record a gait cycle of each subject in a range of 1.5-2.5 meter from the Kinect camera. (Fig. 4.8). This range was reach from the results of the chapter 3 experiments, in order to increase the accuracy of the marker tracking method using Kinect IR camera.

Five healthy adult (male) were selected to perform a normal gait (speed: 0.8 m/s, approximately) toward the Kinect camera while the motion was captured by the Kinect and OptiTrack cameras, simultaneously. The walking test was repeated ten times for each trail.

Table 4.6: Pearson correlation (r) between the low-cost motion capture systems and OptiTrack data. r is interpreted as perfect linear correlation ($r = 1$), no linear correlation ($r = 0$), and strong negative correlation ($r = -1$).

	Correlation r (95%CI) Perception Neuron	Correlation r (95%CI) Kinect SDK	Correlation r (95%CI) Kinect+Markers	Correlation r (95%CI) Kinect+Markers +Constraints
Pelvic Tilt	-0.5268	0.5574	0.1816	0.7675
Pelvic Obliquity	-0.0396	-0.2601	-0.2901	0.5512
Pelvic Rotation	-0.2693	0.2669	0.8833	0.9600
Hip Flex/Ext	0.9056	0.9622	0.9991	0.9996
Hip Ad/Ab	-0.1870	-0.6944	0.3843	0.5942
Hip Rotation	0.0989	-0.7276	0.8066	0.9692
Knee Flex	0.9804	0.9900	0.9998	0.9999
Knee Ad/Ab	0.8698	-	0.9882	0.9983
Knee Rotation	0.9708	-	0.8429	0.9573
Ankle Flex/Ext	0.7166	-0.0201	0.9909	-
Average	**0.3519**	**0.3074**	**0.6787**	**0.8797**

The definition of the lower body joint angles based on the Euler rotation angles of the segment coordinate systems is shown in Fig. 4.10. The average of

Table 4.7: Lin's concordance correlation (CCC) between the low-cost motion capture systems and OptiTrack data. CCC is interpreted as perfect (CCC>0.99), substantial (0.95−0.99), moderate (0.90−0.95), and poor (CCC<0.90).

	Correlation r_c (95%CI) Perception Neuron	Correlation r_c (95%CI) Kinect SDK	Correlation r_c (95%CI) Kinect+Markers	Correlation r_c (95%CI) Kinect+Markers +Constraints
Pelvic Tilt	-0.3359	0.0199	0.0710	0.4145
Pelvic Obliquity	-0.0053	-0.0436	-0.0666	0.1097
Pelvic Rotation	-0.0739	0.0087	0.6572	0.8532
Hip Flex/Ext	0.8657	0.9615	0.9837	0.9927
Hip Ad/Ab	-0.1501	-0.0723	0.2229	0.4450
Hip Rotation	0.0297	-0.1270	0.5237	0.8638
Knee Flex	0.9731	0.8495	0.9988	0.9997
Knee Ad/Ab	0.8650	-	0.8788	0.9747
Knee Rotation	0.5930	-	0.6632	0.8928
Ankle Flex/Ext	0.4425	-0.0181	0.9551	-
Average	**0.3204**	**0.3579**	**0.5888**	**0.7546**

the captured lower body joint angles of ten trials of all five subjects are shown in Figs 4.11 and 4.12. Also, a single trial of a random subject is shown in Figs 4.13 and 4.14. The thin line indicates the angles captured by the Kinect after applying the global optimization method with joint constraints, the dotted line indicates the Kinect captured data without the joint constraints, the dash-dotted line indicates the estimated joint angles using the Kinect SDK (chapter 2), and the thick line indicates the OptiTrack data (approximated ground-truth). Figures 4.11 and 4.13 illustrate the capability of the proposed method in tracking the lower body joint angles. Moreover, it indicates the advantage of employing the global optimization method with joint constraints on the captured positions of the markers to reconstruct the poses of the lower body segments, which results in accurate joint angles estimation while using skin markers [44].

The statistical analysis results derived from the Kinect, Perception Neuron, and OptiTrack data are shown in Tables 4.6-4.10. Table 4.6 presents the Pearson correlation, Table 4.7 presents Lin's correlation, Table 4.8 presents the Absolute

Table 4.8: Absolute agreement (ICC) between the low-cost motion capture systems and OptiTrack data. ICC is interpreted as excellent (ICC>0.75), good (0.6−0.75), fair (0.4−0.6), and poor (ICC<0.4).

	Agreement ICC (95%CI) Perception Neuron	Agreement ICC (95%CI) Kinect SDK	Agreement ICC (95%CI) Kinect+Markers	Agreement ICC (95%CI) Kinect+Markers +Constraints
Pelvic Tilt	-0.3450	0.0203	0.0723	0.4193
Pelvic Obliquity	-0.0054	-0.0445	-0.0681	0.1117
Pelvic Rotation	-0.0755	0.0088	0.6617	0.8557
Hip Flex/Ext	0.8680	0.9623	0.9841	0.9929
Hip Ad/Ab	-0.1535	-0.0738	0.2264	0.4499
Hip Rotation	0.0303	-0.1298	0.5286	0.8661
Knee Flex	0.9736	0.8520	0.9988	0.9997
Knee Ad/Ab	0.8673	-	0.8809	0.9752
Knee Rotation	0.5978	-	0.6676	0.8947
Ankle Flex/Ext	0.4474	-0.0185	0.9559	-
Average	**0.3205**	**0.3577**	**0.5908**	**0.7565**

agreement, and Table 4.9 presents the Consistency between the motion capture and robot data (ground truth). The Root-Mean-Square-Error (RMSE) of the different methods are shown in Table 4.10. The correlation coefficients and RMSE of the lower body joint angles indicate the advantage of the proposed method in capturing lower body joint angles of human over Perception Neuron and Kinect SDK. Furthermore, the captured joint angles' accuracy improved by applying the joint constraints to the Kinect data.

The study goal was to improve and examine the capability of the novel Kinect-based motion capture system in measuring lower extremity kinematics during gait. Our results indicate that the proposed method is an acceptable tool for capturing the pelvic, hip, knee, and ankle joint angles across the gait cycle; however, significant limitations existed when assessing humanoid kinematics, and the method needs improvement. Additionally, the capacity of the Kinect to assess kinematic variables was enhanced compared to previous motion capturing techniques. These findings support the system's ability to evaluate lower extremity gait patterns ob-

Table 4.9: Consistency (ICC) between the low-cost motion capture systems and OptiTrack data. ICC is interpreted as excellent (ICC>0.75), good (0.6−0.75), fair (0.4−0.6), and poor (ICC<0.4).

	Consistency ICC (95%CI) Perception Neuron	Consistency ICC (95%CI) Kinect SDK	Consistency ICC (95%CI) Kinect+Markers	Consistency ICC (95%CI) Kinect+Markers +Constraints
Pelvic Tilt	-0.4580	0.1132	0.1319	0.5846
Pelvic Obliquity	-0.0076	-0.1471	-0.2732	0.4568
Pelvic Rotation	-0.1073	0.2039	0.8804	0.9571
Hip Flex/Ext	0.8904	0.9617	0.9984	0.9993
Hip Ad/Ab	-0.1592	-0.3845	0.3107	0.5582
Hip Rotation	0.0399	-0.2206	0.7407	0.9428
Knee Flex	0.9800	0.8502	0.9988	0.9997
Knee Ad/Ab	0.8697	-	0.8793	0.9748
Knee Rotation	0.6646	-	0.8413	0.9573
Ankle Flex/Ext	0.4506	-0.0200	0.9659	-
Average	**0.3163**	**0.3357**	**0.6474**	**0.8431**

Table 4.10: RMSE between the low-cost motion capture systems and OptiTrack data.

	RMSE (°) Perception Neuron	RMSE (°) Kinect SDK	RMSE (°) Kinect+Markers	RMSE (°) Kinect+Markers +Constraints
Pelvic Tilt	1.6866	2.5628	1.5450	1.0815
Pelvic Obliquity	5.7947	2.9494	1.1894	1.9083
Pelvic Rotation	6.0919	5.2861	0.9149	0.5489
Hip Flex/Ext	4.4675	2.5233	1.6500	1.1025
Hip Ad/Ab	1.9322	7.0425	1.9566	1.2411
Hip Rotation	3.3387	26.0926	2.8100	1.4050
Knee Flex	3.1983	6.0413	0.6503	0.3251
Knee Ad/Ab	2.7713	-	2.1615	1.0808
Knee Rotation	6.7912	-	3.5339	1.7669
Ankle Flex/Ext	6.3234	11.3454	2.1570	-
Average	**4.2396**	**6.5843**	**1.8569**	**1.1460**

jectively; however, the limitations identified in this study must be addressed before it can be widely adopted.

4.4 Summary

Capturing 3D kinematic parameters of human body motion using Kinect SDK was challenging for the Kinect sensor. By tracking the 3D coordinates of the area between two adjacent reflective markers placed on the lower body's anatomical landmarks using Kinect IR camera, 3D kinematics of the lower body was captured during gait. Moreover, estimating the lower body links' optimal pose using the global optimization method with joint constraints improved the motion capturing accuracy.

First, the accuracy of the proposed method was assessed in a humanoid walking test, and the results were compared with the OptiTrack's results (humanoid data as ground truth). The correlation coefficients with 95% confidence intervals (95%CI) indicated excellent intraclass correlations coefficients (ICC>0.75), a substantial concordance correlation (CCC>0.95) between OptiTrack and the humanoid data for all join angles except for Hip Yaw and Ankle Roll. Similar results were achieved using the proposed method; however, the OptiTrack proved to perform slightly better by comparing the arithmetic average of the correlations excluding the Hip Yaw data (OptiTrack's CCC (0.9686)>proposed method's CCC (0.9565)). Moreover, the effectiveness of applying joint constraints to the Kinect data was demonstrated by lightly better average correlations for the Kinect+Markers+Constraints (proposed method) (Table 4.1-4.5). Furthermore, the accuracy of the proposed method was compared with the result of other low-cost motion capture systems (Perception Neuron and Kinect SDK) in capturing the kinematic parameters of human gait (Table 4.6-4.10). In the human gait analysis, the OptiTrack data were used as ground truth since the OptiTrack's capability was proved in the humanoid test. The average of correlation results of the proposed method demonstrated the advantage of proposed method over other low-cost motion capture systems (proposed method's ICC (0.7565)>Kinect SDK's ICC (0.3577)>Perception Neuron's ICC (0.3205)). Although the intraclass correlations coefficients indicated excellent cor-

relations between proposed method and OptiTrack (ICC>0.75), the concordance correlation indicated poor correlation(CCC<0.90). The poor average concordance correlation was caused mainly due to the error in estimating the Pelvic joint angles. The average of the proposed method's RMSE in capturing both human and humanoid joint angles indicated less than 2° error, which is a substantial result for a low-cost motion capture system.

Although the indicated accurate motion capturing by tracking reflective markers using Kinect depth sensor, there are limitations in this method as mentioned in Chapter 3. Since the markers should be placed in a range between 150 cm and 200 cm from the Kinect IR camera in order to minimize the error, the motion was only captured in this range which limits the freedom of the motion. Moreover, the motion needed to be performed slowly due to the lower sampling rate of Kinect (30 Hz) compared to other systems. Importantly, the operators needed to make sure that the markers were not occluded during the motion, resulting in data loss and increases in error. It is also recommended to cover reflective surfaces in the experiment room to reduce IR reflections that affect Kinect's depth sensor. In order to avoid tracking reflective surfaces like the markers directly, it is recommended to use redundant markers and capture the position of the center point of the two adjacent markers.

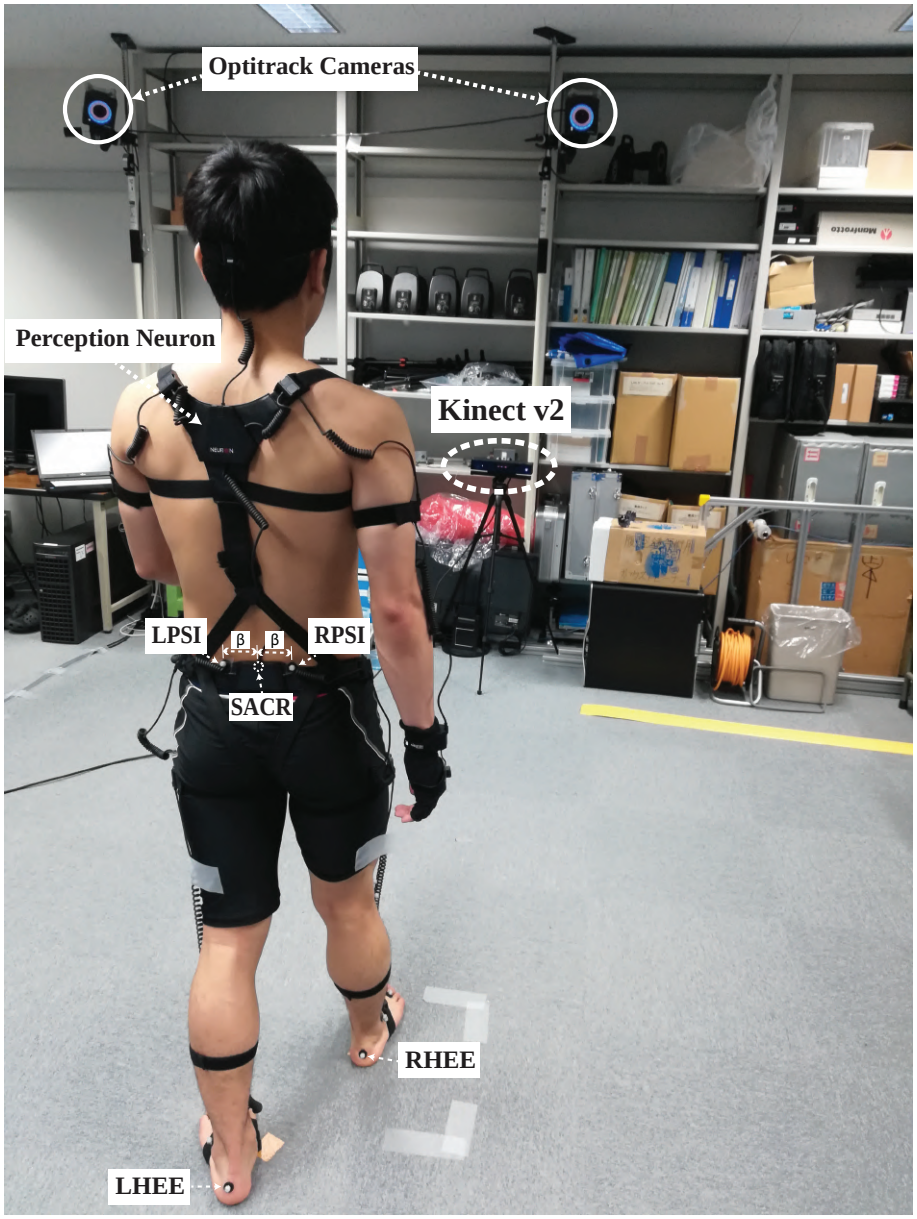


Figure 4.2: A human gait is captured by Kinect, Perception Neuron, and OptiTrack.

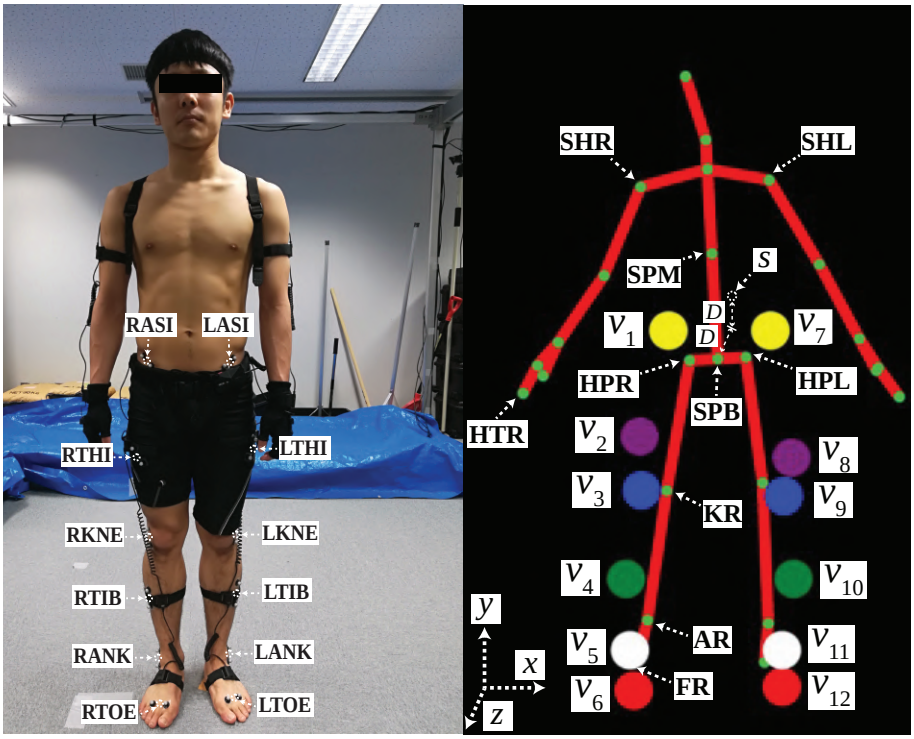


Figure 4.3: Plug-in-Gait lower body marker set with redundant markers (left), and detected virtual markers in specific search domains defined using the Kinect skeleton (right).

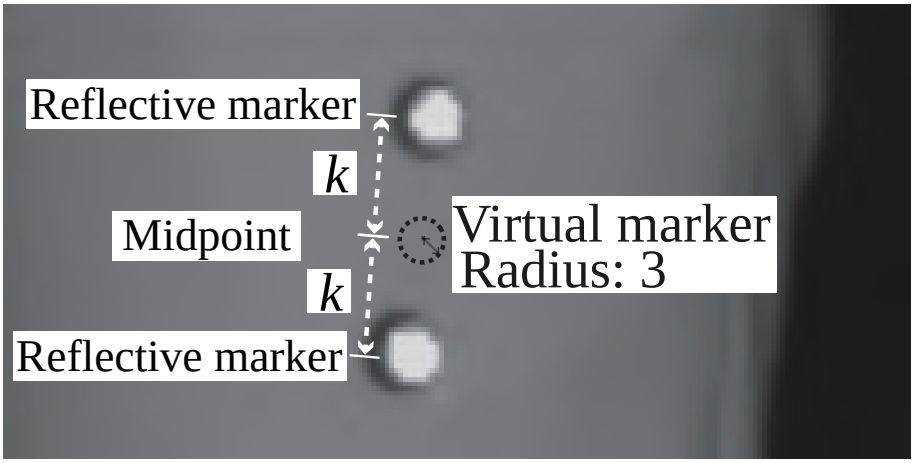


Figure 4.4: A virtual marker with a radius of 3 pixels is estimated in the midpoint of two adjacent reflective markers in a Kinect IR image (this technique was used in the human walking test using redundant markers).

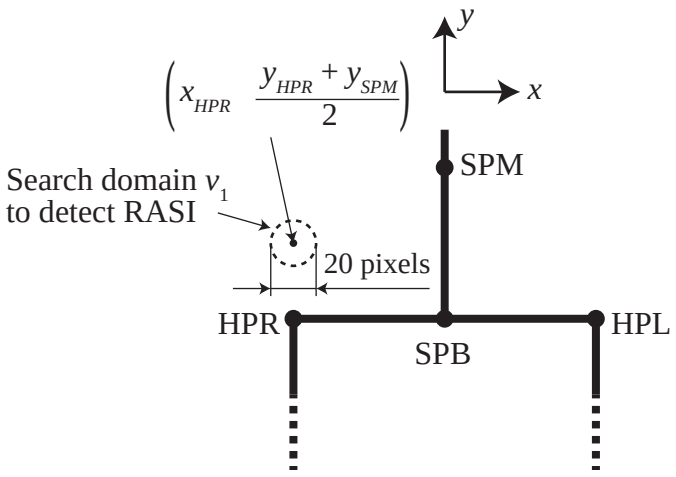


Figure 4.5: Search domain v_1 to detect the landmark RASI.

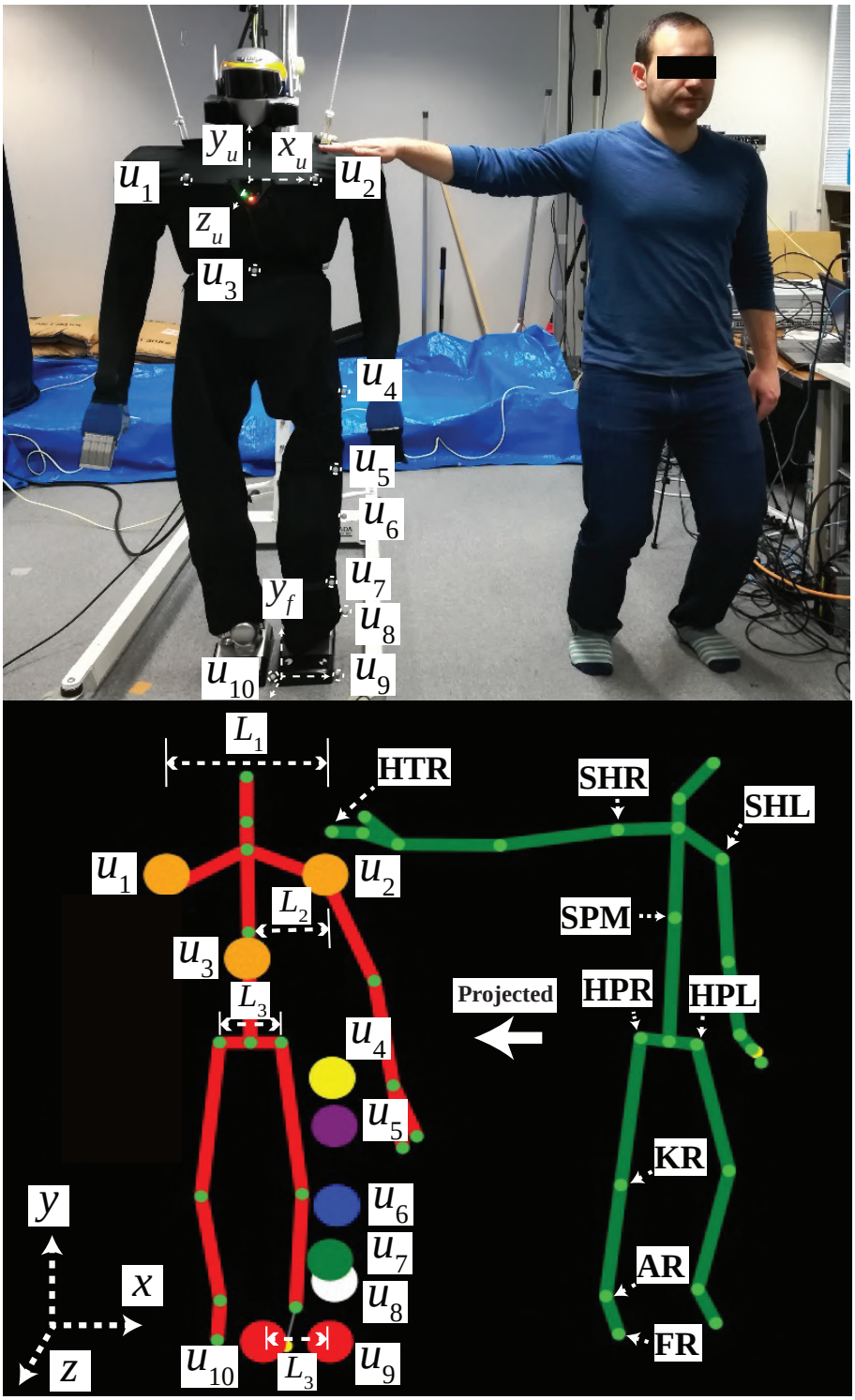


Figure 4.6: Plug-in-Gait marker set placed on the main and lower part of an HRP-2 humanoid robot (up), and estimated markers in specific search domains defined using the human skeleton (down).

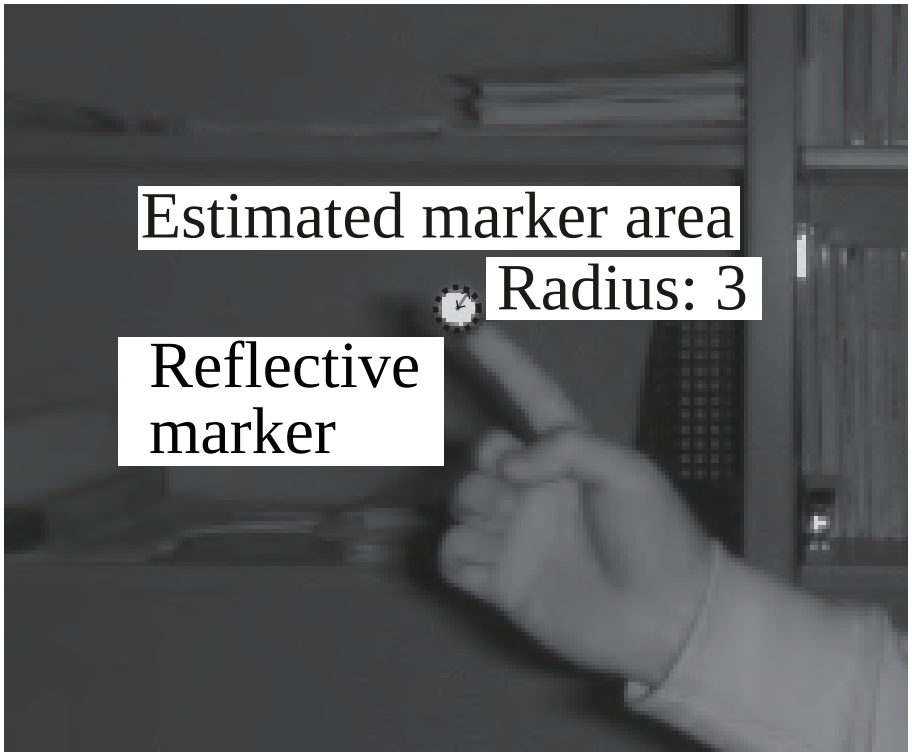


Figure 4.7: An imaginary marker area (blob) with a radius of 3 pixels is estimated around the center of a reflective marker in a Kinect IR image (this technique was used to track the markers placed on the humanoid robot in the robot walking test).

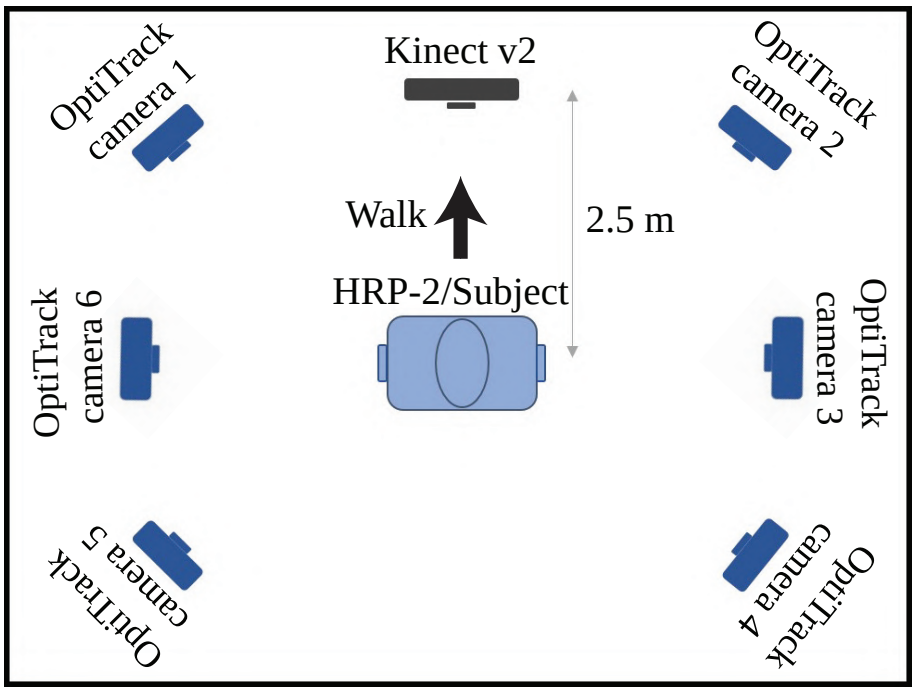


Figure 4.8: Layout of the gait experiment. A Kinect and six OptiTrack cameras were used to capture the gait motion of a humanoid robot (in the robot walking test) and five human subjects (in the human walking test).

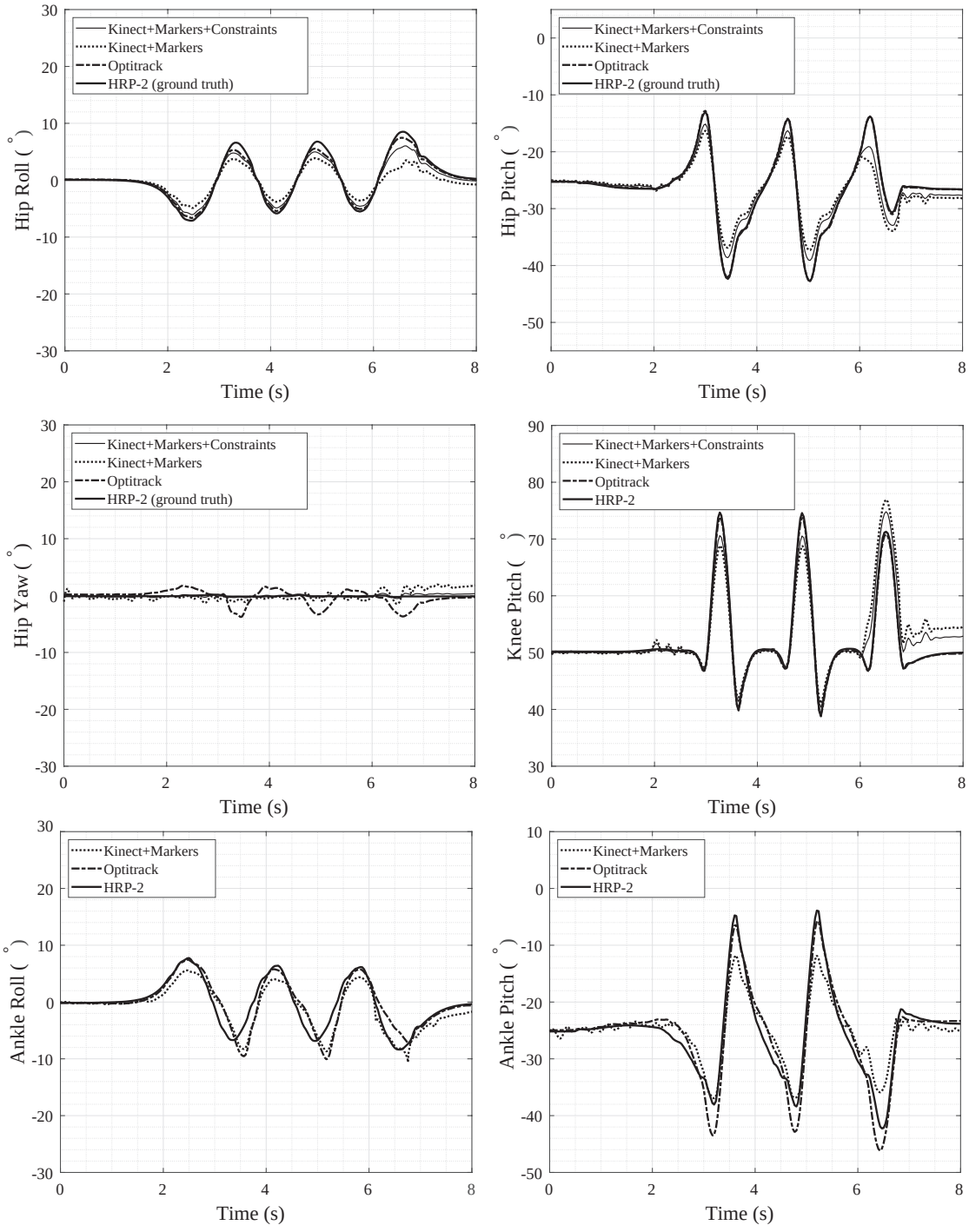


Figure 4.9: Captured joint angles of the humanoid during a walking test.

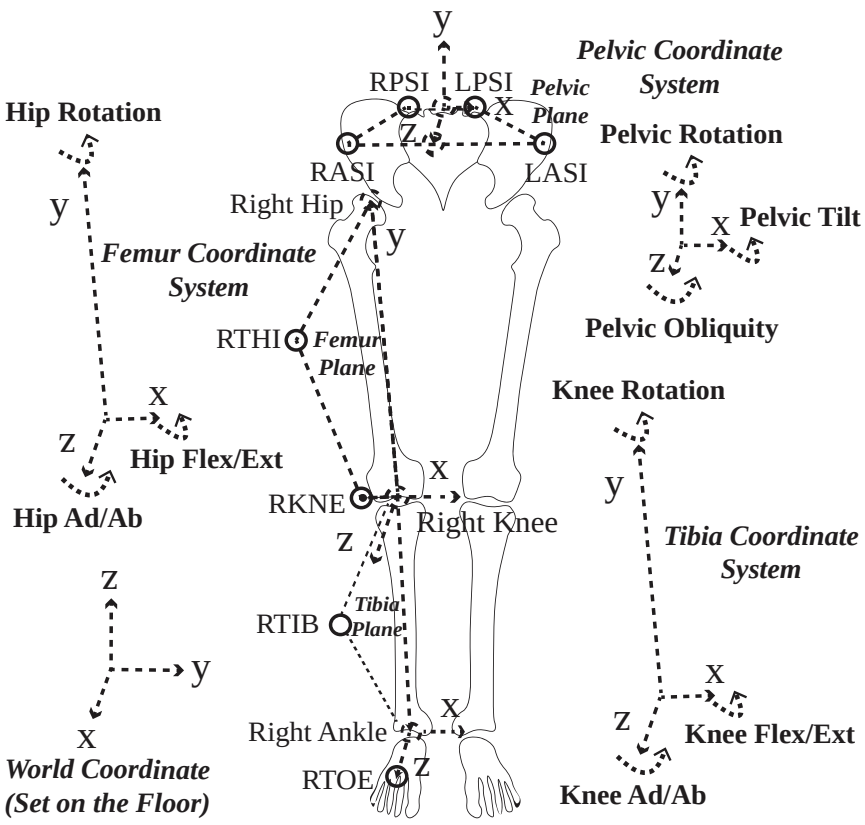


Figure 4.10: Definition of the lower body joint angles using the segments' coordinate systems.

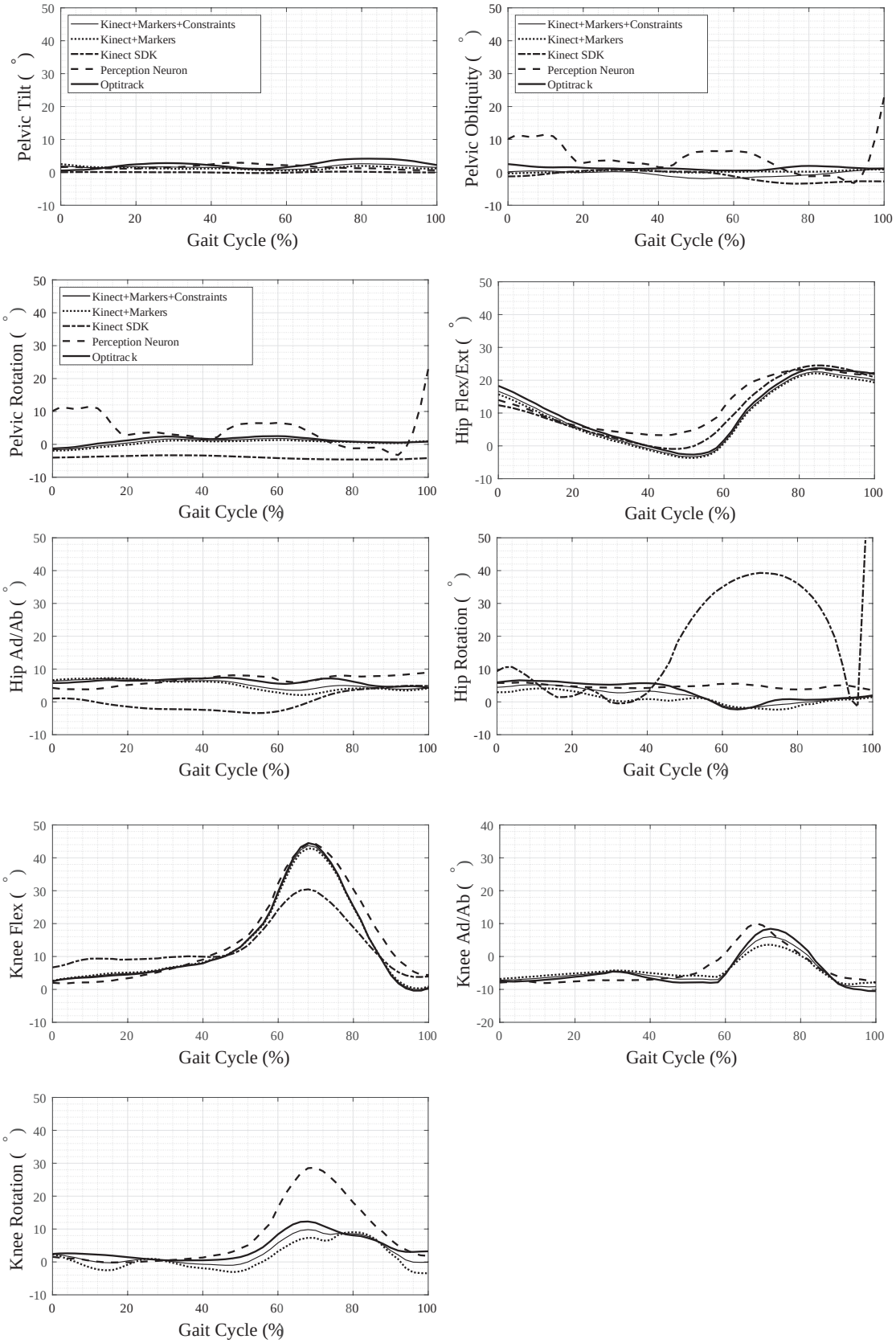


Figure 4.11: Average results of the lower body joint angles of ten trials of the five human subjects.

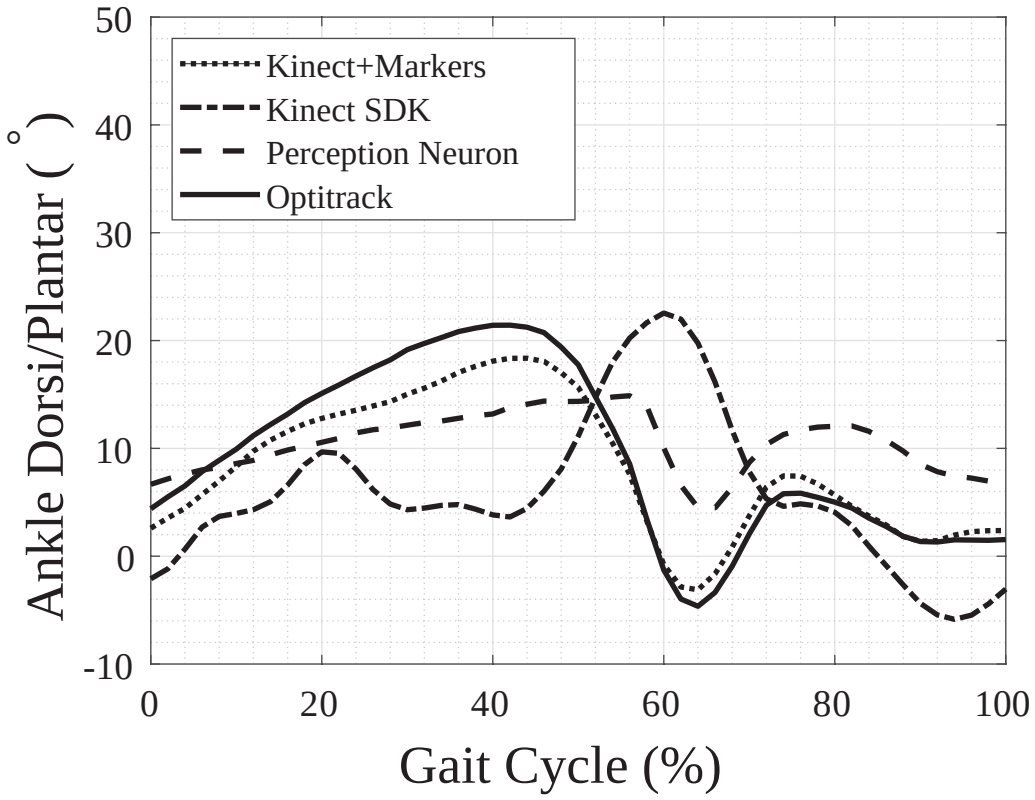


Figure 4.12: Average results of the ankle joint angles of ten trials of the five human subjects.

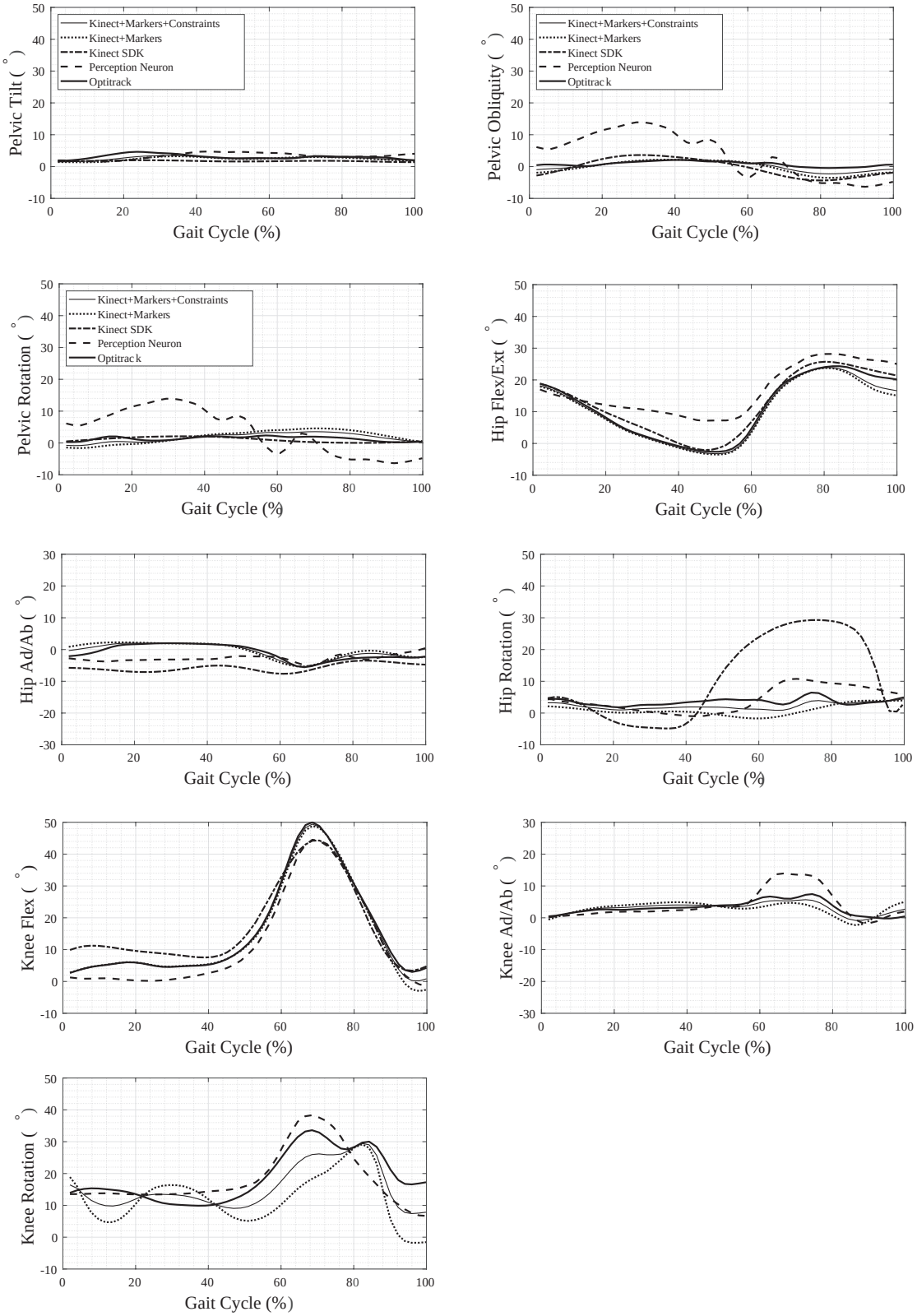


Figure 4.13: Results of the lower body joint angles of a single trial of a random subject.

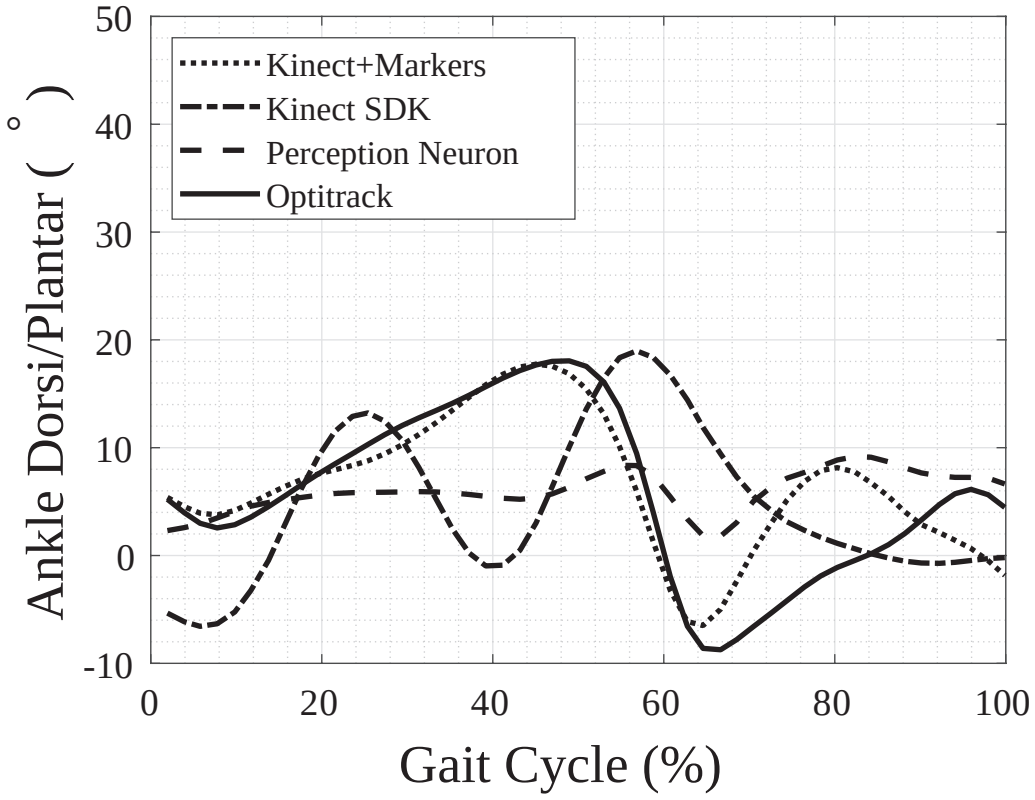


Figure 4.14: Results of the ankle joint angles of a single trial of a random subject.

Chapter 5. Conclusion

5.1 Conclusion

5.1.1 Summary

In this thesis, accuracy assessment of motion capturing using Kinect v2 is studied. In order to capture the lower body joint angles of a human, two methods are described. One is the motion capturing using Kinect skeleton model, and the other is marker tracking using Kinect IR sensor. In chapter 4, the accuracy of both methods was evaluated by capturing the motions of five human subjects, and a humanoid robot during gait experiments.

Markerless motion capturing using Kinect skeleton

The Microsoft Kinect has been implemented as a low-cost motion capture camera in biomedical and robotics applications. Therefore, accuracy assessment of Kinect as a motion capture system is required in order to provide a reference for the future researches. The Kinect skeleton model provided by the Kinect software gives valuable information of the human joint positions. However, calculating kinematic parameters like joint angles needs to be done by processing the skeleton data. Unlike previous studies, we used inverse kinematics to estimate the lower body joint angles of five subjects, and compared the results to a commercial motion capture system (Vicon) as ground-truth. In order to determine the relation of the Kinect's position and the subject, we evaluated the Kinect's accuracy in several positions. Lastly, the correlation between the subjects' movement speed and the accuracy of the Kinect skeleton model in capturing human joint angles was presented.

The results indicated good accuracy of the Kinect skeleton in capturing the knee

flexion/extension. However, estimating lower joint angles using only the Kinect skeleton model consists of significant errors compared to a traditional motion capture system. In these experiments, it is recommended to put the Kinect sensor in front of the subject and it is better the subject performs the motions in low-speed since the Kinect sampling rate is limited to 30 Hz.

Marker-based Motion capturing using Kinect skeleton and depth sensor

In order to improve the motion capturing capability of the Kinect sensor, we introduced a marker tracking method using Kinect IR camera. The Kinect depth data provided more accurate estimation of the landmarks compared to the Kinect skeleton model. Using the Kinect IR camera as a pinhole camera, and combining the depth data to achieve the 3D positions of retroreflective markers helped to determine the markers' positions. The Kinect skeleton model was used to help to detect multiple markers on the subjects in the Kinect IR image and reduced the complexity of the marker tracking.

In order to study the accuracy of the proposed method, the motion capturing capability of the marker-based motion capture system was assessed in several human and humanoid tests. The humanoid data were used as ground-truth, where the accuracy of the proposed method and a commercial motion capture system (OptiTrack) could be evaluated. Statistical analysis of the captured results was done in order to demonstrate the advantage of the proposed method over well-known motion capture systems.

the results indicated the advantage of the proposed method over motion capturing using Kinect skeleton. Moreover, the joint constraints applied to the Kinect data helped to improve the accuracy of the proposed method. The advantage of this method over an IMU system (Perception Neuron) was also presented. In this method, it is recommended to put the Kinect sensor in front of the subject and the motion be performed in a range between 150 cm and 200 cm from the Kinect IR camera. The reflective surfaces needed to be covered so the unnecessary reflections would not affect the Kinect's depth sensor. Finally, The operator needed to make sure markers' occlusion does not occur during the motion capturing.

5.1.2 Future work

Markerless motion capturing using Kinect skeleton

Kinect skeleton model can be used to estimate the knee and hip felxion/extension. However, the results of the accuracy assessment of chapter 2 indicate further improvements in the Kinect's technical specs are needed. Moreover, the Kinect data are noisy and more filtering and data processing researches are required to provide more smooth data using Kinect sensor. In recent studies, an AI-based motion capturing technique proved to replace Kinect in markerless motion capture. However, The AI-based system requires multiple 2D cameras, and the deep learning algorithm needs to be trained over 500,000 annotated image data [7]. Further improvement of such algorithms to enable single-eye cameras to capture complex motions is needed in future researches.

Marker-based Motion capturing using Kinect skeleton and depth sensor

This thesis introduced a marker tracking technique to improve joint angles estimation using a single Kinect. However, some markers placed on the subject could be missed due to occlusion. Using machine learning regression and sequential data prediction algorithms to solve the occlusion issue could help achieve more accurate results. Moreover, the improvement of estimating the Pelvic plane using the detected markers and accuracy assessment of the proposed method in capturing whole-body motions of various subjects remain as future work. Importantly, the Kinect camera needed to be placed in front of the motion and the markers placed in a range that minimizes the 3D position estimation of the markers. With the improvement of Kinect specs in the future, these limitations would improve (such as Azure Kinect DK).

Acknowledgements

The development and accuracy analysis of a human motion capturing method using a single IR camera was conducted in intelligent robots and system laboratory of Hokkaido university graduate school of information science and technology. First, I would like to thank my adviser, Professor Konno, for his valuable and kind supports. I nourished scientific approaches and ideas for research. He also supported me with not only research but also mental encouragement. I would like to thank Professor Takayuki Tanaka and Professor Hajime Igarashi for their advices about my research.

I want to thank Dr. Shunsuke Komizunai and Prof. Tsujita for their continued supports during my Ph.D. I am greatly thankful to all the members of intelligent robots and systems laboratory. I also would like to thank Professor Behzadipour for the support and useful advice on human motion capture.

I am also profoundly grateful to the members of the humanoid research group, Dr. Yaung, Mr. Ogawa, Mr. Hamada, Mr. Saito, Mr. Tsukimoto, Mr. Tanaka, Ms. Inoue, and Mr. Sawanobori, for their discussions and support about my research. Especially I am grateful to Mr. Ogawa and Mr. Saito for the robot programming support.

Besides, I am also profoundly grateful to the Hokkaido University graduate school of information science and technology as a research assistant.

At last, I would like to thank my parents for their encouragement and support during my whole life. Especially, I would like to thank my brother, who always believed in me, and my best friend Amin Arefi for his non-stopping supports. Without their encouragement, this research would not have been accomplished.

References

[1] H. Zhou and H. Hu, "Human motion tracking for rehabilitation: A survey," *Biomedical signal processing and control*, 2008.

[2] K. Miura, M. Morisawa, F. Kanehiro, S. Kajita, K. Kaneko, and K. Yokoi, "Human-like walking with toe supporting for humanoids," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4428–4435, 2011.

[3] S. Kagami, M. Mochimaru, Y. Ehara, N. Miyata, K. Nishiwaki, T. Kanade, and H. Inoue, "Measurement and comparison of human and humanoid walking," in *Proceedings 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation. Computational Intelligence in Robotics and Automation for the New Millennium (Cat. No. 03EX694)*, vol. 2, pp. 918–922, 2003.

[4] J. Lee and K. H. Lee, "Precomputing avatar behavior from human motion data," *Graphical models*, vol. 68, no. 2, pp. 158–174, 2006.

[5] S. Calinon, F. D'halluin, E. L. Sauser, D. G. Caldwell, and A. G. Billard, "Learning and reproduction of gestures by imitation," *IEEE Robotics & Automation Magazine*, vol. 17, no. 2, pp. 44–54, 2010.

[6] J. G. Richards, "The measurement of human motion: A comparison of commercially available systems," *Human movement science*, vol. 18, no. 5, pp. 589–602, 1999.

[7] R. M. Kanko, E. K. Laende, E. M. Davis, W. S. Selbie, and K. J. Deluzio, "Concurrent assessment of gait kinematics using marker-based and markerless motion capture," *Journal of biomechanics*, vol. 127, p. 110665, 2021.

[8] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *CVPR 2011*, pp. 1297–1304, 2011.

[9] S. Zennaro, M. Munaro, S. Milani, P. Zanuttigh, A. Bernardi, S. Ghidoni, and E. Menegatti, "Performance evaluation of the 1st and 2nd generation kinect for multimedia applications," in *2015 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2015.

[10] M. Jebeli, A. Bilesan, and A. Arshi, "A study on validating kinect v2 in comparison of vicon system as a motion capture system for using in health engineering in industry," *Nonlinear Engineering*, vol. 6, no. 2, pp. 95–99, 2017.

[11] X. Xu and R. W. McGorry, "The validity of the first and second generation microsoft kinect for identifying joint center locations during static postures," *Applied ergonomics*, vol. 49, pp. 47–54, 2015.

[12] T. M. Guess, S. Razu, A. Jahandar, M. Skubic, and Z. Huo, "Comparison of 3d joint angles measured with the kinect 2.0 skeletal tracker versus a marker-based motion capture system," *Journal of applied biomechanics*, vol. 33, no. 2, pp. 176–181, 2017.

[13] M. do Carmo Vilas-Boas, H. M. P. Choupina, A. P. Rocha, J. M. Fernandes, and J. P. S. Cunha, "Full-body motion assessment: Concurrent validation of two body tracking depth sensors versus a gold standard system during gait," *Journal of biomechanics*, vol. 87, pp. 189–196, 2019.

[14] M. Eltoukhy, J. Oh, C. Kuenze, and J. Signorile, "Improved kinect-based spatiotemporal and kinematic treadmill gait assessment," *Gait & posture*, vol. 51, pp. 77–83, 2017.

[15] M. Kharazi, A. Memari, A. Shahrokhi, H. Nabavi, S. Khorami, A. Rasooli, H. Barnamei, A. Jamshidian, and M. Mirbagheri, "Validity of microsoft kinecttm for measuring gait parameters," in *2015 22nd Iranian Conference on Biomedical Engineering (ICBME)*, pp. 375–379, 2015.

[16] Z. Jamali and S. Behzadipour, "Quantitative evaluation of parameters affecting the accuracy of microsoft kinect in gait analysis," in *2016 23rd Iranian Conference on Biomedical Engineering and 2016 1st International Iranian Conference on Biomedical Engineering (ICBME)*, pp. 306–311, 2016.

[17] S. Springer and G. Yogev Seligmann, "Validity of the kinect for gait assessment: A focused review," *Sensors*, vol. 16, no. 2, p. 194, 2016.

[18] P. N. Pathirana, S. Li, H. M. Trinh, and A. Seneviratne, "Robust real-time bio-kinematic movement tracking using multiple kinects for tele-rehabilitation," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 3, pp. 1822–1833, 2015.

[19] 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, p. e0139913, 2015.

[20] S. Moon, Y. Park, D. W. Ko, and I. H. Suh, "Multiple kinect sensor fusion for human skeleton tracking using kalman filtering," *International Journal of Advanced Robotic Systems*, vol. 13, no. 2, p. 65, 2016.

[21] F. Destelle, A. Ahmadi, N. E. O'Connor, K. Moran, A. Chatzitofis, D. Zarpalas, and P. Daras, "Low-cost accurate skeleton tracking based on fusion of kinect and wearable inertial sensors," in *2014 22nd European Signal Processing Conference (EUSIPCO)*, pp. 371–375, 2014.

[22] A. Timmi, G. Coates, K. Fortin, D. Ackland, A. L. Bryant, I. Gordon, and P. Pivonka, "Accuracy of a novel marker tracking approach based on the low-cost microsoft kinect v2 sensor," *Medical engineering & physics*, vol. 59, pp. 63–69, 2018.

[23] N. S. Pollard, J. K. Hodgins, M. J. Riley, and C. G. Atkeson, "Adapting human motion for the control of a humanoid robot," in *Proceedings 2002 IEEE international conference on robotics and automation (Cat. No. 02CH37292)*, vol. 2, pp. 1390–1397, 2002.

[24] S. Nakaoka, A. Nakazawa, F. Kanehiro, K. Kaneko, M. Morisawa, and K. Ikeuchi, "Task model of lower body motion for a biped humanoid robot to imitate human dances," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3157–3162, 2005.

[25] S. Schaal, A. Ijspeert, and A. Billard, "Computational approaches to motor learning by imitation," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 358, no. 1431, pp. 537–547, 2003.

[26] L. Mündermann, S. Corazza, and T. P. Andriacchi, "The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications," *Journal of neuroengineering and rehabilitation*, vol. 3, no. 1, pp. 1–11, 2006.

[27] H. J. Luinge and P. H. Veltink, "Measuring orientation of human body segments using miniature gyroscopes and accelerometers," *Medical and Biological Engineering and computing*, vol. 43, no. 2, pp. 273–282, 2005.

[28] J. Sell and P. O'Connor, "The xbox one system on a chip and kinect sensor," *IEEE Micro*, vol. 34, no. 2, pp. 44–53, 2014.

[29] S. Zennaro, "Evaluation of microsoft kinect 360 and microsoft kinect one for robotics and computer vision applications," 2014.

[30] J.-H. Lee *et al.*, "Full-body imitation of human motions with kinect and heterogeneous kinematic structure of humanoid robot," in *2012 IEEE/SICE International Symposium on System Integration (SII)*, pp. 93–98, 2012.

[31] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: A review," *IEEE transactions on cybernetics*, vol. 43, no. 5, pp. 1318–1334, 2013.

[32] A. Pfister, A. M. West, S. Bronner, and J. A. Noah, "Comparative abilities of microsoft kinect and vicon 3d motion capture for gait analysis," *Journal of medical engineering & technology*, vol. 38, no. 5, pp. 274–280, 2014.

[33] M. Eltoukhy, C. Kuenze, J. Oh, S. Wooten, and J. Signorile, "Kinect-based assessment of lower limb kinematics and dynamic postural control during the star excursion balance test," *Gait & posture*, vol. 58, pp. 421–427, 2017.

[34] S. Bronner, "Differences in segmental coordination and postural control in a multi-joint dance movement: developpe arabesque," *Journal of Dance Medicine & Science*, vol. 16, no. 1, pp. 26–35, 2012.

[35] G. Wu and P. R. Cavanagh, "Isb recommendations for standardization in the reporting of kinematic data," *Journal of biomechanics*, vol. 28, no. 10, pp. 1257–1261, 1995.

[36] X. Xu, R. W. McGorry, L.-S. Chou, J.-h. Lin, and C.-c. Chang, "Accuracy of the microsoft kinect for measuring gait parameters during treadmill walking," *Gait & posture*, vol. 42, no. 2, pp. 145–151, 2015.

[37] C. D. Metcalf, R. Robinson, A. J. Malpass, T. P. Bogle, T. A. Dell, C. Harris, and S. H. Demain, "Markerless motion capture and measurement of hand kinematics: validation and application to home-based upper limb rehabilitation," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2184–2192, 2013.

[38] A. Mobini, S. Behzadipour, and M. Saadat Foumani, "Accuracy of kinect skeleton tracking for upper body rehabilitation applications," *Disability and Rehabilitation: Assistive Technology*, vol. 9, no. 4, pp. 344–352, 2014.

[39] A. Schmitz, M. Ye, R. Shapiro, R. Yang, and B. Noehren, "Accuracy and repeatability of joint angles measured using a single camera markerless motion capture system," *Journal of biomechanics*, vol. 47, no. 2, pp. 587–591, 2014.

[40] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[41] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Proceedings of the 1997 IEEE Computer Soci-*

ety Conference on Computer Vision and Pattern Recognition, pp. 1106–1112, 1997.

[42] J. Bouguet, "Matlab calibration toolbox," *URL: http://www. vision. caltech. edu/bouguetj/calib_doc*, 2010.

[43] D. Pagliari and L. Pinto, "Calibration of kinect for xbox one and comparison between the two generations of microsoft sensors," *Sensors*, vol. 15, no. 11, pp. 27 569–27 589, 2015.

[44] T.-W. Lu and J. O'Connor, "Bone position estimation from skin marker coordinates using global optimisation with joint constraints," *Journal of biomechanics*, vol. 32, no. 2, pp. 129–134, 1999.

[45] M. P. Kadaba, H. Ramakrishnan, and M. Wootten, "Measurement of lower extremity kinematics during level walking," *Journal of orthopaedic research*, vol. 8, no. 3, pp. 383–392, 1990.

[46] M. Naeemabadi, B. Dinesen, O. K. Andersen, and J. Hansen, "Influence of a marker-based motion capture system on the performance of microsoft kinect v2 skeleton algorithm," *IEEE Sensors Journal*, vol. 19, no. 1, pp. 171–179, 2018.

[47] M. Dunn, D. Pavan, P. Ramirez, S. Rava, and A. Sharin, "An automated method to extract three-dimensional position data using an infrared time-of-flight camera," in *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 2, no. 6, p. 502, 2018.

[48] K. Kaneko, F. Kanehiro, and S. Kajita, "H.(2004) humanoid robot hrp-2," *IEEE/RSJ Int. Con. on Robotics and Automation, USA*, pp. 1083–1090.

[49] R. B. Davis III, S. Ounpuu, D. Tyburski, and J. R. Gage, "A gait analysis data collection and reduction technique," *Human movement science*, vol. 10, no. 5, pp. 575–587, 1991.

[50] I. Lawrence and K. Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics*, pp. 255–268, 1989.

[51] J. J. Bartko, "The intraclass correlation coefficient as a measure of reliability," *Psychological reports*, vol. 19, no. 1, pp. 3–11, 1966.

[52] C.-C. Chen and H. X. Barnhart, "Comparison of icc and ccc for assessing agreement for data without and with replications," *Computational Statistics & Data Analysis*, vol. 53, no. 2, pp. 554–564, 2008.

List of Publications

Journal Publications

[1] **A. Bilesan**, S. Komizunai, T. Tsujita, A. Konno, "Improved 3D Human Motion Capture Using Kinect Skeleton and Depth Sensor," *Journal of Robotics and Mechatronics*, vol. 33, no. 6, 2021.

International Conferences (Full Review)

[1] **A. Bilesan**, M. Owlia, S. Behzadipour, S. Ogawa, T. Tsujita, S. Komizunai and A. Konno, "Marker-based motion tracking using Microsoft Kinect," in *Proceedings of the 12th IFAC Symposium on Robot Control (SYROCO)*, 2018, Budapest Hungary.

[2] **A. Bilesan**, S. Behzadipour, T. Tsujita, S. Komizunai and A. Konno, "Markerless Human Motion Tracking Using Microsoft Kinect SDK and Inverse Kinematics," in *Proceedings of the 12th Asian Control Conference (ASCC)*, 2019, Kitakyushu Japan.

Appendix A Statistical analysis

In previous researches, the Pearson correlation coefficient (r) was used to assess the correlation between two observers' results. However, this index only measures the linear relationship between the continuous observed data, and it does not take agreement into account [50].

Lin's concordance correlation coefficient (CCC) (also known as r_c) was introduced to assess the agreement without analysis of variance (ANOVA) assumptions [50]. Intraclass correlation coefficient (ICC) has been used for evaluating agreement and reliability between multiple observers during a data collecting task [51]. The ICC coefficient is widely used to indicate whether a new observer (motion capture system) can reproduce the same results as an existing gold-standard system in clinical researches. Extensions of ICC lead to other types of ICCs based on ANOVA models [52]. However, choosing a specific ICC in a study without considering the ANOVA assumptions may lead to biased correlation results. Chen et al. recommended using ICC_3 to evaluate the agreement between different observers, which is based on two-way ANOVA model with interaction between observer and subject [52].

The two-way ANOVA model with interaction is defined as

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \tag{A.1}$$

where X_{ij} indicates the kth replicated reading for the ith joint angle data captured by the jth motion capture system ($i = 1, ..., N$, $j = 1, ..., J$, $k = 1, ..., K$), μ indicates the mean of the joint angles, α_i indicates the random effect of the ith joint angle data (independent and identically distributed (*i.i.d.*) with normal distribution $N(0,\sigma_\alpha^2)$), β_j indicates the random effect of motion capture system j (*i.i.d.* with $N(0,\sigma_\beta^2)$), γ_{ij} indicates the random effect of the observer-data interaction (*i.i.d.* with $N(0,\sigma_\gamma^2)$), and e_{ij} indicates the random error (*i.i.d.* with $N(0,\sigma_e^2)$). Since the motion capture systems are unique and they are widely used in other re-

searches too, according to two-way ANOVA model, β_j is assumed as fixed (two-way mixed model); therefore, $\sum_{j=1}^{J}\beta_j$ and $\sum_{j=1}^{J}\gamma_j = 0$, while $\sigma_\beta^2 = \sum_{j=1}^{J}\ \beta_j^2/(J-1)$. The ICC_3 with fixed observers and its estimator ($\widehat{ICC_3}$) without replication ($k = 1$) is given by

$$ICC_3 = \frac{\sigma_\alpha^2 - \sigma_\gamma^2/(J-1)}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2 + \sigma_e^2}, \tag{A.2}$$

$$\widehat{ICC_3} = \frac{MS_\alpha - MS_\gamma}{MS_\alpha + MS_\gamma + J(MS_\beta - MS_\gamma)/N}, \tag{A.3}$$

where $MS_\alpha = \frac{JK}{N-1}\sum_{i=1}^{N}(\bar{X}_{i..} - \bar{X}_{...})^2$ is the means of the sums of squares from the two-way ANOVA model for between subjects (joint angle data), $MS_\beta = \frac{NK}{J-1}\sum_{j=1}^{J}(\bar{X}_{.j.} - \bar{X}_{...})^2$ is the means of the sums of squares for between observers (motion capture systems), and $MS_\gamma = \frac{K}{(J-1)(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{J}(\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...})^2$ is the means of the sums of squares for observer-subject interaction.

Since each motion capture system has its own sampling rate, we applied a spline fitting over the joint angles data and extracted the data on specific times ($N = 51$ and $N = 235$ in the human and humanoid test, respectively). Also, since we compared the results of each cost-effective system to reference system, $J = 2$ in Eq. (A.1), and to assess the reliability of each system individually, ICC_3 without replication ($K = 1$) was chosen.

If the systematic differences are not considered in the correlation assessment, then ICC for consistency is the main coefficient to study the correlation between data. ICC_3c which is known as ICC for consistency is similar to ICC_3. However, σ_β^2) is omitted from the denominator and the result would be greater than ICC_3 which is known as ICC for agreement.

Lin's concordance correlation coefficient (CCC) which is also a measurement of agreement has been developed and used in many researches. this coefficient takes both precision (degree of scatter) and accuracy (scale shift) into consideration. CCC is specially used to evaluate correlation between paired observers ($J = 2$) for data without replications ($K = 1$).

Absolute agreement and consistency between the motion capture systems and the ground truth were assessed using ICC_3 (with 95% confidence intervals (95%CI)) as the average captured data of the subjects in human test, and one trial data in the humanoid test were used in the statistical analysis [51].