Tracking People with Multiple Kinects

University of
St Andrews

CS4099: Major Software Project

Author:
Chi-Jui Wu

Supervisor:
Dr. David
Harris-Birtill

April 7, 2015

Abstract

The current work is about tracking people with multiple Kinects. The goal is to reliably track people in uncertain, occluded real-world environments. The final submission contains an interactive application that demonstrates the tracking system as well as a series of user studies reporting the success of the system. The strengths and limitations of the system are discussed. An outline of future work to make the current system deployable in the real life is described.

I declare that the work submitted is my own unless otherwise stated.

The report is XX,XXX words long.

I retain the copyright of this work.

Contents

List of Figures

CHAPTER 1

Introduction

People detection and tracking in realtime are essential in surveillance, interactive systems, medical imaging, and humanoid robotics.

The current project proposes an algorithm for tracking people with multiple Kinects which resolves the problem of occlusion.

1.1 Problem Statement

The task of detecting and tracking moving targets is non-trivial. There are many sources of tracking errors, including raw sensor data noise, illumination levels, changing backgrounds, and occlusion. Real-world environments are unpredictive and complex, thus making the task much harder. The system attempts to solve the problme of occlusion with multiple Kinects.

Occlusion occurs when the tracked target is masked by other objects in the scene. The masked target would not exist in the field of view of one or more cameras. If a person were occluded, his precise joint positions and movements would be unknown. Resolving the problem of occlusion would provide any tracking system with more spatial and physiological information about the tracked people.

There are two types of occlusions: static and dynamic. They are defined as:

Static occlusion Occlusion caused by stationary objects in the environment

Dynamic occlusion Occlusion caused by people interactions in the environment

The project aims to resolve both types of occlusion.

A simple instance of the problem is illustrated in Figure 1.1. In the figure, both skeletons are invisible to the front Kinect but visible to the side Kinect. They are occluded by the red obstacle. When they step out of the obstacle into the views of both Kinects, the system should merge the skeletons of the same person from different perspectives. The main objective of the project is to avoid occlusion by extending the field of view of the system. The proposed algorithm would combine depth sensor information from multiple Kinects to achieve this goal.

Figure 1.1: The occlusion problem

1.2 Contributions

The contributions of the current work are...

1. Replicate, validate, and extend current research
2. A Kinect BodyFrame serialization library
3. A Kinect client-server framework
4. Track people with multiple Kinects
5. Display tracking skeletons from different Kinects fields of view
6. Integrate joints information from multiple Kinects to resolve the occlusion problem
7. User studies showing the strengths and weaknesses of the current system

CHAPTER 2

Background

CHAPTER 3

Objectives

3.1 Primary

3.2 Secondary

CHAPTER 4

Current Approach

4.1 Overview

4.2 Calibration

CHAPTER 5

Design

CHAPTER 6

Implementation

CHAPTER 7

Testing

7.1 Tracking

7.2 Occlusion

CHAPTER 8

Studies

8.1 Motivation

A series of user studies are designed to evaluate the system's accuracy at tracking people in different scenarios. The accuracy of the tracking algorithm, or essentially the coordinate transformation algorithm, is measured by the differences in the joint coordinate between multiple potential skeletons of the same person. The studies will require participants to move around in front of multiple Kinects alone and with other participants. The software will log participants' positions from tracking, and these data will provide a quantitative feedback on the accuracy of the algorithm in different Kinect configurations and user scenarios.

To reiterate, a potential skeleton is a skeleton from a single Kinect field of view. One person may have multiple potential skeletons when they are visible to many Kinects. The application is most useful for its ability to transform any potential skeleton into any Kinect's camera space. The potential skeleton in the current Kinect field of view would be unaffected, but the other potential skeletons that were in other Kinects fields of view would have slight deviations in their joint coordinates. The user studies attempt to capture such deviations in all possible cases.

8.2 Hypotheses

CHAPTER 9

Results

This chapter summarizes the results in each user study and makes comparisons between a number of different studies. The results should provide insights into the reliability and accuracy of the tracking algorithm in different settings. In additiona, it would show the coordinates and joints that yield the most stable results during normal human activities.

9.1 Accessing the data

The complete dataset and plots are publicly available at `https://github.com/cjw-charleswu/KinectMultiTrackDataset`

9.2 Analysis

The analysis was done in Matlab 2015a. The analysis scripts are available at `https://github.com/cjw-charleswu/KinectMultiTrack/tree/master/Analysis`.

9.2.1 Cleaning the data

The logging data are post-processed for ease of plotting the results. The initial logging data, for all aforementioned studies, contain 244,527 rows and 87 columns in total. The final data for evaluation contain 243,550 rows and 87 columns. 977 rows are deleted for various reasons documented below.

There is also an error in code where a scenario id is logged incorrectly. In the second part of scenario 8, when the second participant is asked to walk around the first participant, the scenario id is falsely written as 4. This logging error is corrected by replacing all occurrences of scenario id 4 that are immediately after scenario id 8 and before scenario id 5, which is the next task in line for the participants.

The tracker time is stored as the server's current time in milliseconds. The times for each user task (scenario) are reset. The current timestamps refer to the amount of time passed in each scenario, for a particular Kinect configuration and experiment.

The timesalso converted from milliseconds to seconds. The joint coordinates are converted from meters to centimeters.

The studies are interested in the amount of coordinate differences between multiple skeletons after transformation during tracking. Thus, evaluation requires the joint positions of more than one skeletons (from different Kinect fields of view) at any given timestamp.

9.3 Definitions

Δx, Δy, Δz, Δd, Avg., and Std. are defined as:

Δx The distance, or difference, between the x coordinates of a joint of multiple skeletons representing the same person from different Kinects fields of view, expressed in centimeters.

Δy The distance, or difference, between the y coordinates of a joint of multiple skeletons representing the same person from different Kinects fields of view, expressed in centimeters.

Δz The distance, or difference, between the z coordinates of a joint of multiple skeletons representing the same person from different Kinects fields of view, expressed in centimeters.

Δd The distance, or difference, between the x, y, and z coordinates of a joint of multiple skeletons representing the same person from different Kinects fields of view, expressed in centimeters.

Avg. Average (mean)

Std. Standard deviation

These values quantify the amount of differences produced by the tracking algorithm when transforming multiple skeletons of the same person to a single Kinect field of view.

Coordinates and joints distances are defined as:

Coordinates distances The Δx, Δy, Δz, and Δd distances averaged over a person's entire set of joints, expressed in centimeters.

Joints distances The Δx, Δy, Δz, and Δd distances for each of a person's joints, expressed in centimeters.

9.4 Structure

There are three main results sections, one for each of the primary tasks, namely the Stationary, Steps, and Walk tasks. Each of these sections contains three basic visualizations showing the effects of the particular scenario on coordinates transformation errors when tracked with Kinects at different locations.

Each main section will begin with a figure showing changes in the average coordinates and joints distances with Parallel, 45° and 90° apart Kinects. The x asis will be the position of Kinect placement. The y axis will be the average coordinates distances, in terms of Δx, Δy, Δz, and Δd (as defined in 9.3). The figures will also entail the average coordinates distances over all of the Kinect placements.

The second figure will give more details about the average case shown in the previous figure. It will show changes in the average coordinates distances for each of the joint types. The x asis will be the joint type. The y axis will be the average joints distances, also in terms of Δx, Δy, Δz, and Δd.

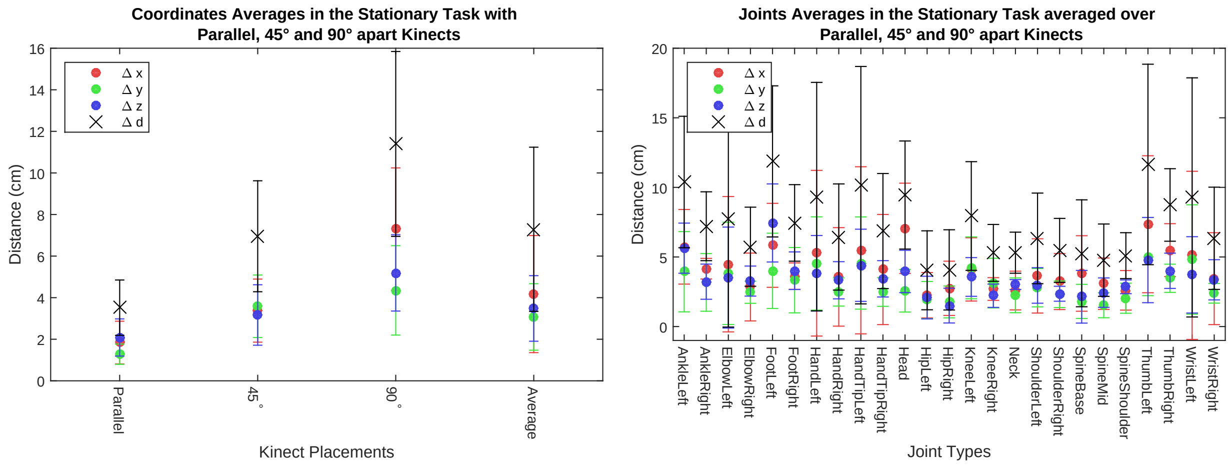
Then, each section will finish with a table summarizing the actual values of the average coordinates distances for the presented task with each of the Kinect placements, as well as averaged over all different placements. The figures show how Kinect placements affect the average coordinates and joints distances during the Steps task.

At last, the results obtained from different scenarios and Kinect placements are compared. Figures will show the average coordinates and joints differences over time.

9.5 Stationary

This section reports average coordinates and joints distances in the Stationary task with parallel, 45° and 90° apart Kinects.

The Δd distance in the Stationary task with parallel Kinects is 3.52. The Δd distance in the Stationary task with 45° apart Kinects is 3.95. The Δd distance in the Stationary task with 90° apart Kinects is 11.39. The average Δd distance in the Stationary task over all Kinect placements is 7.9, with a standard deviation of 3.95.



(a) Average coordinates distances in the Stationary task with Parallel, 45° and 90° apart Kinects.

(b) Average joints distances in the Stationary task averaged over Parallel, 45° and 90° apart Kinects.

Figure 9.1: Plots showing the results in the Stationary task with different Kinect placements.

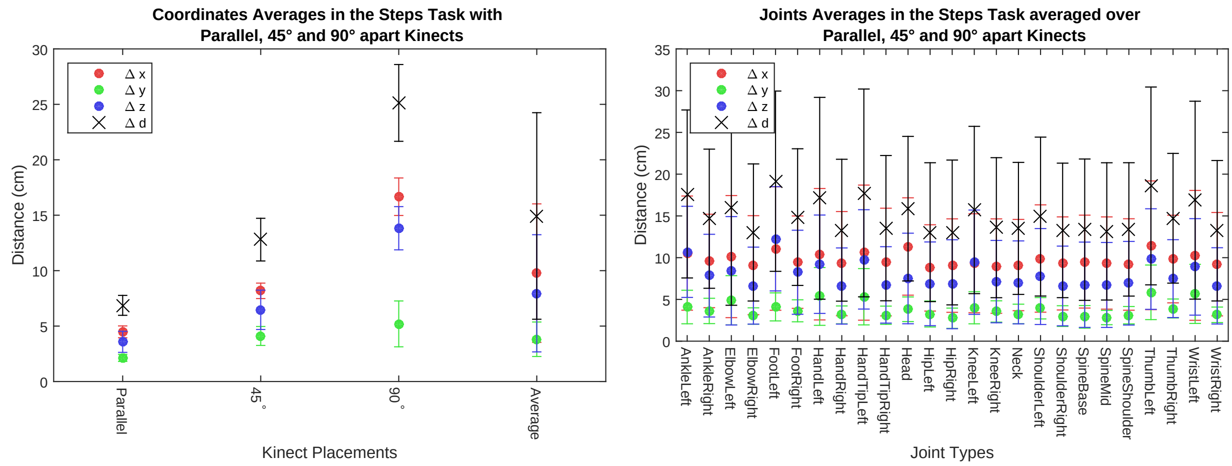
Distances	Parallel	45°	90°	Average
Avg. Δx	1.84	3.38	7.30	4.17
Std. Δx	1.03	1.52	2.94	2.82
Avg. Δy	1.28	3.59	4.35	3.07
Std. Δy	0.49	1.50	2.15	1.60
Avg. Δz	2.08	3.17	5.1917	3.48
Std. Δz	0.89	1.45	1.84	1.58
Avg. Δd	3.52	6.95	11.39	7.29
Std. Δd	1.33	2.67	4.45	3.95

Table 9.1: Average coordinates distances in the Steps task with Parallel, 45° and 90° Kinects, as well as the average case. The means and standard deviations for Δx, Δy, Δz, and Δd are reported.

9.6 Steps

This section reports average coordinates and joints distances in the Steps task with parallel, 45° and 90° apart Kinects.

The Δd distance in the Steps task with parallel Kinects is 6.87. The Δd distance in the Steps task with 45° apart Kinects is 12.80. The Δd distance in the Steps task with 90° apart Kinects is 25.13. The average Δd distance in the Steps task over all Kinect placements is 14.93, with a standard deviation of 9.32.



(a) Average coordinates distances in the Steps task with Parallel, 45° and 90° apart Kinects.

(b) Average joints distances in the Steps task averaged over Parallel, 45° and 90° apart Kinects.

Figure 9.2: Plots showing the results in the Steps task with different Kinect placements.

Distances	Parallel	45°	90°	Average
Avg. Δx	4.48	8.18	16.7	9.78
Std. Δx	0.53	0.70	1.70	6.25
Avg. Δy	2.13	4.11	5.20	3.81
Std. Δy	0.32	0.86	2.07	1.56
Avg. Δz	3.58	6.47	13.83	7.96
Std. Δz	0.95	1.77	1.95	5.28
Avg. Δd	6.87	12.80	25.13	14.93
Std. Δd	0.90	1.92	3.46	9.32

Table 9.2: Average coordinates distances in the Steps task with Parallel, 45° and 90° Kinects, as well as the average case. The means and standard deviations for Δx, Δy, Δz, and Δd are reported.

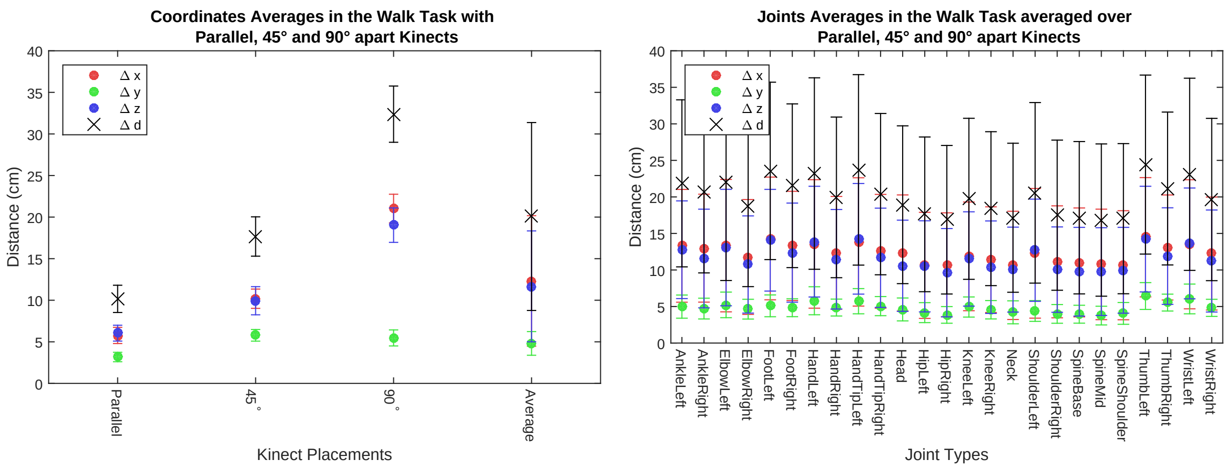
9.7 Walk

This section reports average coordinates and joints distances in the Walk task with parallel, 45° and 90° apart Kinects.

The Δd distance in the Walk task with parallel Kinects is 10.17. The Δd distance in the Walk task with 45° apart Kinects is 17.67. The Δd distance in the Walk task with 90° apart Kinects is 32.38. The average Δd distance in the Walk task over all Kinect placements is 20.07, with a standard deviation of 11.30.

9.8 Stationary, Steps, Walk

This section summarizes the results in the Stationary, Steps, and Walk tasks with Parallel, 45° and 90° apart Kinects.



(a) Average coordinates distances in the Walk task with Parallel, 45° and 90° apart Kinects.

(b) Average joints distances in the Walk task with Parallel, 45° and 90° apart Kinects.

Figure 9.3: Plots showing the results in the Walk task with different Kinect placements.

Distances	**Parallel**	**45°**	**90°**	**Average**
Avg. Δx	5.76	10.18	21.02	12.32
Std. Δx	0.97	1.16	1.73	7.85
Avg. Δy	3.17	5.78	5.47	4.81
Std. Δy	0.57	0.70	0.96	1.42
Avg. Δz	6.04	9.94	19.03	11.67
Std. Δz	0.95	1.69	2.07	6.67
Avg. Δd	10.17	17.67	32.38	20.07
Std. Δd	1.64	2.37	3.38	11.30

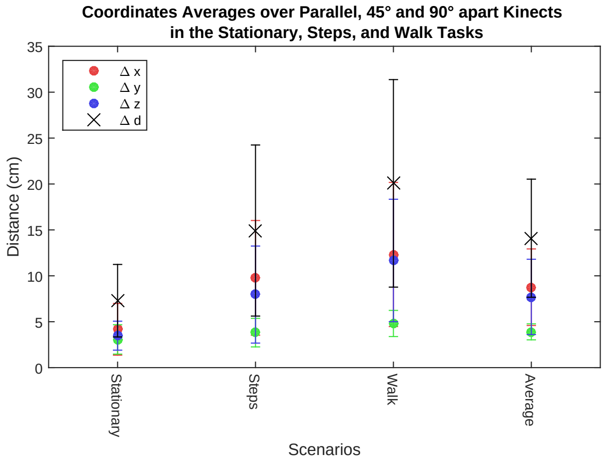
Table 9.3: Average coordinates distances in the Walk task with Parallel, 45° and 90° Kinects, as well as the average case. The means and standard deviations for Δx, Δy, Δz, and Δd are reported.

Figure 9.4 shows two different plots. Firstly, it shows the average coordinates and joints distances in the Stationary, Steps, and Walk tasks averaged over different Kinect placements (Parallel, 45° and 90° apart Kinects). The reverse is also shown. The figure also shows the average coordinates and joints distances with Parallel, 45° and 90° apart Kinects averaged over different tasks (Stationary, , Steps, and Walk). The figure shows how the complexity of tasks and the placement of the Kinects, respectively, affect the accuracy of the tracking algorithm.

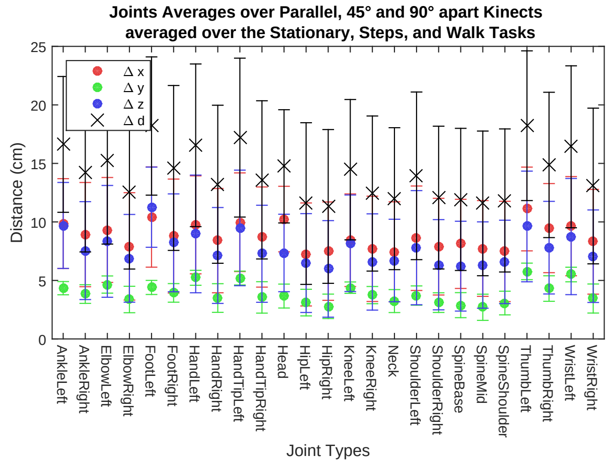
9.9 Obstacle

9.10 Interactions

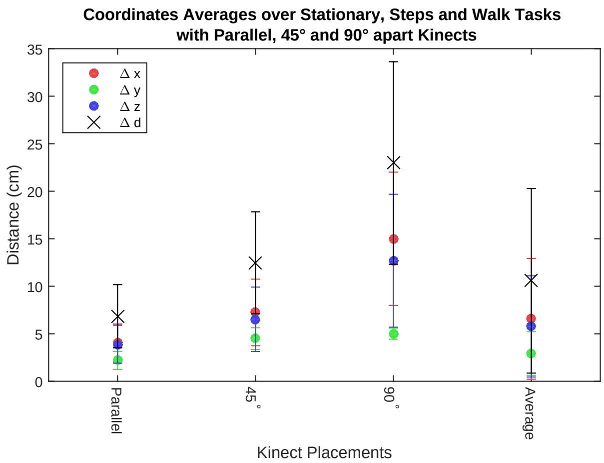
9.11 Overall



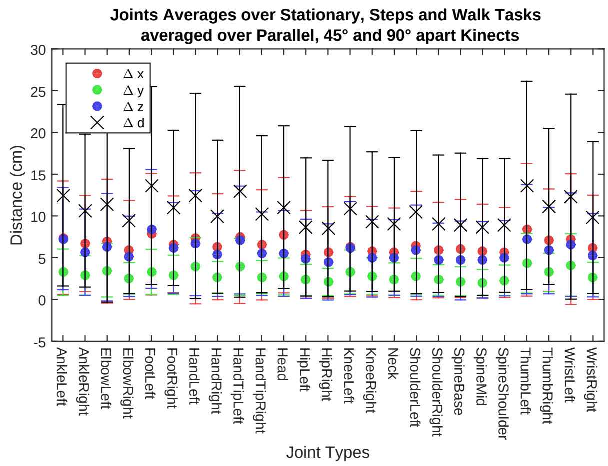
(a) Average coordinates distances in the Stationary, Steps, and Walk tasks averaged over Parallel, 45° and 90° apart Kinects.



(b) Average joints distances in the Stationary, Steps, and Walk tasks averaged over Parallel, 45° and 90° apart Kinects.



(c) Average coordinates distances with Parallel, 45° and 90° apart Kinects averaged over Stationary, Steps, and Walk tasks.



(d) Average joints distances with Parallel, 45° and 90° apart Kinects averaged over Stationary, Steps, and Walk tasks.

Figure 9.4: Plots showing the overall results in the Stationary, Steps, and Walk tasks with different Kinect placements.

Distances	Stationary	Steps	Walk	Average
Avg. Δx	0	0	0	0
Std. Δx	0	0	0	0
Avg. Δy	0	0	0	0
Std. Δy	0	0	0	0
Avg. Δz	0	0	0	0
Std. Δz	0	0	0	0
Avg. Δd	0	0	0	0
Std. Δd	0	0	0	0

Distances	Parallel	45°	90°	Average
Avg. Δx	4.03	7.25	15.00	6.57
Std. Δx	2.00	3.50	7.00	6.35
Avg. Δy	2.19	4.49	5.01	2.92
Std. Δy	0.95	1.15	0.58	2.30
Avg. Δz	3.90	6.53	12.68	5.78
Std. Δz	2.00	3.39	6.99	5.33
Avg. Δd	6.85	12.47	22.97	10.57
Std. Δd	3.32	5.36	10.66	9.71

Table 9.4: Table showing coordinates distances in the Walk task with Parallel, 45° and 90° Kinects, as well as the average case. The means and standard deviations for Δx, Δy, Δz, and Δd are reported.

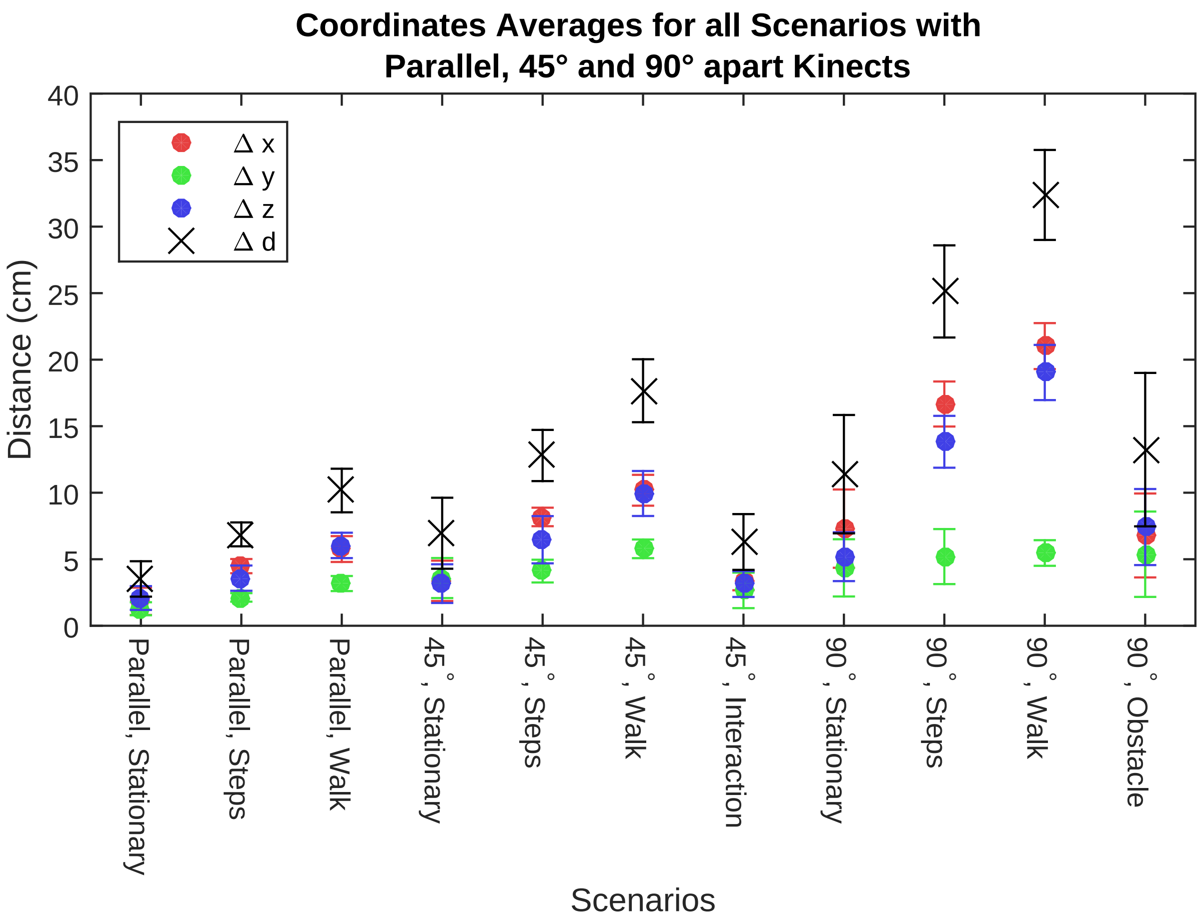


Figure 9.5: Plot showing the overall results for all scenarios in the studies

Setup	Avg. Δx	Avg. Δy	Avg. Δz	Avg. Δd
Parallel, Stationery	1.84	3.38	7.30	4.17
Parallel, Steps Δy	0.49	1.50	2.15	1.60
Parallel, Walk Δy	0.49	1.50	2.15	1.60
45°, Stationery	1.03	1.52	2.94	2.82
45°, Steps	0.89	1.45	1.84	1.58
45°, Walk	0.89	1.45	1.84	1.58
45°, Interaction	0.89	1.45	1.84	1.58
90°, Stationery	1.28	3.59	4.35	3.07
90°, Steps	3.52	6.95	11.39	7.29
90°, Walk	3.52	6.95	11.39	7.29
90°, Obstacle	1.33	2.67	4.45	3.95

Table 9.5: Table showing the overall average coordinates distances

CHAPTER 10

Discussion

10.1 Evaluation

10.2 Future Work

CHAPTER 11

Conclusion

CHAPTER 12

Notes

12.1 Project

12.2 Software

CHAPTER 13

Ethics

CHAPTER 14

Acknowledgements

The author would like to thank Dr. Michael Weir and Dr. Erion Plaku for their valuable insight throughout the development of this project and the School of Computer Science for a truly phenomenal undergraduate education. The author would also like to thank his wife, Jessica, for putting up with the late nights and limited contact whilst this project slowly enveloped his life.

CHAPTER 15

Appendix

Bibliography