P9

Jan Sundgaard Schultz & Simon Christian Munk Sørensen

January 11, 2013

Faculty of Engineering and Science
Aalborg University

Department of Computer Science

TITLE:

Identity in Proxemic Interaction

PROJECT PERIOD:
SW9, INF9,
1^{st} September 2012 -
11^{th} January 2013

PROJECT GROUP:
sw909e12

GROUP MEMBERS:
Jan Sundgaard Schultz
Simon Christian Munk Sørensen

SUPERVISOR:
Mikael Skov

NUMBER OF COPIES: 4

REPORT PAGES: 63

APPENDIX PAGES: 29

TOTAL PAGES: 92

ABSTRACT:

This project report is part one of a two-part project spanning two semesters. The goal of this report is to investigate if it is possible, using the XBox Kinect, to recognise a person from a set of five or less people. A large amount of data was collected during a two-day experiment which involved tracking and measuring users as they placed themselves in a couch. The data collected trough this experiment was then used as input for a recognition algorithm. The result was that two out of four tests were successful, and a user was correctly recognised. There was however no conclusive evidence as to which parameters impact the possibility of a successful recognition. The data does suggest that the height of a person may have an impact, but there was no decisive evidence of this.

CONTENTS

CHAPTER 1 INTRODUCTION

This project will explore ubiquitous computing with a focus on proxemic interaction. This report is part one of a larger master project spanning the 9th and 10th semester at Computer Science at Aalborg University. Part two will be the focus of the 10th semester, and will be discussed in Chapter 8.

The goal of these two semesters is to produce a system, that uses an a Xbox Kinect and proxemic interaction in order to make a TV more intelligent. Furthermore the aim is to be able to use the system in the context of a living room.

Looking at the features that Smart TV's currently posses they already show signs of intelligence. Samsung has created a television that uses both voice commands and can identify a person by face recognition, and thereby logging a person into his or her account [28]. This however does not work seamlessly, a command has to be given in order to initiate the face recognition.
According to Greenberg et al. [5] there are five dimensions to proxemic interaction: Distance, Orientation, Movement, Identity and Location. This 9th semester report will focus on the identity dimension, which enables identification of an object, in this case, a person. The identification of a person opens up for some interesting interaction possibilities, for example if a person can be identified it will also be possible to direct specific or even personal information towards this person.

There are many ways of recognizing a person, Preis et al. [26] has shown that gait recognition can be successful in identifying a person, Turk and Pentland [29] has shown that face recognition is also able to recognise a person. Both of these are biometric methods that can distinguish a single person from a large set of people.

One of the sub-goals of this 9th and 10th semester is to produce a system, that is capable of seamlessly recognising a person when he or she is watching TV in their living room. There exists numerous facial recognition methods, but these have the disadvantage that they do not work well when there is limited light [8], which can certainly be the case when watching TV. Gait recognition could also prove difficult to use since the user is unlikely to do much walking in front of the TV.

Facial and Gait recognition methods are both able to uniquely identify a person from thousands of others, making them valid biometric systems. The intended use of this system is as already mentioned, the living room of a normal household in Denmark. According to Danmarks Statistik [2] there are a total of 2.593.553 households in Denmark. We have chosen to limit this system to households of 5 persons or less, which make up 98.6% of the total number of households. This means that the use of soft biometrics would most likely be sufficient to identify and recognise a person uniquely among a pool of 5 persons or less. Examples of soft biometrics could be a persons height, weight, color of the eyes, etc.. A soft biometric is a biometric, that is not unique to only one person, e.g. two people can have the same height, this means that it is not able to uniquely identify one person from thousands of others. If one soft biometric is not accurate enough it is possible to combine them and increase the possibility of recognising a person. [10]

As mentioned the system should make use of an Xbox Kinect. The XBox Kinect is a device that uses a depth sensing sensor and RGB cameras to determine if a user is present in front of it. This makes it usable in the living room and in front of the TV because it works well in normal or dimmed lighting conditions. The Kinect is also able, via its software and cameras, to identify if a person is present and determine the location of his or her joints, thereby making it possible to calculate for example the height of a person, or even the length of a limb such as the arm.

This report will investigate, what kind of data can be produced using the Kinect, e.g. how accurate it is, and if there is a possibility, that this data can be used to determine who a user is. This leads us to the initiating problem:

Problem Statement

> ”Is it possible, using the Microsoft XBox Kinect, to uniquely recognise a person based on the skeleton data provided by the Windows Kinect SDK?”

CHAPTER 2

BACKGROUND

As mentioned in the introduction the goal of 9th and 10th semester is to investigate proxemic interaction. Therefore the first section is a short introduction to this area, and it will introduce the four proxemic zones, that was coined by Edward Hall. This will lead into the next section, proxemics in HCI, which will explain, how proxemics can be used in HCI and ubiquitous computing. This section also defines five areas of proxemics HCI, one of which is identify, which is also the area, this project will look further into. Identification will be explored in further detail in the recognition section, where different ways of identifying people will be explained. Lastly this chapter will look into the Xbox Kinect, which is the sensor, that will be used to identify people.

2.1 Proxemics

Proxemics is a part of nonverbal communication, which is generally defined as communication without spoken words [13]. Proxemics describes the distances people will have to each other, in different contexts. Anthropologist Edward T. Hall found, that people usually keep four different distances to other people; Intimate, Personal, Social and Public. Originally Hall and linguistic scientist George Trager found eight distances, but after further observations, he adjusted it to four zones, each with a close and a far phase. [6] The four zones are as follows.

2.1.1 Intimate distance

The zone spans from 0 - 18 inches (0 - 0.45 m). This distance is used for physical contact, for example for making love, comforting and protecting. In the close phase different body parts can touch each other, and arms can be wrapped around the other person, sharp vision is mostly blurred. In the far phase hands can touch another person's extremities also the eyes can easily focus, but objects may seem enlarged and distorted. This enlargement and distortion can also be the reason, why Americans feels discomfort when people, with whom they are not intimate, step into the zone. [6]

2.1.2 Personal distance

The zone spans from 1.5 - 4 feet (0.45 - 1.20 m). The personal zone, is like a bubble, that a person keeps to other people. In the close phase of the zone, one can touch or grasp another person. There is no longer a visual distortion of features and they can be seen clearly. In the far phase, another person is just out of arm's reach, but details in face and clothes are easily seen. [6]

2.1.3 Social distance

The zone spans from 4 - 12 feet (1.20 - 3.65 m). In this zone, it is no longer possible to touch another person. In the close phase, the head size is still perceived as being of normal size, at four feet the upper body can be seen, and at seven feet the whole person can be observed. This distance is used for impersonal business and social gatherings. The far phase is used for business and social events of a more formal character. The finest details of the face and clothes are lost. The far phase can be used to screen of people from each other, thereby enabling them to do the work at hand, without feeling the need to converse. [6]

2.1.4 Public distance

The zone spans from 12 - 25 or more feet (3.65 - 7.60 or more m). This zone is usually used, when in presents of important public figures, but since the person is out of direct involvement, he/she has to exaggerate the verbal and nonverbal communication. In the close phase, you can no longer see the fine details of the eyes and skin, also the contours of the body are lost and begin to look flat. At twelve feet, it is also possible to take evasive/defensive action if threatened. In the far phase persons have to exaggerate (non)verbal communication. The zone ends, when you can no longer make out objects as people. [6]

Hall notes, that there are zones or territories which people keep, like bubbles of space around them, but also, that these zones can vary from culture to culture. This means that the personal distance could be smaller in some cultures or depend on the relationship to a person (family/not family). [6]

2.2 Proxemics in HCI

Proxemics has been used in different HCI contexts, for example in Obata and Sasaki [25] where OfficeWalker, a kind of video chat room, that uses proxemics in order to be less intrusive, when contacting a person or in Kastanis and Slater [11] where proxemics is used by avatars in a virtual world, to manipulate users to move to a desired area. In the Greenberg et al. [5] article ”Proxemics Interactions: The New Ubicomp?”, they used proxemics as an element in ubiquitous computing (ubicomp). In order to use the proxemics concept Greenberg et al. [5] operationalised it and proposed five dimensions of proxemics in ubicomp. The five dimensions are; *Distance*, *Orientation*, *Movement*, *Identity* and *Location* which will be explained below.

Distance can be thought of as a continuous measure for example from zero to one meter, or as a discrete measure like Hall's proxemic zones, where you can determine, which zone a person is in, in relations to another person or technology. *Orientation* can also be split into continuous and discrete measurements. The continuous measure could for example be, the angle between to objects, and the discrete could be, if the object is oriented towards, semi-towards or away from another object. *Movement* can be seen as distance and orientation of an object over time, is the object moving closer or further away, and is it oriented towards or away from another object. *Identify* describes an object, which can range from minimal measures like making a distinction between one object to another, to detailed measures of exact identity and attributes. *Location* describes the environment or context an object is in, for example the kitchen or the living room. Location is also important for the other dimensions, since interaction can differ between contexts. [15]

2.2.1 Six design challenges

In the later article ”Informing the Design of Proxemic Interactions” Marquardt and Greenberg [15] proposed six design challenges associated with proxemics related ubicomp. Five of the challenges will be explained in short and the sixth, managing privacy and security, will be explained in greater detail below.

Challenge one concerns *revealing interaction possibilities*. In traditional GUI design the different elements are in the foreground of the user's attention, in ubicomp however, the elements will be nearly invisible until needed [15]. The challenge is therefore:

> ”how can technology be designed to reveal the interaction possibilities appropriate when it is not only in the background of a persons attention, but during the transition of it moving into the foreground?” [15]

The second challenge concerns how a user is *directing actions*, in for example PC's, the input to the system usually comes from another device, like a keyboard, but in ubicomp other input methods can be used, such as gestures or voice commands. But this calls for a device, that is able to make distinctions between input directed towards the system and non-input. The third challenge is about *establishing connections* between different devices, because ubicomp functions seamlessly, it is challenging to connect devices to each other, while still safeguarding privacy and security. Meaning that functions like sharing, should only be possible in certain situations, and not in others, for example in the personal zone, but not in the social zone, furthermore not in all personal zone encounters. Challenge four concerns the issue of *providing feedback*. In traditional HCI feedback is very important, but in ubicomp, it will be even more important to give feedback about status of applications, user input and errors, due to the very nature of ubicomp. Challenge five looks at *avoiding and correcting mistakes*, because many ubicomp systems are comprised of a sensing technology controlled by computers, they will inevitably make mistakes, it is therefore important for these systems to handle or avoid mistakes. [15]

The last of the six challenges concerns *managing privacy and security*, and since ubicomp systems provide a potential of multiple people interacting with the system, the need for privacy and security increases. [15] But:

> "The question is how can the system protect privacy sensitive information and handle the access to information, while at the same time not get in the way of all the positive offerings of ubicomp mentioned in Challenges 1-5?" [15]

Marquardt and Greenberg also give examples on different scenarios, on how to answer this question. With *proximity-dependent authentication* the combination of an authentication token, and proximity are used in order to grant access or connect to another device. This is also true for *distance-dependent information disclosure*, but an added security is, that the closer an object moves to another object, the more information is revealed. With *proxemic-aware privacy mechanisms* more of the proxemic dimensions, could be put in play, for example could orientation be used to reveal or hide information. Location could be used to change the security of a device, at home it would be more relaxed. Identity could also be used to manage privacy by hiding information, if another person enters the objects proximity: [15]

> "a system would be able to use relaxed privacy and security settings when a person is alone, but switch to more restrictive privacy and security settings when it detects any other people or devices around them" [15]

This project will look into the identify dimension of proxemic ubicomp as proposed by Greenberg et al. [5], but also the challenge of managing privacy and security as stated by Marquardt and Greenberg [15]. Because the system will be build for the home environment, it should be fairly small and easy to set up. Therefore the system will use a PC for input/output processing and a Xbox Kinect for providing sensor data. However the main goal of this project is to explore the possibility of identifying different users seamlessly in order to build in a kind of privacy or security into our system. Since the system will be used in a home by a limited amount of people it is proposed, that the Xbox Kinect will give sufficient data in order to identify and differentiate between the users. The next step is therefore to look into general ways of identifying users/people, and afterwards look into ways of identifying people using the Xbox Kinect.

2.3 Recognition

In the real world, when a person recognizes another person, it is usually by the way they look, the face, height or build of a person, from afar it could be the way they walk, and over a telephone it could be the sound of the voice. When computers have to recognize a person, it is done in a similar fashion, and is called a biometric system. Such a system recognizes distinct patterns in biometric data and compares these, to data stored in a database. According to Jain et al. [9] biometric data is comprised of biological measurements, that should satisfy four requirements. They should have *universality*, meaning all people should have the characteristic, it should have *distinctiveness* so two people are different from each other, it should have *permanence* so it does not change over a period of time, and the data should have *collectability* meaning it should be quantifiable. However in order for the system to be more practical, other factors could be considered, for example the *performance* of the system i.e. the recognition speed and accuracy. Furthermore it could be *acceptability* of the system by its users, and the system should also be secured against *circumvention.* [9]

In biometric systems there a two modes of operation *verification* and *identification.* Verification mode does a one-to-one comparison, this is done by giving information about who the person is, and then verifying, if this is the person [9]. An example could be an ID-card, that is used to get into a building. The ID-card tells the system, that a specific person is trying to get into the building, the PIN code tells the system, that this is in fact the person. Identification mode does a one-to-many comparison, this is done by checking the identity against all the people in a database [9]. An example of this could be, if the same building could scan, for example fingerprints or perhaps DNA, you could let your finger get scanned at the door, and the system checks the data against a database of people, that have access rights in order to find a match.

There are many different biometrics, that can be used for recognition, for example DNA, the shape of the ear, veins in face and hands via infrared thermogram, finger- and palm prints, hand and finger geometry, iris and retina, keyboard keystroke, odor, signature, face, gait and voice recognition. [9] Since the the Kinect has a camera, can calculate joints and skeleton of a person over time and has microphones, this project will look further into face, gait and voice recognition, since these would be possible with the Kinect, but also in the context of the living room.

Face recognition is one of the most common biometric characteristic used by people to identify another person. In biometrics systems it ranges from static (e.g. image) to non-static (e.g. video) identification [9]. Turk and Pentland [29] was some of the first to use eigenfaces in face recognition, and their work is used, as a ground basis for many similar ways of recognition [8]. In the article "Face Recognition Using Eigenfaces" Matthew Turk and Alex Pentland showed, that by extracting relevant information in images of a faces, it is possible to make a database, that captures the variation in faces independent of distinct features. This information can then be used, to evaluate individual faces. They achieved this by firstly creating an "average face" or mean face, out of a number of face images. Then the mean face is subtracted from each of the face images to get the deviation of each image. Then they created the eigenfaces (eigenvectors) from each of the face images which shows the largest similarities between some faces, and the most drastic differences between others. The vector subspace of these eigenfaces can then be used to represent any face as a linear combination of the eigenfaces. To recognize a face it is turned into an eigenface, and then checked against the other faces in the subspace, the recognition is done by calculating the smallest distance to each of the faces in the subspace, thereby finding a match or not. [29].

Face recognition has also been used with the Xbox Kinect like in Leyvand et al. [14] where they used a combination face recognition, height and color of clothes to recognize a player for video games.

Gait analysis is done in a similar fashion as face recognition, but instead of images or video of a face, it is of the way people walk. The analysis can be done by 2D images like [30], where they used algorithms to extract the silhouette of people in images and then using principal component analysis (PCA) and eigenspace transformation in order to recognize people by their walk. Gait analysis has also been done with 3D data like [27] where they used motion capture, in order to gather more precise data of peoples gait. The Xbox Kinect has also been used for gait analysis in [26], they used the skeletal data of the Xbox in order to track 20 different points of the skeleton and used 13 biometric features; height, length of legs, torso, lower legs, thighs, upper and lowers arms and also step length and speed. This resulted in a identification success rate of 91% with nine test persons.

Voice biometrics or speaker recognition [4] depends on a combination of physiology, the shape of a persons voice organs, and behavior, which changes due to age, emotional states or just a common cold. There is two kinds of voice recognition, text-dependent and text-independent. The text-dependent system is based on predefined utterances, text-independent systems does not and are therefore harder to design, but also gives better protection. [9] Voice biometrics is very useful for remote users and systems, via for example telephones [4], they are however sensitive to e.g. background noise [9]. In Galatas et al. [3] they used a combination of video-, voice recognition and Xbox kinect depth data to boost automatic speech recognition.

Biometrics are unique characteristics to a person, however there are also characteristics, that are not unique, these are called soft-biometrics. The information gathered by soft-biometric systems are for example height and other anthropometric measurements, but also weight, gender, age, eye color and ethnicity [10], they also defined these soft biometrics traits as:

> "...characteristics that provide some information about the individual,

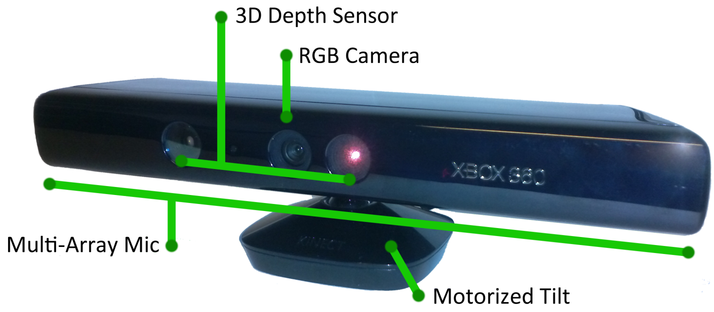


Figure 2.1: An image of the Kinect.

> but lack the distinctiveness and permanence to sufficiently differentiate any two individuals.” [10]

There are however different views on the possible use of soft biometrics in Jain et al. [10] they propose using soft biometrics as ancillary information to improve performance on primary biometrics. In Heckathorn et al. [7] they suggest that using the statistical principle of interchangeability of indicators, that combines multiple low accuracy indicators to produce a highly accurate one. They also find that skeletal measurements for example height, width of ankles and wrists, length of forearms and hat size are among the most reliable.

In order to find suitable way of recognizing a person the hardware and software of the Kinect, will be explored in the next section.

2.4 XBox 360 Kinect Hardware

This section will briefly discuss the hardware provided by the Kinect. Section 2.5 will discuss the features provided by the Kinect for Windows SDK.

2.4.1 Sensors

The XBox 360 Kinect (Kinect) provides a depth image, which allows an application to determine the distance to any point, within its field of view. This image is made possible using two sensors, an IR emitter and a receiver, marked as the 3D Depth Sensors on Figure 2.1[20].

The depth image can achieve a resolution as high as 640x480 pixels, at a frame rate of 30 frames per second[17].
The available RGB camera provides the application with an image stream, that can, similarly to the depth sensors, provide a video stream of 30 frame per second, the resolution can go as high as 1280x960 pixels.
The Kinect also provides a microphone array, making it possible to detect from where sound has originated, a tilt motor which can tilt the Kinect +- 27 degrees, and lastly an accelerometer.

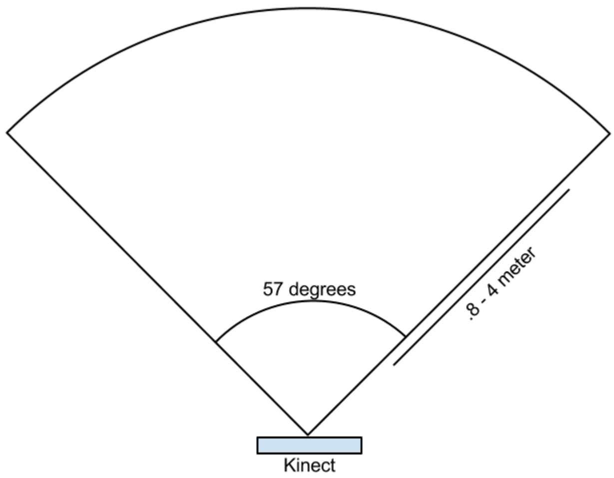


Figure 2.2: A figure showing the horizontal field-of-view of the Kinect.

The viewing angle of the Kinect depth sensor and RGB camera is 43 degrees vertical and 57 degrees horizontal, as can be seen in Figure 2.2.

2.4.2 XBox 360 Kinect vs. Windows Kinect

Two versions of the Kinect has been released by Microsoft, the XBox 360 Kinect used in this project, and the Windows Kinect. The XBox 360 Kinect was released for XBox 360 in November 2010, and is meant to be used together with console games. The Windows Kinect was released by Microsoft in February 2012[24] and is aimed at commercial products, rather than gaming products. The more apparent differences between the Windows Kinect and the XBox 360 Kinect is the Near Mode feature. Instead of having a range going from 80-4000 cm as the XBox 360 Kinect, the Windows Kinect has a range of 40-3000 cm[21].
This project is concerned with people watching TV, which is usually not done at a distance below 80 cm, therefore the XBox 360 Kinect's capabilities have been deemed sufficient for this project. But also taking the proxemics zones into account, it is probably the fewest people, who wants to get "intimate" with their TV.

2.5 Kinect for Windows SDK

Over the past years several unofficial SDK's, such as the Open NI framework, has been available for the development of Kinect applications. In 2011 Microsoft released their own SDK aimed at Kinect development[19]. This SDK does not officially support the XBox 360 Kinect, despite of this, it is still possible to do development using the XBox Kinect, and as described in Section 2.4.2, there are no noteworthy differences between them, in relation to this project.
The Kinect for Windows SDK version 1.6 has been chosen for this project, mainly

Figure 2.3: An image provided a visual representation of depth data, with a pixel and its data defined.

because it is an official product made by Microsoft, which is also in active development and the features are well written and detailed in online documentation. This section will detail some of the features, provided by the SDK, that are relevant to this project.

2.5.1 Depth sensing

As stated in Section 2.4, the Kinect can measure depth values at distances from 0.8 to 4.0 meters from the Kinect. When accessing the depth data through the depth stream, the developer is presented with a depth image, in the form of an array, where each pixel of the image, also contains the distance measured to that point, in millimeters.[16] It is also possible to determine, if a user is present, at that given pixel, this will be further discussed in Section 2.5.2.
Figure 2.3 represents a depth image, where the depth has been converted to shades of grey, and every pixel on top of a player is colored yellow.

The Kinect can identify up to 6 persons based on data provided by the depth stream. The depth values, that are placed in this data, has also been formatted according to Figure 2.4.

2.5.2 Skeletal tracking

The accuracy and precision of the Kinect depth data can vary depending on a number of different parameters. As stated in Section 2.5.1 the Kinect can reliably detect depth values, at a distance of 0.8 to 4.0 meters. Microsoft does however also state, that the practical limits of the Kinect returning accurate and precise data are

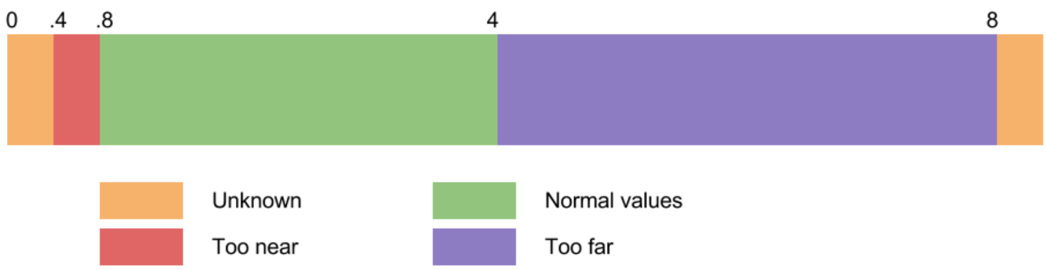


Figure 2.4: Shows which distances the Kinect can capture reliable data at.

within 0.8-2.5 meters[23]. This may be due to the fact, that the longer away from the Kinect a point is, the greater the discrepancy.

The article Khoshelham and Elberink [12] states, that the Kinect is reliable accurate to a distance of 3 meters.

Joint filtering

The depth data will sometimes return noisy data, this will lead to joints being jittery. This can be prevented using the built in smoothing filter. This filter can be used to smooth the skeletal frames over time, removing jitter and stabilising the joints. A skeletal frame is a technical term used in the Kinect SDK, it is a container, from which it is possible to extract the joint positions.[18] The smoothing filter has five parameters which can be set:

- *Smoothing*, will smooth the data over time, making the skeletal joints less jittery. Too much smoothing will also increase the latency towards the raw data.
- *Correction*, a low value will correct more slowly and make the data appear smoother, while a high value will be faster, decreasing latency
- *Prediction*, the number of frames into the future, that should be predicted. This can be used to lower the latency. It may however also result in less accurate data, since the predicted frames are based on previous data.[22]
- *Jitter radius*, setting this radius to x, will cause all jitter beyond x to be clamped to that radius, setting an effective maximum on jitter
- *Max. deviation radius*, when data has been filtered it is likely that they deviate from raw data, setting the maximum deviation radius will clamp the filtered data beyond this point to the radius.

Latency is the time it takes for the filter to process the data and output it. With a high latency, a user looks at his own skeleton on a screen would see it lag behind his own movements. Using no filter will reduce the latency, but also make the data more noisy.

When taking this background chapter and the context of the 9th and 10th semester project into consideration, this project will look into using the skeleton calculated by the Kinect, in order to identify a person. The reason for this is, that the living room context could be too dark for face recognition, and there could be too much noise for voice recognition. Also, though face recognition could be done seamlessly, voice recognition would need some amount of talk/voice, making it less seamless. Gait recognition could be used in both these cases, but this could not be used, if the person sits down before the Kinect tracks the person. The skeleton however can be tracked after the person is seated. The next chapter will detail the experiment, that was conducted in order to explore, if it is possible to use only the skeleton to identify a person.

CHAPTER 3

EXPERIMENT

This chapter will explain, what was done in the two experiments, that was performed. The first experiment was a pilot study, which was conducted in order to find any problems, with the experiment. The pilot study section will partly explain, how it was executed and partly, what was changed in the final experiment. In the experiment section the final experiment will be detailed.

3.1 Pilot study

When conducting experiments different problems can occur, to weed out some of these problems, a pilot study was performed, this section explains, how the pilot study was conducted. Furthermore it details the things, that was changed in the final experiment.

Goal

The goal of the pilot study was to identify any flaws and problems with the test setups and the application used for data collection, but also to establish, how long time the test took. The goal was to have a test, that took around 30 min, there were three reasons for this, firstly if the test took too, long it would make it harder to find people with the time, to do the experiment, secondly since the test involves some amount of walking, the test should not strain the test persons too much, thirdly the amount of data collected, for each test subject, would be very large around 20 Gb/30 min. Furthermore the goal was to collect initial data in order to see, if there were any problem areas.

Participant

There was one person in the pilot test, he was also a writer on this project. The test subject was 28 years old and 184 cm tall. He had used the Kinect before, but does not have a Kinect at home.

Data Collection

The data that was collected, was the raw Kinect data and data of the calculated skeleton. The data that was collected, was done in a the same way as the final experiment, see Section 3.5, though on the basis of the pilot study, some changes have been done to the user interface of the ICS application, mainly reducing the amount of clicks needed to save data.

Location and tasks

In the pilot study the test location was the same as the final experiment, see Section 3.6 for more information.

Procedure

The pilot study experiment consisted of 4 test setups; calibration, corner couch, single couch and perpendiculat to kinect. These will be detailed below.

Calibration

The calibration was done in order to create a baseline for recognizing a person. The baseline consists of measurements of the different limbs/skeleton, that the Kinect can detect, furthermore these measurements should be fairly stable. In the pilot study the calibration test consisted of 5 sub tasks, also see Figure 3.1:

1. Locate the return point. The return point was the distance away from the Kinect, where the entire test subject was inside the field-of-view.

2. Do a full rotation while standing at the return point.

3. Facing the Kinect, walk backwards to the 5-meter mark, then walk forward to the 1-meter mark, and lastly back up to the return point.

4. Do a quarter rotation to the left, then a half rotation to the right and a quarter rotation to the left, ending up facing the Kinect.

In the final experiment the calibration was simplified in order to get less bad data. Bad data would arise, when for example the test person walked out of the Kinects reach. The change also meant, that the calibration took less time in the final experiment.

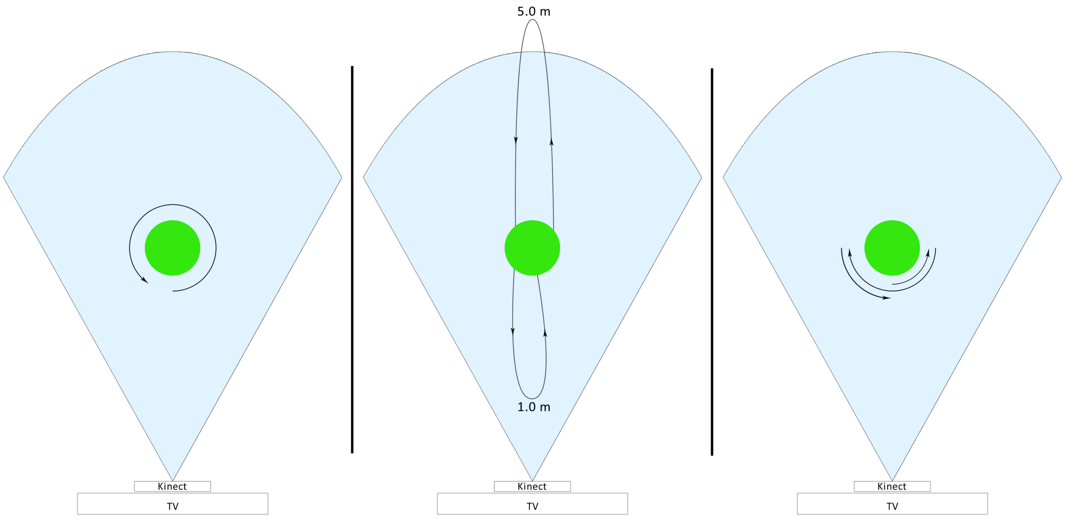


Figure 3.1: Shows the calibration test of the pilot study.

Test 2-4 couches

In test 2 corner couch two, 2-person couches were used, to create a corner couch. The back of the couch was placed at 4.0 m from the Kinect. The test subject walked from positions 1 and 4 to each of the seats A to D, see Figure 3.2. Walking from one position to a seat was also denoted a task, a task could be walking from position 1 to seat A, this is also denoted 1 $\rightarrow$ A. Each task was repeated five times, for the first two times of each seat a sit task was also performed. The sit tasks involved, leaning forward, stretching the arms into the air, picking up an imaginary item to either the left or the right and picking up an imaginary item in front of the test subject. The test took around 28 min.

In the collected video it was noticed, that there was an unnatural way of walking to the seat in task 1 $\rightarrow$ D, this was mainly because the test subject followed the guideline to the end, and then walked to the seat, instead of walking more directly to the seat. To prevent this unnatural way of walking, the position was changed to position 2 in the final experiment, see also Section 3.2. Furthermore since the goal was to achieve an experiment, that took circa 30 min it was decided to cut down on the amount of seats from 4 to 2 in the final experiment.

In test setup 3 single couch a three person couch was used, the couch was placed at 3.0 m from the Kinect to the back of the couch. The tasks was conducted like in test 2, and the list of tasks performed can be seen in Table 3.1. Furthermore an overview of the couch setup can be seen in Figure 3.3. The test took around 10 min.

In the final experiment position 1 was moved to position 2, also the three person couch was changed to a two person couch. This was done because the tasks were similar to each other, but also to cut down on the time used.

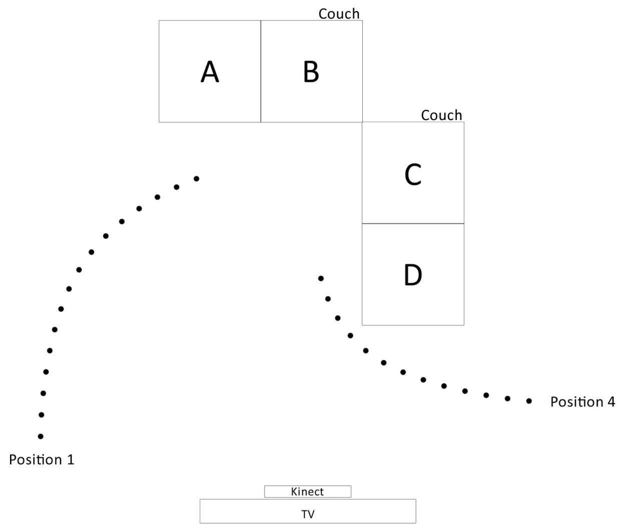


Figure 3.2: Shows the pilot study test 2 corner couch.

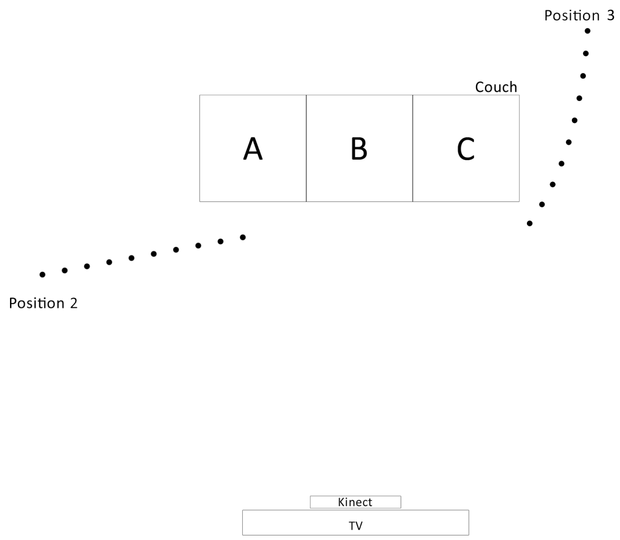


Figure 3.3: Shows the pilot study test 3 single couch.

In test setup 4 perpendicular to Kinect two, 2-person couches was placed at 2.0 m to the armrest closest to the Kinect, and they were placed perpendicular to it. The tasks conducted can be seen in Table 3.1 and an overview can be seen in Figure 3.4. The test took around 21 min to complete.

In the final test two seats was removed. One of the seats that was removed was seat A, this was due to the fact that none of the tasks from position 1 → A yielded any recognition of the test subjects skeleton, the reason for this could be, that the

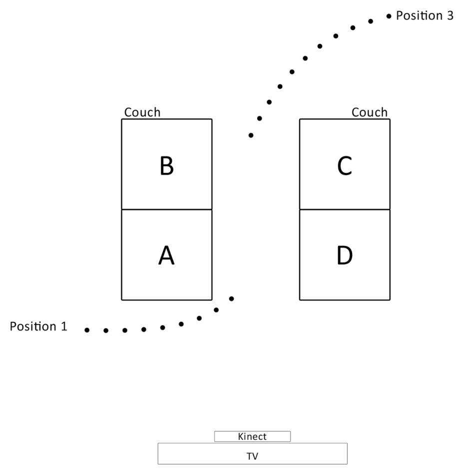


Figure 3.4: Shows the pilot study test 3 single couch.

Kinect cannot calculate the skeleton because of the route, angle of the test person and the amount of time spent in front of the Kinect. We will discuss this issue further in Chapter 6. The other seat, that was removed was seat C, this was done since the seat was placed similar to seat B, but also to cut down time consumption. Also see Appendix D for an overview of the task execution.
All four tests in the pilot study, combined took around one hour to complete, which was 30 min longer than our goal, this is also the main reason, that tasks were cut. This could however affect the test results, meaning that the results would not be as complete as doing the whole test. But since the tasks in the 3 tests cover the different seats, that was cut from the full setup, it should not affect the results.

Another thing, that was found in the pilot study, was that the test leader managing the data collection, was sitting to the left of the Kinect against the wall. This meant, that whenever he did the sit tasks, with the test subjects stretching etcetera, it was found, that the subject tended to look at the test leader. This in return meant, that the test subject was not looking at the Kinect (TV), which could give less accurate readings from the Kinect. In the experiment it was therefore chosen, that the test leader should be seated as close to the Kinect as possible, in order to keep the test more lifelike. Furthermore it was found, that the distance of the couch (seats A+B) in Test 2 felt as though they were unnaturally far away from the "TV", to explore this, the distance to the test subjects own TV at home was needed, this is why the "draw your living room" question was added to the questionnaire. This information helped decide, if the couch setups were plausible, and which new setups could be explored in future work.

Pilot Study Results

After looking through the results of the pilot test, different results were found. In test 2 corner couch it was found, that task 1 $\rightarrow$ A take number 4 (T2 - 1 $\rightarrow$ A

	Walk	Walk
Test 2 - Corner Couch	$2 \rightarrow$ A, Sit A	
	$2 \rightarrow$ B, Sit B	
		$4 \rightarrow$ C, Sit C
		$4 \rightarrow$ D, Sit D
	$2 \rightarrow$ A	$4 \rightarrow$ A
	$2 \rightarrow$ B	$4 \rightarrow$ B
	$2 \rightarrow$ C	$4 \rightarrow$ C
	$2 \rightarrow$ D	$4 \rightarrow$ D
Test 3 - Single Couch	$2 \rightarrow$ A, Sit A	
	$2 \rightarrow$ B, Sit B	
	$2 \rightarrow$ C, Sit C	
	$2 \rightarrow$ A	$3 \rightarrow$ A
	$2 \rightarrow$ B	$3 \rightarrow$ B
	$2 \rightarrow$ C	$3 \rightarrow$ C
Test 4 - Perpendicular	$1 \rightarrow$ A, Sit A	
	$1 \rightarrow$ B, Sit B	
	$1 \rightarrow$ C, Sit C	
		$3 \rightarrow$ D, Sit D
	$1 \rightarrow$ A	$3 \rightarrow$ A
	$1 \rightarrow$ B	$3 \rightarrow$ B
	$1 \rightarrow$ C	$3 \rightarrow$ C
	$1 \rightarrow$ D	$3 \rightarrow$ D

Table 3.1: This table shows the different tasks in tests 2-4, in the pilot study. The number denotes the start position and the letter denotes the seat. To better understand why the tasks look like this, see Appendix D

- 4) calculated 3 skeletons within 3 frames over 1 second. This was however not visible in the skeleton video feedback, so this result was regarded as no skeleton was detected. In (T2 - $2 \rightarrow$ D - 3, 4 and 5) there was no tracking of the skeleton. In test 3 single couch task $1 \rightarrow$ B takes 2 and 5 the Kinect took a long time to calculate the skeleton. In (T3 - $1 \rightarrow$ B - 3) did not calculate a skeleton. In test 4 perpendicular to Kinect task $1 \rightarrow$ A takes 1-5 did not calculate a skeleton. There were in total 108 recordings out of which 10 was not able to generate a skeleton, which means a $\frac{(108-10)}{(108)\times 100} = 90.74\%$ success rate. This also gave some indication of the possibility, of at least tracking a person.

3.2 The Experiment

This section will explain how the final experiment was conducted.

3.3 Goal

The goal of the project was to explore a way of identifying a person using the Kinect. The context of the identification would be in a persons home in front of the TV. In order to do explore if this was possible, an experiment was conducted. The goal of the experiment was to test the Kinects ability to track a person, when he/she was walking towards a couch and sat down or was already sitting down. When a person got tracked, measurements of the persons skeleton would be calculated. Another goal was therefore to collect these measurements from different people, in order to see if these were similar enough to a baseline collected in the first test, also called the calibration. This was done in order to find, how large the difference was, and to get an idea of how precise the measurements were. Furthermore a goal was to find, which skeleton parts were most often calculated, in order to maybe use these as primary parts for the identification. The main goal of this experiment was therefore to collect data, which could help develop a technique, using a Kinect, to identify a person by his skeleton.

3.4 Participants

In total there were 13 test subjects, 3 women and 10 men ranging from 22 to 29 years of age. All test subjects are friends, family or fellow students of the authors, and were all personally asked if they wanted to participate in the experiment. 12 of the test subjects are students or are former students at the department of computer science at Aalborg University. The height of the test subjects span from 158 - 196 cm, with a male average of 181.2 cm, and a female average of 165 cm. It should be noted that the test subjects did not have their height measured during the experiment, the height referred to is the one provided by the test subjects in the questionnaire. Five of the test subjects had previously used a Kinect before, but none of them owned a Kinect. The questionnaire answers can be found in Appendix B.

3.5 Data Collection

This section will detail the data collection methods, that was used in this experiment. Two types of data was collected from this test:

- Raw Kinect data packed into an .xed file
- Calculated skeletons

Both of these will be explained further in the following sections.

As explained in Section 2.5.2 the Kinect can be fine-tuned using a filter, so it automatically tries to improve on jittery data. As will be discussed in the future works, there are many data filters, that can be applied to the Kinect data. For future data analysis one or several different filters might be used, to improve on

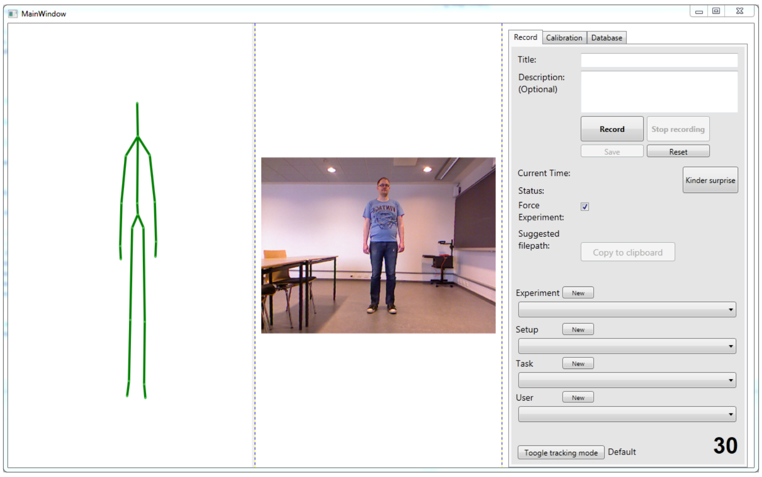


Figure 3.5: Shows a screenshot of the ICS application.

the data returned by the Kinect, which was why, the raw data was recorded using Kinect Studio 1.6.

The calculated skeletons was used in the data analysis to see how reliable the Kinect depth data was.

3.5.1 The ICS application

The application "Identify user and Calculate the Skeleton" (ICS) was developed specifically with this experiment in mind. A screenshot of the application can be seen in Figure 3.5. This section will briefly explain, the core features of ICS together with the underlying data structure.

The functionality concerned with this test, will be explained in the following section. In Figure 3.5 two tabs are shown: Calibration and Database, these tabs does not contain any functionality relevant to this experiment, and are merely there for technical reasons.

Recording

The ICS application was used for identifying and tracking a user, calculate his skeleton based on the joints provided by the skeleton stream, and save it to a database.

Figure 3.6 shows the UML diagram of the underlying data model of the ICS application. The Experiment, Setup and Task entities are self explaining, they are there as to make the ICS application easier to work with.
A Recording consists of zero to many Calculated Skeletons. There was one recording for each repetition of a task. Because of this separation, it was possible to easily distinguish task repetitions from each other, and perform data analysis more easily.
A calculated skeleton consists of 12 limbs, their length and a time stamp.

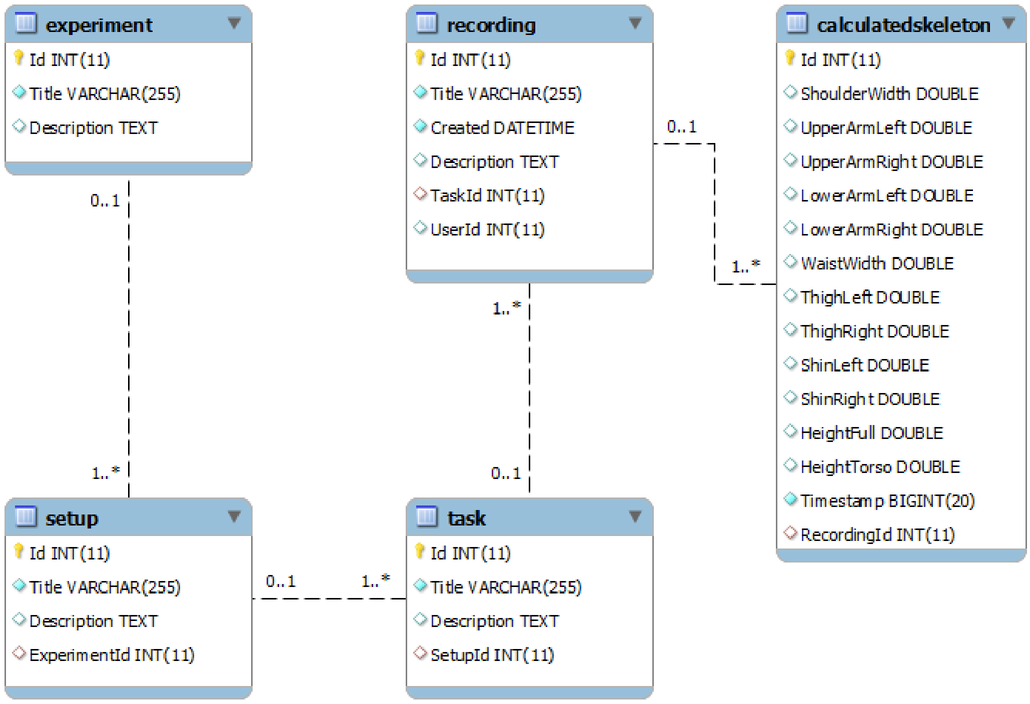


Figure 3.6: Shows the UML diagram of the underlying data model of the ICS application.

Recordings was made using the ICS application. A recording consisted of the measurement, that can be seen on Figure 3.7. Furthermore a recording had a relative time stamp provided by the Kinect.

A recording was associated with an experiment, a setup, a task and the name of the user.

3.5.2 Kinect Studio v1.6

Kinect Studio was used to collect raw Kinect data, this includes among others, a video stream and a stream of depth data, which was saved to a .xed file. Kinect Studio registers any application, that was connected to a Kinect, and is able to act as a man-in-the-middle, collecting any data sent from the Kinect to the application, that initialized it.
Kinect Studio is also able to push data into an application, meaning instead of collecting the data, it will override any data sent from the Kinect, and instead inject previously recorded data.
As explained in Section 2.5.2 it would be possible to use data filters to improve on the raw data, which was why the data was recorded for later use. This avoids the need, for having test persons to constantly test any modifications to the ICS

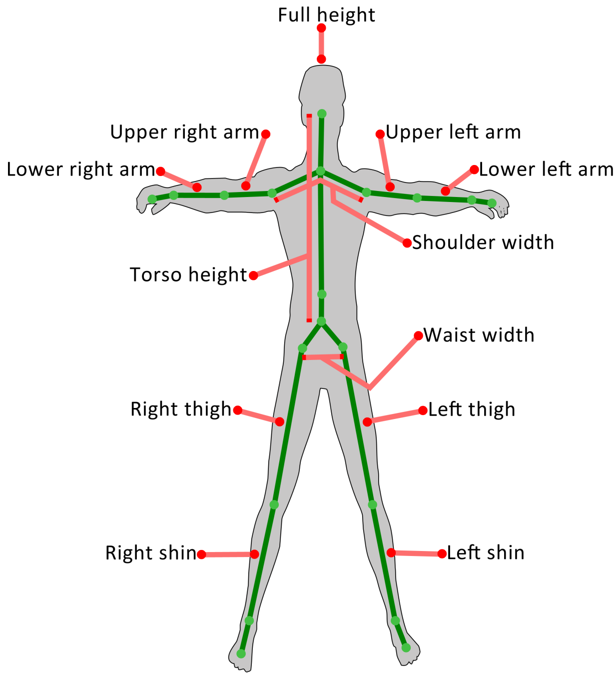


Figure 3.7: Shows the joints and limbs of a skeleton. The ones with a name, was used for identification.

application.

3.6 Setting

There are an endless number of different configurations both for how you have arranged your couch and your TV, the path you take when you go to sit in the couch, and how you sit in the couch. In order to limit the infinite number of combinations, three different couch test setups was chosen, these will be explained in the following section. Each setup had a number of different tasks, which consists of a starting position, a path the users walked towards the couch, and a seat on which they sat. All users was subject to all tasks, and all tasks were performed several times. Each user also partook in the test setup called "Calibration".

The Test Location and Items

The experiment was conducted in a classroom at Aalborg University. The size of the area used for the experiment is 5.50 m times 5.38 m (DxW), the room was however larger than the used area, and a picture both can be seen in Figure 3.8. The couches were 0.60 m times 1.35 m (DxW) and two people can be seated in each of them, the armrest and back of the couch had the same height, 0.75 m. Furthermore the Kinect was placed at a height of 1.16 m and an angle of -7 degrees.

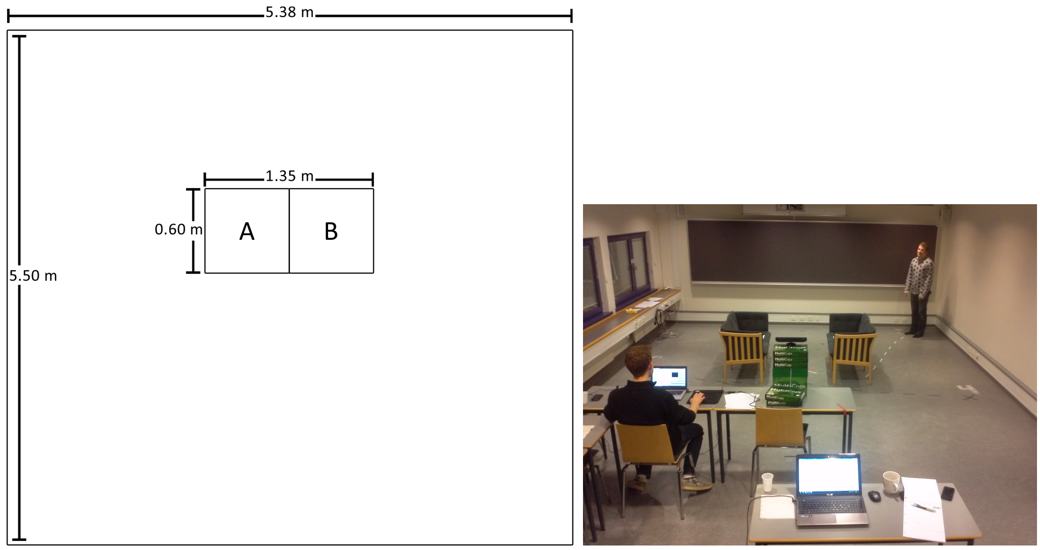


Figure 3.8: The left figure shows the size of the area used to perform the experiment and the right shows a picture of the full room.

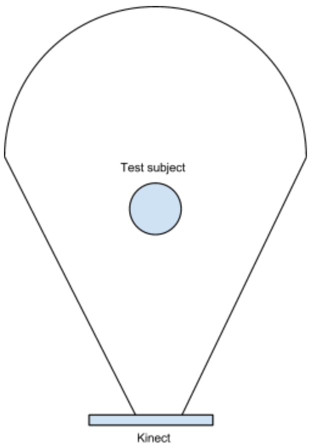


Figure 3.9: Shows where the person was positioned when calibrating.

Test 1 - Calibration

In the calibration test, the test person was placed at a distance of approximately 2.5 m, so that the person was seen in whole by the Kinect. The setup can be seen in Figure 3.9

Test 2 - Corner Couch

In this test the back of the left couch, the one facing the Kinect, was placed at 4.0 m, the other couch was placed perpendicular to the first couch. The Kinect was

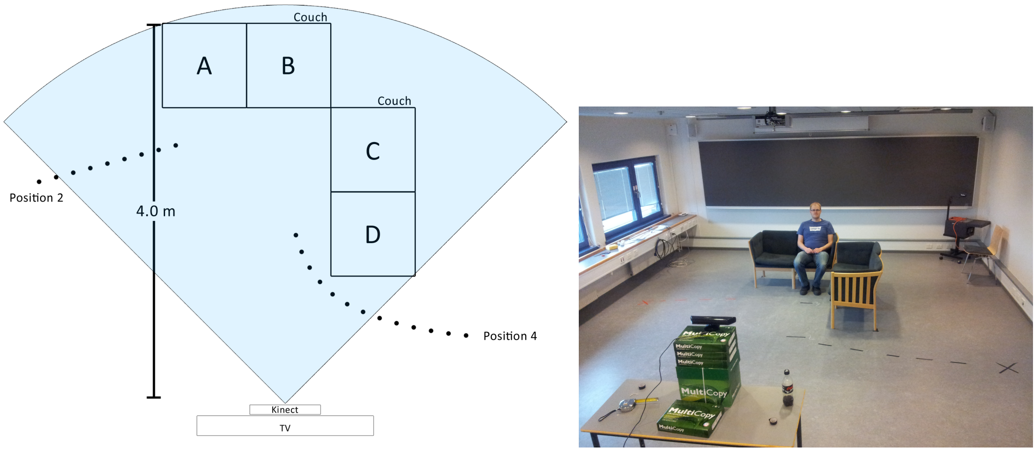


Figure 3.10: The picture to the left shows an overview of test setup 2, the right shows a picture from the experiment.

placed at the centre of the width of the couches, see also Figure 3.10

Test 3 - Single Couch

In test 3 there was a single couch, the back of the couch placed at 3.0 m. The couch was facing the Kinect and the width of the couch was centred to it. The placement can be seen in Figure 3.11.

Test 4 - Perpendicular to Kinect

In test 4 there were two couches placed perpendicular to the Kinect, with the armrest nearest to the Kinect being 2.0 m away. This setting can be seen in Figure 3.12.

3.7 Procedure

Our experiment consisted of four test setups and a questionnaire, this section explains the reason of these particular tests and the procedure of each test

Test 1 - Calibration

The purpose of this project was to investigate the identification capabilities of the Kinect. In order to perform this identification we must first calibrate a user, to the system, by measuring the length of different limbs. The length was calculated using the calibration data gathered in this test. These measurements was then saved to a database (See Section 3.5) for later use. It should be noted, that the data collected

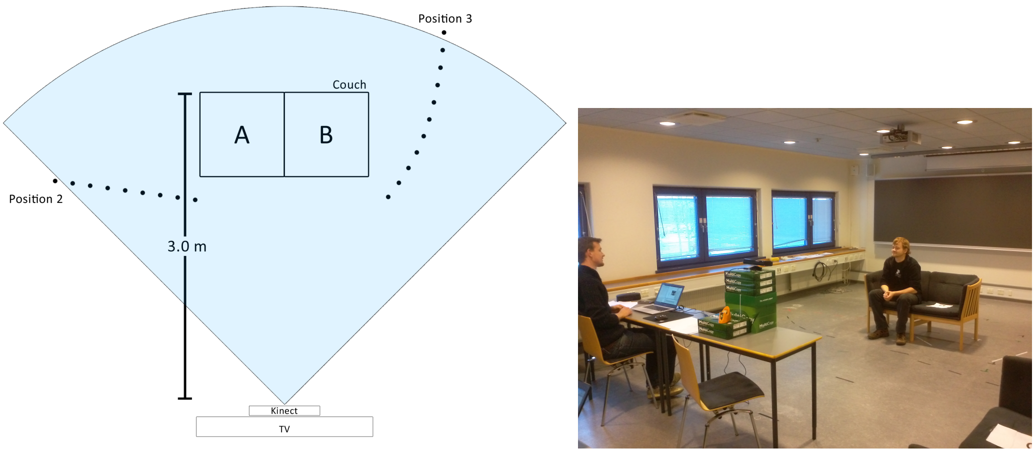


Figure 3.11: The picture to the left shows an overview of test setup 3, the right shows a picture from the experiment.

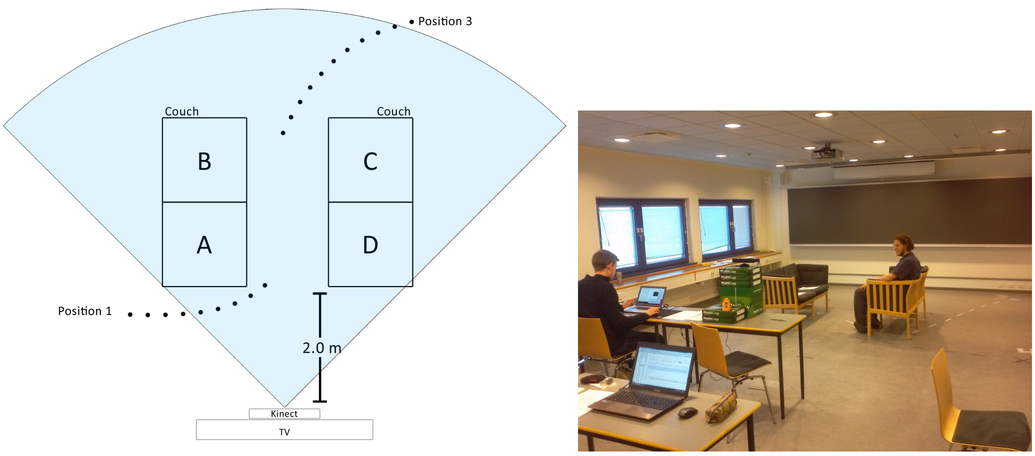


Figure 3.12: The picture to the left shows an overview of test setup 4, the right shows a picture from the experiment.

Figure 3.13: Shows a person performing the calibration tasks.

in this test was not the final calibration of a user, but merely the data foundation for the final calibration, which will be explained in Chapter 4.

In the calibration test the user was placed at a distance of ca. 2.5 m from the Kinect, at which his/her entire body will fit inside the Kinect field-of-view. While on this spot, the user would then be asked to perform some basic movements with his/her arms and legs. The movements performed were as follows (see also Figure 3.13):

1. Slowly raise the arms to the side and subsequently raise them above the head, then take the arms down

2. Kick the feet, so that the entire leg may move

Test 2-4 - Couches

The following three test setups tested the Kinect's ability to identify if there was a user and calculate the skeleton joints of this person. The setups emulated a living room with a couch in front of the TV. The couches varied in position and angle between the different setups. This test showed if and how fast the Kinect, was able to identify and track a user, who was either walking towards or was already seated on a couch. For each test, the test person walked from the two positions to two seats, for example from position 2 to seats A and B and from position 3 to Position A and B. Each task, from position to seat is repeated 5 times. Two times for each seats and while seated, the test person was asked to perform the following tasks:

- Lean slightly forward

- Stretch your arms above your head

- Simulate grabbing an object in front of you

- Simulate grabbing an object to your side, such that you reach over the side of the sofa (If you seated right in the sofa, reach out to your left side)

Tasks for test 2-4 can be seen in Table 3.2.

	Walk	Walk
Test 2 - Corner Couch	2 → B, Sit B	4 → C, Sit C
	2 → B	4 → C
	2 → C	4 → B
Test 3 - Single Couch	2 → A	3 → A
	2 → A, Sit A	
	2 → B	3 → B
	2 → B, Sit B	
Test 4 - Perpendicular	1 → B, Sit B	3 → D, Sit D
	1 → B	3 → D
	1 → D	3 → B

Table 3.2: The different tasks in Tests 2-4. The number denotes the position and the letter denotes the seat.

Also see Appendix C for further details, on how the experiment was conducted.

Questionnaire

Users was also asked to fill out a questionnaire, which asked for; name, age, sex, height, different Kinect info. Furthermore the test person was asked to draw a picture of the couch/TV setup, they had at home. The Questionnaire can be seen here Chapter A.

CHAPTER 4 DATA ANALYSIS

This chapter will present how the data collected during the test was cleaned, what methods are used to analyse it this data, and introduce the technical terms used in this chapter as well as Chapter 5. It should be noted that In this chapter and Chapter 5 the term "calculated skeleton" will be referred to as "skeleton".

4.1 Data Analysis Application

QlikView is a business intelligence (BI) application developed by QlikTech which was used to explore the data collected in the experiment. It can be used to connect to an SQL database or various other types of databases and retrieve data. The application can be used to do a wide range of aggregation and statistics on the data retrieved from a database. It also provides a quick-to use interface which makes it fast to create different kinds of tables and graphs.

4.2 Data Clean-up

This section will describe how the data was cleaned for erroneous entries. The clean-up was divided into three steps:

- Removal of wrong data that was due to human error
- Removal of the Full Height "limb" from the data and results
- Detection and removal of outliers

4.2.1 Erroneous Data

During the experiment, Tests, Tasks, Users, etc. were registered manually by the test leader, occasionally an error would occur and a task would have to be redone,

Skewness for each User, over all Couch Data

User Id	N	Torso	Low-Arm L	Low-Arm R	Shin L	Shin R	Shoulders	Thight L	Thigh R	Up-Arm L	Up-Arm R	Waist
1	534	-0,02	0,10	0,26	-0,54	-0,76	2,27	-0,10	0,15	0,05	0,13	1,70
2	537	1,12	0,98	0,22	-0,92	-0,97	4,03	-0,02	-0,04	0,52	0,14	1,20
3	615	1,07	0,24	0,19	-0,21	-0,36	2,58	-0,59	-0,68	0,08	0,32	1,51
4	590	-0,31	0,66	-0,00	-0,05	-0,03	-0,60	-0,58	-1,01	0,04	-0,26	0,60
5	777	0,41	-0,36	-0,18	-0,88	-1,34	1,63	0,17	0,16	0,06	0,22	0,71
6	845	-0,55	-0,23	-0,26	-0,29	-0,56	-1,04	-1,37	-0,62	-0,57	-0,26	1,80
8	834	-0,09	0,77	0,81	-0,67	-0,54	2,01	-0,15	0,27	0,74	0,71	0,82
9	924	1,00	0,61	0,59	-0,72	-0,81	1,97	0,56	0,79	1,07	0,72	1,00
11	738	0,67	-0,34	-0,29	-1,21	-1,19	1,75	-0,18	-0,10	0,25	0,04	0,64
12	598	0,38	0,34	0,16	-0,84	-0,81	4,20	-0,30	-0,41	0,45	0,40	2,40
13	593	1,56	0,43	0,15	-0,02	-0,21	5,48	0,02	-0,46	0,62	0,52	0,25
14	368	-0,17	0,17	0,18	-1,63	-0,78	2,02	0,05	0,39	0,49	0,27	0,89
15	504	0,69	0,85	2,87	0,77	-0,34	3,45	-0,18	-0,21	2,93	0,19	0,06

Figure 4.1: The skewness of the raw data collected during the Couch tests.

resulting in bad data. These entries have been identified and removed from the data. Furthermore, the Full Height "limb" has also been removed from the results due to erroneous data caused by a programming error in the ICS application. After the clean-up of the data the database contained a total of 13 users, who performed 19 tasks, 18 of which were repeated 5 times, totalling in 793 recordings and 9039 calculated.

4.2.2 Skewness

Data that has been calculated using a variety of sensors can sometimes contain a number of outliers, spikes in values. These outliers are undesirable and should be dealt with. Skewness is a measure of asymmetry in a given dataset. The measure can be both negative and positive. According to Bulmer [1] an absolute skewness between 0.5 and 1 is moderately skewed, an absolute skewness that exceeds 1 indicate a highly skewed distribution. Table 4.1 shows the skewness for each limb for each user. The skewness has been calculated using all the available skeletons to that user. The cells with a red background are those whose absolute value exceeds 0.5. It can be seen that there exist some skewed data, Section 4.3 will discuss how these can be removed.

Calibration data

Calibration data is the data that was collected during the Calibration test. It is of the same type as the data from the couch tests, and as such may also contain skewed data. Table 4.2 shows the skewness calculated for the calibration data, it can be seen that this data is also significantly skewed. It should be noted that it does not make sense comparing skewness between two sets of data since it is based on a number of parameters such as the size of the given dataset, and because skewness is a unit less measure.

Skewness for each User, over all Calibration Data												
User Id	N	Torso	Low-Arm L	Low-Arm R	Shin L	Shin R	Shoulders	Thight L	Thigh R	Up-Arm L	Up-Arm R	Waist
1	49	-4,46	0,00	-0,85	-0,87	-0,98	0,02	-1,84	-0,25	-1,67	-1,79	-0,26
2	37	-2,38	3,10	2,40	2,09	1,88	-2,79	-1,96	-0,86	-0,15	-0,64	-0,99
3	32	-2,58	2,13	2,50	0,90	-0,56	-1,17	0,41	-0,11	-0,55	0,03	-0,34
4	27	0,15	2,36	1,94	0,73	-1,42	-0,44	-0,49	0,97	-0,71	0,39	0,67
5	36	-1,93	1,49	2,32	1,46	2,04	0,54	-0,08	0,46	-0,43	-0,30	1,06
6	39	-1,65	3,78	1,98	1,58	0,83	-0,69	1,44	-0,08	-0,83	-0,62	0,89
8	41	0,13	1,70	1,41	0,56	-2,05	-1,00	-0,33	-0,55	2,08	-0,25	-0,49
9	45	-1,56	1,64	1,40	-0,26	0,45	-1,39	-0,00	-0,35	-0,87	-1,15	1,34
11	47	-1,64	1,75	0,21	0,20	0,44	-1,07	-0,05	-0,08	-0,65	-0,31	1,91
12	64	-3,63	0,44	0,13	-1,06	-1,15	-1,77	-0,66	-1,47	-0,78	-0,54	2,93
13	54	-1,03	2,53	2,21	-0,38	-1,12	-0,18	0,06	-0,48	-0,87	-0,72	0,92
14	67	-1,12	0,33	0,59	-0,08	1,22	-1,09	0,62	0,96	-1,36	-1,28	0,80
15	44	-1,78	1,86	1,92	0,96	-1,32	-1,85	-0,41	-0,28	-1,24	-0,53	0,83

Figure 4.2: The skewness of the raw data collected during the Calibration test.

4.3 Data Reduction

As defined in Section 3.5.1 a recording is a collection of zero or more calculated skeletons. There exists one recording for each repetition of a task. This means there exists a total of 780 recordings for all three couch tests. The number of skeletons is 8457 for all three couch tests. There are 13 recordings from the Calibration test with a total of 582 skeletons. Appendix E shows a sample calculated skeletons, raw data.

In order to make it easier to analyse the data, the skeletons will be aggregated according to their associated recording, this means that instead of 8457 skeletons there will be 780 for the couch tests, and 13 skeletons for the Calibration test. This also establishes a one-to-one relationship between a skeleton and a recording, making them one entity, as such when referring to a recording for the remainder of this chapter, as well as Chapter 5 what is meant is the associated skeleton, which contains medians of the original dataset.

The median has been used as aggregation function. To clarify, the median will find the value separating the bottom half from the top half of the data set. It has been chosen because it is robust against outliers, the value that makes a data set skewed.

4.4 Hit Rate

Hit rate is defined by the percentage of recordings in which the Kinect tracked a person. If there are 20 recordings, during 15 of them the Kinect tracked the user, meaning the hit rate is 75%. The hit rate will always be calculated using recordings, but the results may be aggregated in order to show the hit rate for an entire Test or a User.

4.4.1 Limb Hit Rate

Limb hit rate is defined af the number of times the length of a limb could be calculated versus how many times it could not, to make this possible it required that the associated joints were tracked. Limb lengths are properties of the skeleton, if a limb length is missing it counts as a miss, otherwise it is a hit, tasks where a user was not tracked are not included in this result.

4.5 Accuracy

The accuracy of a measurement is defined as the measurements closeness to its true value. For instance, it is a fact that a cubic meter of water weighs 1000 kg. Our scale weighed it in at 997 kg (the measurement) which is 3 kg from the true value (1000 kg), so the closeness is 3 kg.

4.5.1 Calibration Data

The data collected during the Calibration test will be used as the "true" value in Section 5.2. This value will be used when trying to calculate the average accuracy that can be expected from the data collected during the Couch tests. The true value can also be referred to as the Calibration Median, this is because of the way it is calculated. During the Calibration test the test subject was continuously measured by the ICS application. The data resulting from this test are a lot of recordings and skeletons. The Calibration Median, the "true" value, is the median of all these skeletons, for each user. This means that all 13 users have one median, the true value, for each of the limbs.
There is one disadvantage to using the data from the Calibration test: It is unknown how inaccurate the Kinect really is, and the data from the Calibration test may not reflect reality. This is something that would have to be accounted for in the future, by for example taking manual measurements of test subjects.

4.6 Mean vs. Median

There are several ways of determining the middle value of a dataset, two of which are the mean and median. If a set is symmetrical, e.g. 3, 3, 3, 3, 3 or 1, 2, 3, 4, 5, the median and mean will be the same: 3. As previously stated a median is more robust to outliers, meaning if a high value, e.g. 517, is introduced into a set of numbers like the ones from the previous example, the mean will be heavily influence, the median will not. This is also the reason why the median will be heavily used in Chapter 5 when the choice stands between the mean and the median to display results.

4.7 Prototype Algorithm

This section will introduce a prototype algorithm developed specifically to test if it is possible to recognise a user. 4.7 shows the pseudo-code for this system. The system will have two sets of data in memory at start-up: A set of calibrated users, and an array of 11 constants denoting the accuracy for each limb. The Kinect will at some point provide a measurement of a person, this measurement will act as input to the system. All limbs in this measurement will be converted into a range based on the accuracy for the specific limb, e.g. the Torso was measured to be 35 cm, the Torso accuracy constant is 3, so the Torso range is 32 - 38 cm. The calibration data is then looped, if one of the users are calibrated to a value within this range, that calibrated user receives 1 point. This is done for all limbs that were measured, and the user with the most points is returned, in the event of a draw no user is returned.

```
// The set of calibrated users
Database calibratedUsers = array(U1, U2, ..., Un);

// *, every limb available to the system
// c = the limbs accuracy constant in cm
// Example: Constant TorsoAccuracy = 4;
Constant *Accuracy = c;

Skeleton input = Kinect::Identify();
// Wait until the Kinect has identified and is tracking a user

// Kinect identified and is tracking a user
// for each limb that appears in the measurement
foreach(input as limb)
{
    min = limb->length - limbAccuracy;
    max = limb->length + limbAccuracy;

    foreach(calibratedUsers as user)
    {
        if(user.limb > min AND user.limb < max)
            // It was a match
            user.count++; // All limbs are weighted equally
    }
}

// When all limbs have been looped
// return the user user with highest count
return highestCount(user);
```

Section 5.2.3 will use this algorithm through several tests by selecting a subset of the test subjects from the experiment as the calibrated users. One of these users will then be chosen to be the measured user. The measurements will based on a median for each limb based on this users entire set of skeletons.

CHAPTER 5

FINDINGS

This chapter will provide a description and interpretation of the data collected during the experiment. The focus will mainly be on hit rate, accuracy and the ability to recognise a user.

The data analysed in this section is based on 13 test subjects performing four tasks for each of the three couch tests, repeating each task five times totalling 780 recordings for all users.
It should be noted that in this chapter the term "test subject" and "user" are used synonymously, as are "calculated skeleton" and "skeleton". It should also be noted that User 7 is missing from the results because he did not show up for the test, but had already been registered in the database.

5.1 Hit Rate

Hit rate is defined by the percentage of recordings in which the Kinect tracked a person. In order for the Kinect to recognise a person it has to track him, this is however not always possible. This section will investigate the hit rates for the different setups, the different users, and each of the limbs.

5.1.1 Couch placement

Figure 5.1 provides an overview of how well the Kinect performed during the three couch tests, with regard to hit rate. The hit rate for each test was calculated based on 260 recordings for each of the Couch tests. The highest scoring test was the Single Couch (93.8%), and Corner Couch the lowest (85.4%). A difference of 8.4% means that in the Single Couch test the Kinect managed to track 22 task repetitions more than the Corner Couch. Across the entire experiment the Kinect had a hit

Hit Rate per Test		
Test	**N**	**Hit rate**
Corner Couch	260	85,4%
Perpendicular to Kinect	260	87,7%
Single Couch	260	93,8%
Total	**780**	**89,0%**

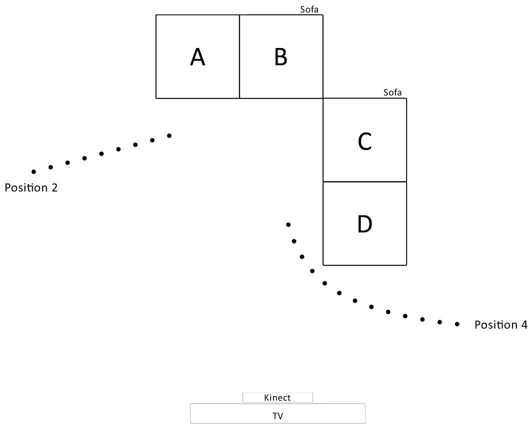
Figure 5.1: The hit rate for all Couch Tests based on 260 recordings for each of the setups.

Hit Rate per Task			
Test	**Task**	**N**	**Hit rate**
Corner Couch	2 -> B	39	59,0%
	2 -> B, Sit B	26	65,4%
	4-> C, Sit C	26	92,3%
	4 -> C	39	92,3%
	2 -> C	65	93,8%
	4 -> B	65	93,8%
Perpendicular to Kinect	1 -> D	65	64,6%
	1 -> B	39	87,2%
	1 -> B, B	26	88,5%
	3 -> D	39	97,4%
	3 -> B	65	100,0%
	3 -> D, D	26	100,0%
Single Couch	2 -> B	39	69,2%
	2 -> B, Sit B	26	92,3%
	2 -> A	39	94,9%
	2 -> A, Sit A	26	100,0%
	3 -> A	65	100,0%
	3 -> B	65	100,0%

Figure 5.2: The hit rate for all Tasks in their respective Tests.

rate of 89%.

Each setup consisted of six tasks, Figure 5.2 gives an overview of how these tasks performed. There is a total of 18 tasks, 12 of those scored a Hit Rate above 90%. Four tasks scored a hit rate below 85%, while five tasks scored a hit rate of 100%, which means that all users were tracked in all repetitions of these tasks.
The tasks 2 → B together with 2 → B, Sit B in the Corner Couch test, was the ones who performed the worst, dragging the overall Hit Rate down below 90% for that test. This is also the case with 2 → B in the Single Couch test, although 2 → B, Sit B scored considerably higher in this test compared to Corner Couch. See Figure 5.3 for the corner couch and single couch setup.



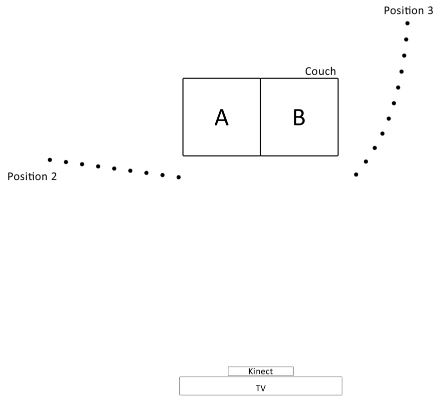


Figure 5.3: To the left is test 2 corner couch and to the right is task 3 single couch.

Hit rate per User per Test				
User Id	Corner Couch (N=20)	Single Couch (N=20)	Perpendicular to Kinect (N=20)	Total (N=60)
1	90,0%	85,0%	90,0%	88,3%
2	85,0%	85,0%	95,0%	88,3%
3	95,0%	100,0%	80,0%	91,7%
4	95,0%	95,0%	85,0%	91,7%
5	85,0%	95,0%	90,0%	90,0%
6	100,0%	100,0%	100,0%	100,0%
8	100,0%	100,0%	100,0%	100,0%
9	95,0%	100,0%	95,0%	96,7%
11	85,0%	95,0%	100,0%	93,3%
12	75,0%	100,0%	90,0%	88,3%
13	95,0%	90,0%	75,0%	86,7%
14	30,0%	80,0%	50,0%	53,3%
15	80,0%	95,0%	90,0%	88,3%
Mean	**85,4%**	**93,8%**	**87,7%**	**89,0%**

Figure 5.4: The hit rate for each User in each Couch Test, a total hit rate of 89%

5.1.2 Users

This section will focus on the hit rate each user achieved. Figure 5.4 shows how the Kinect performed on each User for each Test. The Total hit rate varies between the lowest of 53.3% for User 14, and the highest at 100% for Users 6 and 8. User 14 scored the lowest hit rate in all the three tests at 30%, 80% and 50% for the Corner Couch, Single Couch and Perpendicular Couch setup respectively.

Why a certain couch setup or a User receives better or worse results than others, can be attributed to many different parameters, with origins in how the Kinect

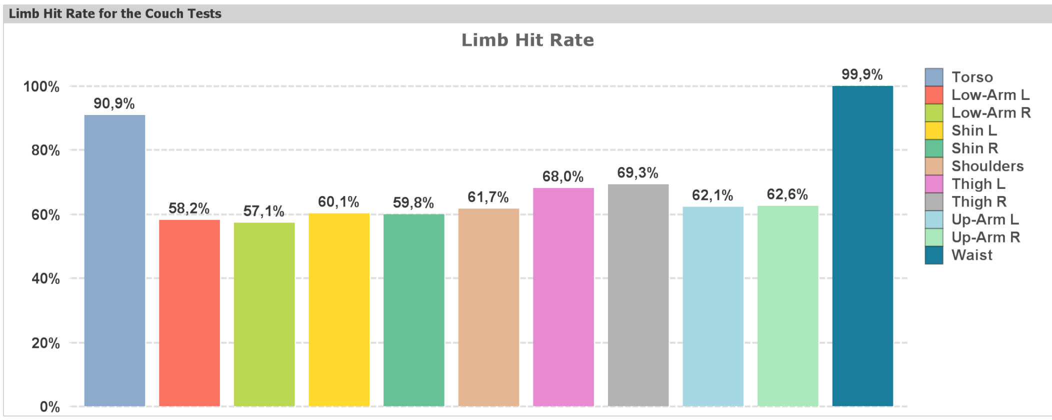


Figure 5.5: The hit rate of each limb excluding the results from the Calibration test.

identifies a person. This matter will be discussed in Chapter 6.

5.1.3 Limb Hit Rate

Figure 5.5 shows the hit rate for each limb across all tests and users (excluding the Calibration test). It is interesting to see which limbs are calculated more often than others, because this gives an idea of which limbs could be useful when trying to recognise a person. The top-four candidates are Torso Height (90.9%), Waist Width (99.9%) and Thigh Right/Thigh Left (69.3%, 68% respectively). The difference between the highest and lowest scoring of the remaining limbs is 5.5%. The top-four limbs and their accuracy will be highlighted Section 5.2.

For comparison, the Calibration test had significantly better results, as can be seen in Figure 5.6, as was also intended. The lowest scoring limb was the lower left arm at 93.8%, and the waist width being the highest at 100%. As expected these results have a high hit rate, because the user was positioned in front of the Kinect and only made small movements.

5.2 Accuracy

As explained in Section 4.5 accuracy is defined as the closeness a measurement has to its true value, in this section the measurements will come from the couch tests and the true value will be based on data from the Calibration test.

5.2.1 Calibration data

Figure 5.7 shows an example of what calibration data looks like, this example shows that of User 5. As previously mentioned, calibration data is based on the recordings made during the Calibration test. It should be understood that these values are not

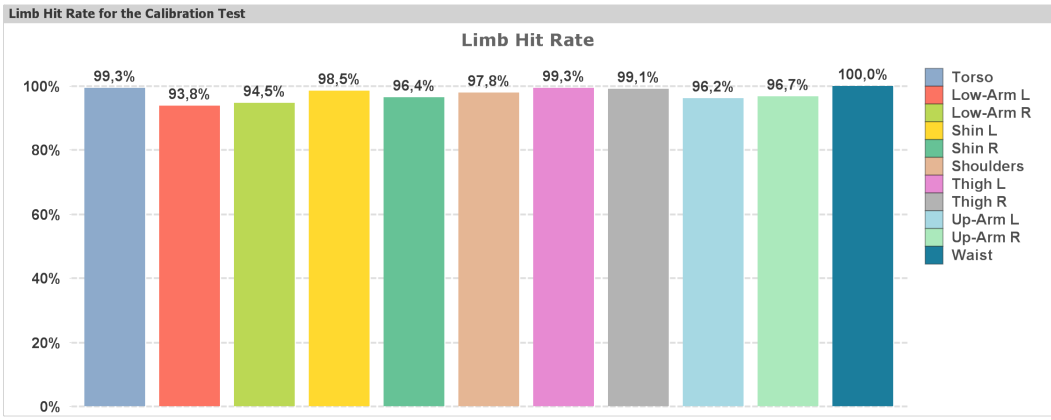


Figure 5.6: The hit rate of each limb including only the results from the Calibration test.

Calibration Data for User 5	
Limb	**Length (cm)**
Height Torso	65,3
Lower Arm Left	25,6
Lower Arm Right	25,5
Shin Left	42,0
Shin Right	41,2
Shoulder Width	44,8
Thigh Left	61,6
Thigh Right	60,9
Upper Arm Left	29,4
Upper Arm Right	29,7
Waist Width	17,6

Figure 5.7: The calibration of User 5, this is was is used as the true value.

the size of the body in the real world, for example user 5's waist is calibrated to 17.6 cm, but in the real world user 5 has a much larger waist.

5.2.2 Couch test data accuracy

This section will investigate the accuracy of the recordings made during the couch tests.

Figure 5.8 shows a table with a selection of five users, they give a glimpse of the tendency the full set of users have. It displays the accuracy of the Torso measurements made during the couch tests. The number of elements the mean and median are based on varies due to a varying hit rate between the users. The minimum mean and median are 3.4 cm and 2.5 cm respectively. There is a clear tendency of the median being smaller than the mean, 14 out of 15 times the median is lower, which is better because it means a higher accuracy.

Torso Closeness for User				
User Id	**Test**	**N**	**Mean (cm)**	**Median (cm)**
1	Corner Couch	18	6,7	4,9
	Perpendicular to Kinect	18	7,4	4,8
	Single Couch	17	3,4	2,5
3	Corner Couch	19	8,0	6,6
	Perpendicular to Kinect	16	10,1	5,7
	Single Couch	20	7,3	6,5
6	Corner Couch	20	6,5	4,0
	Perpendicular to Kinect	20	6,5	3,8
	Single Couch	20	5,0	4,7
8	Corner Couch	20	6,2	6,0
	Perpendicular to Kinect	20	6,2	2,2
	Single Couch	20	6,6	6,4
12	Corner Couch	15	4,4	3,2
	Perpendicular to Kinect	18	6,4	4,9
	Single Couch	20	7,1	7,4

Figure 5.8: The mean and median accuracy of the Torso, for each of the three couch tests. N denotes the number of recordings the aggregation is based on, it varies due to the hit rate.

Closeness (in cm) per User											
User Id	**Torso**	**Low-Arm L**	**Low-Arm R**	**Shin L**	**Shin R**	**Shoulders**	**Thigh L**	**Thigh R**	**Up-Arm L**	**Up-Arm R**	**Waist**
1	4,4	1,7	2,7	3,7	4,0	3,5	6,9	6,3	2,1	2,1	1,2
2	5,0	1,5	1,3	1,5	2,6	2,4	12,0	11,8	2,2	1,9	1,3
3	6,2	1,1	1,3	1,4	2,6	1,7	7,8	6,5	1,9	2,1	1,2
4	3,7	1,1	1,4	2,5	8,2	1,5	4,9	3,1	2,3	2,2	1,0
5	6,9	1,8	1,3	2,3	1,5	2,3	9,3	9,3	3,2	3,2	1,2
6	4,3	2,1	1,3	2,1	1,8	3,7	7,3	6,5	3,8	2,2	0,9
8	5,7	1,0	1,0	1,5	2,2	1,6	9,7	7,9	1,5	2,2	0,9
9	5,1	1,7	1,7	2,4	2,1	1,8	10,0	11,1	2,6	4,4	1,6
11	7,4	2,0	1,8	1,5	1,4	4,8	11,3	11,0	3,1	2,9	1,4
12	5,5	1,6	1,0	1,2	1,7	2,1	10,5	10,0	1,6	2,2	1,4
13	4,2	1,4	1,4	3,7	3,5	1,4	10,0	10,5	4,4	3,3	1,4
14	7,6	1,0	1,2	1,6	1,5	4,1	7,3	9,2	3,2	2,0	1,0
15	7,7	1,7	1,5	2,4	2,1	2,2	2,5	4,5	2,3	2,3	1,2
Total	**5,5**	**1,6**	**1,4**	**1,9**	**2,3**	**2,4**	**8,4**	**8,0**	**2,4**	**2,5**	**1,2**

Figure 5.9: The median accuracy for each user and limb. Also shown is the "total" median, the median for all recordings made during the couch test.

Because of the tendency in Figure 5.8, that the median shows higher accuracy, the median is also being used in Figure 5.9. This is theoretically the same data as the previous figure, just aggregated over a larger dataset. This figure shows the median accuracy for all users and for all limbs. It also shows the total accuracy calculated based on the set of all 780 recordings, made during the couch tests.

What can be derived from this figure is that, when using the median to present an overview of the accuracy 8 out of 11 limbs has an accuracy of 2.5 cm or better,

Accuracy Constants										
Torso	Low-Arm L	Low-Arm R	Shin L	Shin R	Shoulders	Thigh L	Thigh R	Up-Arm L	Up-Arm R	Waist
5,5	1,6	1,4	1,9	2,3	2,4	8,4	8,0	2,4	2,5	1,2

Figure 5.10: Accuracy constants for all limbs.

furthermore it is seen that torso and both thighs are the most inaccurate. Using the median to aggregate the data only presents one view, Section 4.6 explains why the median was chosen.

5.2.3 Recognising a Person

This section will investigate if it is possible, using the prototype algorithm in Section 4.7, to recognise a person based on a subset of the test subjects who participated in the experiment. Each test will have 3-5 users "calibrated" to the system, one of these users will then be chosen to act as a measurement. The accuracy constants for each limbs can be seen in Figure 5.10, they are the same as the Total medians calculated in Figure 5.9.

Test 1 - Four Chosen at Random

- *Calibrated Users:* 1, 3, 5, 8
- *Measured User:* 3
- *Recognised User:* 3

Figure 5.11 and Figure 5.12 shows the calibrations and measurement, respectively. Table 5.1 shows the resulting comparison. In this test, the correct user was recognised with a total of 8 points, 3 points ahead of the runner up.

Calibration Data (in cm)											
User Id	Torso	Low-Arm L	Low-Arm R	Shin L	Shin R	Shoulders	Thigh L	Thigh R	Up-Arm L	Up-Arm R	Waist
1	62,4	23,7	27,7	39,4	39,0	37,1	52,8	52,9	25,4	25,9	13,6
3	63,0	23,1	29,8	38,2	38,8	38,7	53,1	52,5	25,2	25,5	14,6
5	65,3	25,6	28,1	42,0	41,2	44,8	61,6	60,9	29,4	29,7	17,6
8	65,7	23,3	30,4	37,0	39,5	43,7	53,8	52,9	24,8	25,9	15,8

Figure 5.11: Calibrated Medians for Users 1, 3, 5, 8

Range for Measurement based on Accuracy Constant										
Torso	Low-Arm L	Low-Arm R	Shin L	Shin R	Shoulders	Thigh L	Thigh R	Up-Arm Left	Up-Arm R	Waist
48,6 - 59,6	22,3 - 25,5	21,6 - 24,4	34,5 - 38,3	33,3 - 37,9	35,4 - 40,2	36,4 - 53,2	36,7 - 52,7	20,7 - 25,5	20,5 - 25,5	13,4 - 15,8

Figure 5.12: Measurement for User 3

User Id	*Total*
User 1	5
User 3	8
User 5	0
User 8	4

Table 5.1: The results from running the prototype algorithm on users 1, 3, 5 and 8, with user 3 as the measured user. The resulting recognised user is user 3, the correct one.

Test 2 - Four Tallest People

- *Calibrated Users:* 5, 9, 13, 14
- *Measured User:* 9
- *Recognised User:* None

Figure 5.13 and Figure 5.14 shows the calibrations and measurement, respectively. Table 5.2 shows the resulting comparison. In this test no user was recognised because three of them scored an equal amount of points, the measured user scored the least points.

Calibration Data (in cm)											
User Id	Torso	Low-Arm L	Low-Arm R	Shin L	Shin R	Shoulders	Thigh L	Thigh R	Up-Arm L	Up-Arm R	Waist
1	62,4	23,7	27,7	39,4	39,0	37,1	52,8	52,9	25,4	25,9	13,6
3	63,0	23,1	29,8	38,2	38,8	38,7	53,1	52,5	25,2	25,5	14,6
5	65,3	25,6	28,1	42,0	41,2	44,8	61,6	60,9	29,4	29,7	17,6
8	65,7	23,3	30,4	37,0	39,5	43,7	53,8	52,9	24,8	25,9	15,8

Figure 5.13: Calibrated Medians for Users 5, 9, 13, 14

Range for Measurement based on Accuracy Constant										
Torso	Low-Arm L	Low-Arm R	Shin L	Shin R	Shoulders	Thigh L	Thigh R	Up-Arm Left	Up-Arm R	Waist
48,6 - 59,6	22,3 - 25,5	21,6 - 24,4	34,5 - 38,3	33,3 - 37,9	35,4 - 40,2	36,4 - 53,2	36,7 - 52,7	20,7 - 25,5	20,5 - 25,5	13,4 - 15,8

Figure 5.14: Measurement for User 9

User Id	*Total*
User 5	4
User 9	3
User 13	4
User 14	4

Table 5.2: The results from running the prototype algorithm on users 5, 9, 13 ,14 with user 9 as the measured user. No user was recognised because of a tie between highest scoring users.

Test 3 - Three Users with Highest Overall Hit Rate

- *Calibrated Users:* 6, 8, 9
- *Measured User:* 6
- *Recognised User:* 6

Figure 5.15 and Figure 5.16 shows the calibrations and measurement, respectively. Table 5.3 shows the resulting comparison. In this test, the correct user was recognised with a total of 5 points, 3 points ahead of the runner up.

Calibration Data (in cm)											
User Id	Torso	Low-Arm L	Low-Arm R	Shin L	Shin R	Shoulders	Thigh L	Thigh R	Up-Arm L	Up-Arm R	Waist
6	58,9	22,3	28,4	35,6	36,5	38,5	52,2	52,8	27,5	26,5	13,7
8	65,7	23,3	30,4	37,0	39,5	43,7	53,8	52,9	24,8	25,9	15,8
9	66,5	25,6	32,5	40,9	41,0	43,0	61,0	61,9	29,4	30,7	16,1

Figure 5.15: Calibrated Medians for Users 6, 8, 9

Range for Measurement based on Accuracy Constant										
Torso	Low-Arm L	Low-Arm R	Shin L	Shin R	Shoulders	Thigh L	Thigh R	Up-Arm Left	Up-Arm R	Waist
46 - 57	20,5 - 23,7	20,3 - 23,1	32 - 35,8	32,4 - 37	32,3 - 37,1	35,5 - 52,3	35,7 - 51,7	20 - 24,8	20,5 - 25,5	12,7 - 15,1

Figure 5.16: Measurement for User 6

User Id	*Total*
User 6	5
User 8	2
User 9	0

Table 5.3: The results from running the prototype algorithm on users 6, 8, 9 with user 6 as the measured user. The resulting recognised user is user 6, the correct one.

Test 4 - Five Shortest People

- *Calibrated Users:* 2, 3, 4, 6, 12
- *Measured User:* 2
- *Recognised User:* 4

Figure 5.17 and Figure 5.18 shows the calibrations and measurement, respectively. Table 5.4 shows the resulting comparison. In this test, the correct user was not recognised, User 4 was recognized with a total of 9 points, 2 points ahead of the runner up, and 5 points ahead of the correct user.

Calibration Data (in cm)											
User Id	Torso	Low-Arm L	Low-Arm R	Shin L	Shin R	Shoulders	Thigh L	Thigh R	Up-Arm L	Up-Arm R	Waist
2	60,6	22,4	28,5	37,0	39,3	40,1	54,6	53,9	26,0	26,7	15,5
3	63,0	23,1	29,8	38,2	38,8	38,7	53,1	52,5	25,2	25,5	14,6
4	57,6	21,7	26,0	37,8	43,0	40,2	46,1	41,6	24,9	25,8	14,8
6	58,9	22,3	28,4	35,6	36,5	38,5	52,2	52,8	27,5	26,5	13,7
12	62,3	21,7	27,0	36,8	37,5	39,9	56,1	55,5	24,7	25,0	15,0

Figure 5.17: Calibrated Medians for Users 2, 3, 4, 6, 12

Range for Measurement based on Accuracy Constant										
Torso	Low-Arm L	Low-Arm R	Shin L	Shin R	Shoulders	Thigh L	Thigh R	Up-Arm Left	Up-Arm R	Waist
48,5 - 59,5	20,8 - 24	21,3 - 24,1	34,5 - 38,3	33,8 - 38,4	36,3 - 41,1	34,6 - 51,4	34,2 - 50,2	20,8 - 25,6	21,6 - 26,6	13,8 - 16,2

Figure 5.18: Measurement for User 2

User Id	*Total*
User 2	4
User 3	6
User 4	9
User 6	6
User 12	7

Table 5.4: The results from running the prototype algorithm on users 2, 3, 4, 6, 12 with user 2 as the measured user. The recognized user, User 4, was not the correct one.

5.3 Questionnaire Results

This section will present the drawings which were provided by the test subject in the questionnaire. The drawings can be seen in Appendix B. The drawings resemble two of the test setups in the experiment namely corner couch and single couch. The drawings that resemble corner couch are the ones from user 1, 4, 8, 9, 11, 13 and 15. User 3's drawing also resembles corner couch, the TV is however placed to the right of the couch. Furthermore there are indications that at least some of the couches have a chaise longue for example user 4, 8 and 15, also when asked user 11 said his couch had one. This is however note stated in the drawing. The drawings that resemble single couch are the ones from user 2, 6, 12 and 14. From the drawings it can be seen, that the estimated distances from the couch to the TV vary between 1.5 to 4.0 m. Furthermore it can be seen that 9 of the 12 drawings includes a table, which is mostly placed between the couch and the TV. 10 of the 12 drawings has the route used by the test person to the couch. Three of these show a route similar to the one from position 2 in the corner couch setup, these are the ones from user 1, 2 and 13, user 15's route resembles the route from position 4. The rest of the drawings show a route from behind the TV, these are the ones from user 2, 3, 6, 8 ,9 and 14.

CHAPTER 6 DISCUSSION

This chapter will discuss the findings from the experiment.

6.1 Hit Rate

Firstly the Kinect hit rates will be discussed.

Hit Rate per Test Setup

In the overview of the couch setups it can be seen, that test setup 2 corner couch and 4 perpendicular couch score a lower hit rate than test 3 single couch. The reason for this could be, that those two setups have couches, that are perpendicular to the Kinect, which perhaps makes it harder for the Kinect to detect a person and thereby calculate a skeleton. But since the corner couch setup has a couch facing towards the Kinect, like the single couch, and the single couch scores the highest hit rate, it was interesting to see, that corner couch setup scores the lowest hit rate, a reason for this could be, that at 4.0 m the couch was to far away for the Kinect, to get the best possible work conditions. The reason why the single couch gets the highest score was most likely, that it at 3.0 m away it was placed at an optimal position from the Kinect, but also, that the couch was facing the Kinect. Furthermore this could indicate, that the placement of the Kinect affects the hit rate.

Hit Rate per Task

When looking at the hit rate for each task the most obvious result is, that when the test person was asked to do the sit tasks the hit rate is higher, than when not doing the sit tasks. The reason for this could be, that the test person uses more time in front of the Kinect which means, that it has a greater chance of tracking this person.

Furthermore when looking at corner couch tasks 2 → B and 2 → B sit and single couch task 2 → B it can be seen, that they have a low hit rate. The reason for this could be, that the short distance affected the results negatively, but when looking at single couch 2 → A, which has an even shorter distance, similar result would be expected, this was not the case however. This could indicate, that there were a difference in the way the test subjects sat in seat A and seat B.

Four out of five tasks, that started from position 3 had a hit rate of 100%. The reason for this could be both the amount of time spent in front of the Kinect, but also, that the test subjects walked more or less directly towards the it. Furthermore when looking at tasks, that started from position 4 these also show high hit rates, 92.3% or above, the reasons for this could be the same as for position 3, but here the test subject walked away from the Kinect, which could explain the difference. Both these results could indicate, that the placement of the Kinect, to the test person, affects the hit rate.

Hit Rate per User per Setup

When looking at the hit rate per user, per setup It can be seen, that user 6 and 8 score 100% which means, that in every task they performed, there was a calculated skeleton, however, it does not tell us how accurate these skeletons are. This was also true for the rest of the measurements. User 14 had the lowest amount of calculated skeletons and the reason for this was not known, it could be because of the clothes he was wearing or maybe the shape of the body made it difficult for the Kinect to track.

Limb Hit Rate

If the limb hit rate of the experiment is compared to the one from the calibration it is clearly seen, that the later scores the highest across limbs, this could indicates, that it is far easier for the Kinect to track the user and his limbs, but also that the calibration can be used as the true value.

Furthermore when looking at the hit rate of the limbs in the experiment it is seen, that torso height and waist width has the highest hit rate. This could mean, that these two measurements could be used as primary measurements and possible have more weight in an algorithm. But since this is the hit rate, it is not known if these would also be more accurate than the limbs, that have a lower score. Both left and right thigh score a little higher than the rest of the measurements, the reason for this could be, that since the waist width has such a high hit rate, this rubs of on the thigh hit rate because one of the joints used in the calculation of the thigh, was the same as in waist width.

Likewise it is seen, that both upper arms have around 4-5% higher hit rate than the lower arms the reason for this could be, that the upper extremities, both upper arms and thighs are more stationary when in movement, which could give the Kinect better opportunity to track them.

6.2 Accuracy

When looking at the accuracy of the different limbs it can be seen, that 8 out of 11 limbs scored an accuracy of 2.5 cm or better (better being closer to 0), which is better than expected considering all the data was collected with the doing a wide range of different movements. The remaining three limbs, torso, left thigh and right thigh all scored an accuracy worse than 5 cm. Interestingly enough of the four limbs that scored the highest hit rates, torso, left thigh, right thigh were number 2, 3 and 4 respectively. This could indicate, that a high hit rate is not equal to good results nor the opposite, because the waist, scored highest in hit rate but also had the best accuracy at 1.2 cm.

An explanation for why the torso is so inaccurate (5.5 cm) despite of its high hit rate may be attributed to the fact that its value is based on a total of four different joints. Because of this, the high inaccuracy is not completely unexpected because each joint will be subject to a certain amount of inaccuracy, and having a limb consisting of four joints will most likely have a larger inaccuracy than a limb of only two joints.

The thighs were the limbs with the worst accuracy (left: 8.4 cm, right: 8.0 cm), we attribute this inaccuracy to the fact that during the experiment the users spent a considerable amount of time seated in the couch, furthermore 12 out of the 60 repetitions involved the test person sitting in the couch over a longer period doing sit tasks. It is a known fact that the Kinect has problems accurately tracking a user when he or she is sitting down, which could explain the large inaccuracy.

Besides measuring the accuracy we also tested the possibility of recognising a person using a simple prototype recognition algorithm created specifically for this purpose. The algorithm simply checked a calibrated user was within range of the measured limb, if so, that user was given a point. The user with the most points was accepted as the measured user. Of the four tests, 1-4, two of them, test 1 and 3 recognised the correct user. Table 6.1 shows the results from performing the accuracy tests. There is no definitive connection between why a test was successful or why it was not, it does however prove that it was possible to get a correct identification:

- In test 1 three of the users are of fairly evenly height, this includes the measured user, and the test succeeded by recognising the correct user.
- In test 3 the three users were of very different height, and the test succeeded by recognising the correct user.
- In test 2 three of the users were of fairly even height, this includes the measured user, and the test failed.
- In test 4 three of the users were of fairly even height, this includes the measured user, and the test failed by recognising the wrong user, the user farthest away when comparing height.

	User	Hit Rate	Height (cm)	Score
Test 1	1	88.3%	173	5
	3	**91.7%**	**170**	**8**
	5	90%	194	0
	8	100%	175	4
Test 2	5	90%	194	4
	9	**96.7%**	**196**	**3**
	13	86.7%	192	4
	14	53.3%	185	4
Test 3	**6**	**100%**	**164**	**5**
	8	100%	175	2
	9	96.7%	196	0
Test 4	**2**	**88.3%**	**170**	**4**
	3	91.7%	170	6
	4	91.7%	158	9
	6	100%	164	6
	12	88.3%	170	7

Table 6.1: The results from performing the accuracy tests from Section 5.2.3 along with each users hit rate from Figure 5.4. The rows in bold are the users who were measured.

It was expected there would be some sort of correlation between the height and the possibility of recognising the correct user, because the height of a person has an impact of the size of his limbs. There is however no definitive proof that the height is a decisive parameter.

6.3 Questionnaire Results

When looking at the participants in the experiment, see Appendix B it can be seen that all of the participants are in their twenties, this could mean, that the data collected for these participants would not be the same as for younger or older people. For example kids would most likely be smaller than most of the test persons in the experiment, this could maybe have some affect on the results. Also elderly could have some affect on the measurements, maybe they move more slowly towards the couch giving the Kinect longer time to track the skeleton, but if the elderly person had a hunched back it would be harder for the Kinect to calculate for example the torso height.

When looking at the hight were the shortest test person was 1.58 m and the tallest being 1.96 m, this seems to be a decent spread, at least when looking at the adult population.

There were 5 of the 13 test persons who had used a Kinect before, but since none of them had a Kinect at home it was not believed that the test subject was to affected by this fact. If the test persons had used an Kinect more often it could maybe have affected the results, because they would have an idea about, how the Kinect tracked a person and therefore changed their behaviour accordingly in order to get better tracking.

When looking at the drawings of the test persons living rooms it is seen that most of them look similar to the two test setups, corner couch and single couch, in the experiment. This could mean that the test setups, that were chosen, are fairly lifelike. However in hindsight the setups should probably have been chosen based on a preliminary questionnaire.

CHAPTER 7 CONCLUSION

This report set out to answer the following question:

> "Is it possible, using the Microsoft XBox Kinect, to uniquely recognise a person based on the skeleton data provided by the Windows Kinect SDK?"

On the basis of the experiment and the subsequent findings we can conclude that it is indeed possible to recognise a user based on the skeleton data provided by the XBox Kinect. The findings did however not offer any definitive evidence as to why two of the performed recognition tests succeeded and why two of them failed. There is some data that suggests height may be involved in how often it is possible to recognise a user, two of the tests were done on people with similar height and they both failed, a third test was done on a set of people with different height, this test succeeded. The last test was done on a set of people with similar height, this test also succeeded, against our expectations.

Furthermore we can conclude that the overall hit rate of the Kinect was 89%. The hit rate is imperative in the efforts of recognising a person since the ability to track a persons limbs is basis for the actual recognition. There are also strong indication that the placement of the Kinect in relation to the person we wish to track has an impact on the hit rate. Based on the findings it can be concluded that the hit rate does not have any apparent impact on the accuracy of the measurements.

Finally we can conclude that 8 out of 11 limbs can be measured with an accuracy of 2.5 cm or better. The only limbs to have worse accuracy, are the torso and thighs. The low accuracy of the torso measurement is most likely due to it being calculated based on four joints, instead of two as the remaining limbs. The inaccuracy of the thigh measurements can most likely be attributed to the Kinects inability to properly track the associated joints when a person is in a seated position, as would be expected.

CHAPTER 8 10TH SEMESTER

This chapter will outline what will explored in the 10th semester project.

Since no filters were applied to the Kinect data in the present study, it would be interesting to explore the different filter in order to see, if this would stabilize the data and lessen the problem of outliers. One of the filters that could be used is *Smoothing* which makes the joints less jittery, this could mean that accuracy of the calculated limbs could become higher, making it more likely that a person can be identified. It could also be possible to apply multiple filters to improve the data. The issue of using filters will be explored in February.

In March when the best filter(s) has been found, the hit rate will be analysed again to see if the filters has made a difference. Furthermore since some tasks showed lower hit rates than others, by exploring the video recorded from the Kinect it could be possible to find some common denominator, that explains these findings, for example the way people move towards a couch, and how they sit in their seats.

Based on the findings of applying filters, an algorithm will be developed that has a high success rate in recognising people. This algorithm implemented into a system that also supports proxemic interaction, which will make it possible to further explore the identify dimension as an interaction form. Parts of the system will be developed through February and March but most of the development will be in April and the first two weeks of May. In the last two weeks of May different proxemic interaction forms will be explored in the lab, with identification as the main focus. Furthermore in order to get the best possible conditions for the experiment the analysis of the hit rate will help, with the best placement of furniture.

BIBLIOGRAPHY

[1] Bulmer, M. G. (1979). *Principles of Statistic*. Dover Publications, Inc., Mineola, 2th edition.

[2] Danmarks Statistik (2013). Husstande 1. januar efter omrde, tid og hustandsstørrelse. [online, accessed 10 Jan. 2013] `http://www.dst.dk/da/`.

[3] Galatas, G., Potamianos, G., and Makedon, F. (2012). Audio-visual speech recognition incorporating facial depth information captured by the kinect. In *20th European Signal Processing Conference*, pages 2714–2717, Bucharest, Romania. EUSIPCO 2012 (European Signal Processing Conference).

[4] González-Rodríguez, J., Toledano, D. T., and Ortega-García, J. (2008). *Handbook of Biometrics*, chapter Voice Biometrics, pages 151–170. Springer US.

[5] Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R., and Wang, M. (2011). Proxemic interactions: The new ubicomp? *Interactions*, XVIII:42–50.

[6] Hall, E. T. (1966). *The Hidden Dimension*. Doubleday, New York, 2nd edition.

[7] Heckathorn, D. D., Broadhead, R. S., and Sergeyev, B. (2001). A methodology for reducing respondent duplication and impersonation in samples of hidden populations. *Journal of Drug Issues*, 31(2):543–564.

[8] Jafri, R. and Arabnia, H. R. (2009). A survey of face recognition techniques. *Journal of Information Processing Systems*, 5(2):41–68.

[9] Jain, A., Ross, A., and Prabhakar, S. (2004a). An introduction to biometric recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):4 – 20.

[10] Jain, A. K., Dass, S. C., and Nandakumar, K. (2004b). Soft biometric traits for personal recognition systems. In Zhang, D. and Jain, A. K., editors, *Biometric*

Authentication, volume 3072 of *Lecture Notes in Computer Science*, pages 731–738. Springer Berlin Heidelberg.

[11] Kastanis, I. and Slater, M. (2012). Reinforcement learning utilizes proxemics: An avatar learns to manipulate the position of people in immersive virtual reality. *ACM Trans. Appl. Percept.*, 9(1):3:1–3:15.

[12] Khoshelham, K. and Elberink, S. O. (2012). Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454.

[13] Knapp, M. L. and Hall, J. A. (2010). *Nonverbal Communication in Human Interaction*. Wardsworth, Centage Learning, Boston, 7th edition.

[14] Leyvand, T., Meekhof, C., Wei, Y.-C., Sun, J., and Guo, B. (2011). Kinect identity: Technology and experience. *Computer*, 44(4):94 –96.

[15] Marquardt, N. and Greenberg, S. (2011). Informing the design of proxemic interactions. *Pervasive Computing, IEEE*, 11(2):14–23.

[16] Microsoft. Depth stream. [online, accessed 17 Dec. 2012] `http://msdn.microsoft.com/en-us/library/jj131028.aspx`.

[17] Microsoft. Depthimageformat enumeration. [online, accessed 17 Dec. 2012] `http://msdn.microsoft.com/en-us/library/microsoft.kinect.depthimageformat.aspx`.

[18] Microsoft. Joint filtering. [online, accessed 17 Dec. 2012] `http://msdn.microsoft.com/en-us/library/jj131024.aspx`.

[19] Microsoft. Kinect for windows sdk to arrive spring 2011. [online, accessed 17 Dec. 2012] `http://blogs.technet.com/b/microsoft_blog/archive/2011/02/21/kinect-for-windows-sdk-to-arrive-spring-2011.aspx`.

[20] Microsoft. Kinect for windows sensor components and specifications. [online, accessed 17 Dec. 2012] `http://msdn.microsoft.com/en-us/library/jj131033.aspx`.

[21] Microsoft. Near mode: What it is (and isnt). [online, accessed 17 Dec. 2012] `http://blogs.msdn.com/b/kinectforwindows/archive/2012/01/20/near-mode-what-it-is-and-isn-t.aspx`.

[22] Microsoft. Skeletal joint smoothing white paper. [online, accessed 17 Dec. 2012] `http://msdn.microsoft.com/en-us/library/jj131429.aspx`.

[23] Microsoft (2012a). Skeletal tracking. [online, accessed 17 Dec. 2012] `http://msdn.microsoft.com/en-us/library/hh973074.aspx`.

[24] Microsoft (2012b). Starting february 1, 2012: Use the power of kinect for windows to change the world. [online, accessed 17 Dec. 2012] `http://blogs.msdn.com/b/kinectforwindows/archive/2012/01/09/kinect-for-windows-commercial-program-announced.aspx`.

[25] Obata, A. and Sasaki, K. (1998). Officewalker: a virtual visiting system based on proxemics. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, CSCW '98, pages 1–10, New York, NY, USA. ACM.

[26] Preis, J., Kessel, M., Werner, M., and Linnhoff-Popien, C. (201-). Gait recognition with kinect. [online, accessed 16 Dec. 2012] `http://www1bpt.bridgeport.edu/~jelee/courses/CS550_F12/Paper/7.GaitRecognition.pdf`.

[27] Razali, N. S. and Manaf, A. A. (2009?). Gait analysis for criminal identification based on motion capture. [online, accessed 12 Dec. 2012] `http://www.nauss.edu.sa/acit/PDFs/f3203.pdf`.

[28] Samsung (2013). Interactive smart tv with face recognition. [online, accessed 10 Jan. 2013] `http://www.samsung.com/us/2012-smart-tv/`.

[29] Turk, M. and Pentland, A. (1991). Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, pages 586 –591.

[30] Wang, L., Tan, T., Ning, H., and Hu, W. (2003). Silhouette analysis-based gait recognition for human identification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(12):1505 – 1518.

APPENDIX A

QUESTIONNAIRE

Spørgeskema

Hvad er dit fulde navn?

____

Hvor høj er du?

_____________________cm

Hvad er din alder?

_____________________år

Hvad er dit køn? (sæt ring)

Mand / Kvinde

Har du prøvet at bruge en Kinect før? (sæt ring)

Ja / Nej

Har du en Kinect i dit hjem? (sæt ring)

Ja / Nej

Hvis du har en Kinect, hvor ofte bruger du den så? (sæt ring)

Dagligt Ugentligt Månedligt Sjældnere

Tegn din TV-opsætning som den er i dit hjem, hvis der er flere, tegn den du bruger oftest.

- Skriv gerne (ca.) mål på tegningen.
- Tegn hvilken rute du typisk tager hen til sofaen / stolen
- Hvis TV'et står tæt på en væg, må du gerne tegne dette også

Hvis vi i fremtiden må kontakte dig vedrørende denne test, skriv da venligst et tlf. nr. eller email.

____

Samtykkeerklæring

Undervisningslokale 0.2.90 på Aalborg Universitet

Jeg, ____________________________, forstår den præsenterede information til dette system eksperiment. De spørgsmål, som jeg har stillet, er blevet besvaret tilfredsstillende.

Jeg samtykker at deltage i dette system eksperiment værende bevidst om, at jeg kan trække mig ud af eksperimentet på et vilkårligt tidspunkt.

Jeg samtykker, at eksperimentet bliver optaget på video under forudsætning, at dette materiale kun benyttes i forsknings- og undervisningssammenhænge.

Jeg samtykker, at det udfyldte spørgeskema må anvendes under forudsætning, at dette materiale kun benyttes i forsknings- og undervisningssammenhænge.

Navn på deltager: ____________________________

Underskrift: ____________________________

Dato: ____________________________

Navn på forsker: ____________________________

Underskrift: ____________________________

Dato: ____________________________

APPENDIX B
QUESTIONNAIRE RESULTS

In this appendix the results of the questionnaire will be presented. The answers to the questions can be seen in Table B.1 the drawings of the test subjects TV setups can be seen afterwards, note that user 5 does not use a TV and there is therefore no drawing.

User ID	Height	Age	Sex	Have used a Kinect before	Have Kinect at home	how often used
User 1	173	22	W	No	No	
User 2	170	25	M	No	No	
User 3	170	24	M	No	No	
User 4	158	24	W	No	No	
User 5	194	24	M	Yes	No	
User 6	164	22	W	No	No	
User 8	175	29	M	Yes	No	
User 9	196	25	M	Yes	No	
User 11	180	26	M	Yes	No	
User 12	170	26	M	No	No	
User 13	192	24	M	No	No	
User 14	185	25	M	No	No	
User 15	180	25	M	Yes	No	

Table B.1: The answers to the questionnaire.

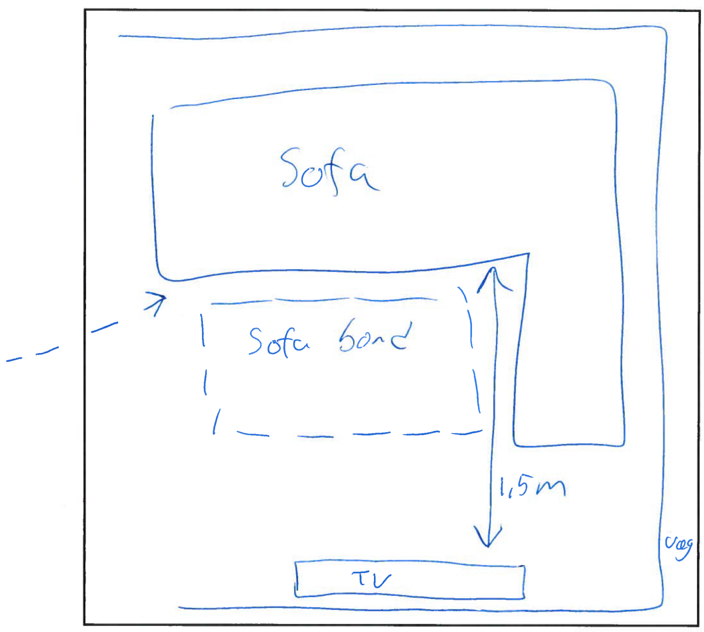


Figure B.1: Drawing from user 1.

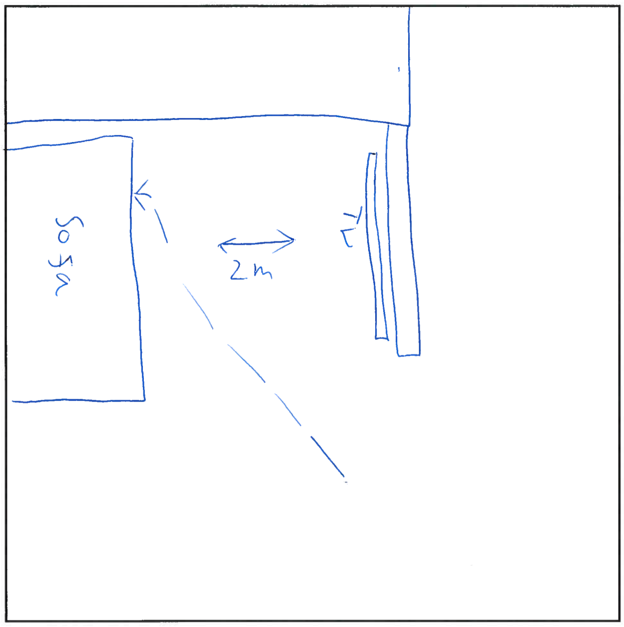


Figure B.2: Drawing from user 2.

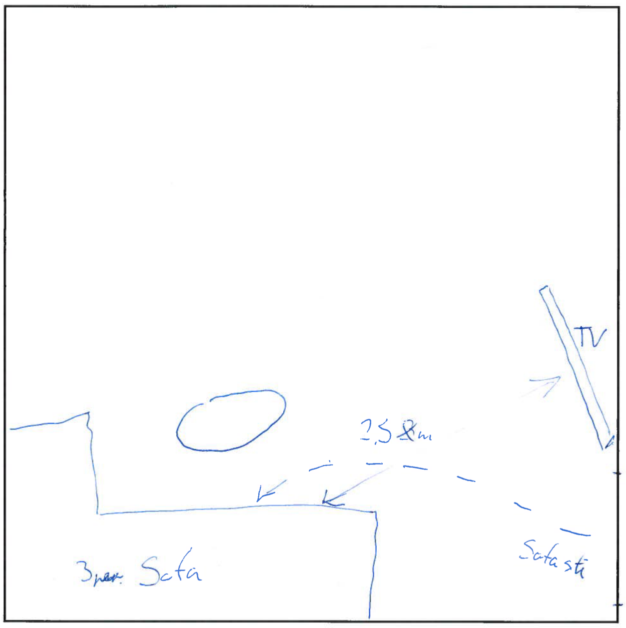


Figure B.3: Drawing from user 3.

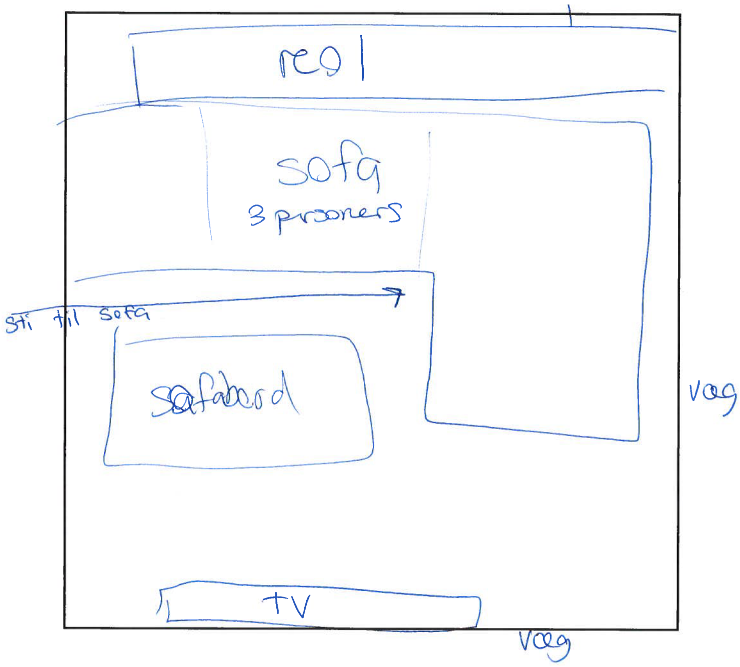


Figure B.4: Drawing from user 4.

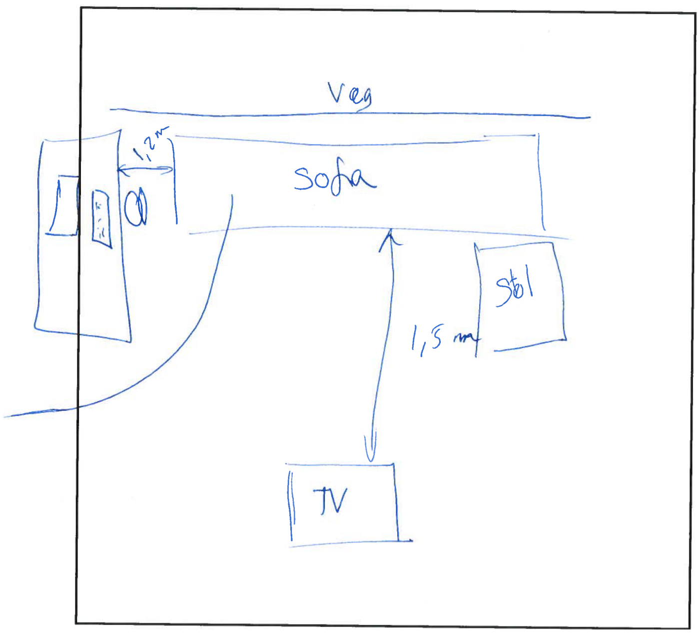


Figure B.5: Drawing from user 6.

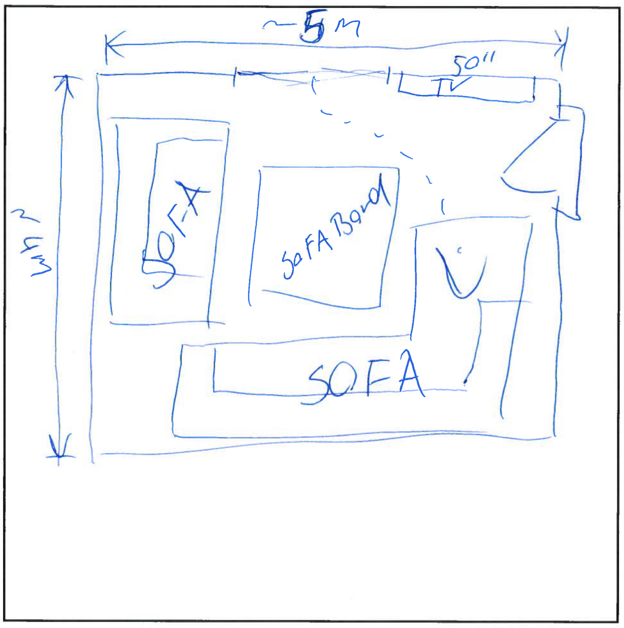


Figure B.6: Drawing from user 8.

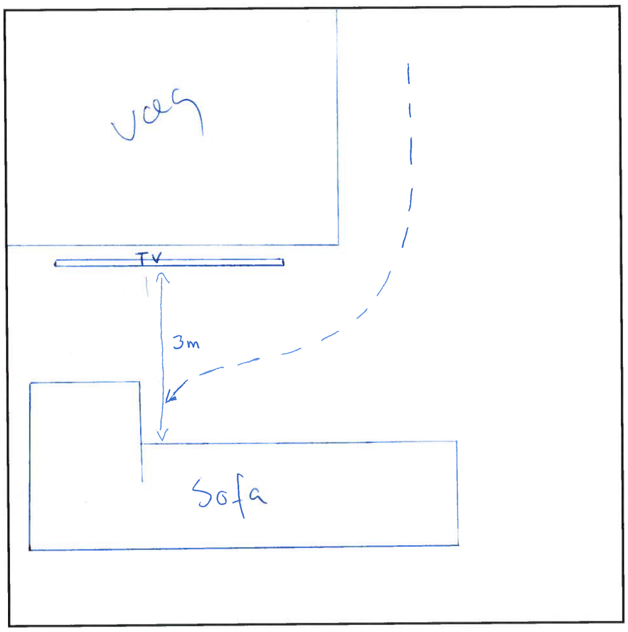


Figure B.7: Drawing from user 9.

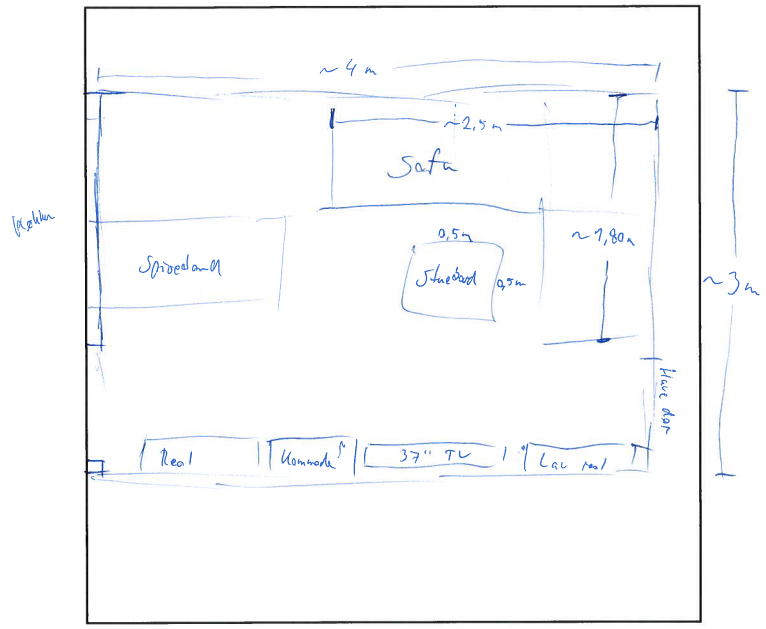


Figure B.8: Drawing from user 11.

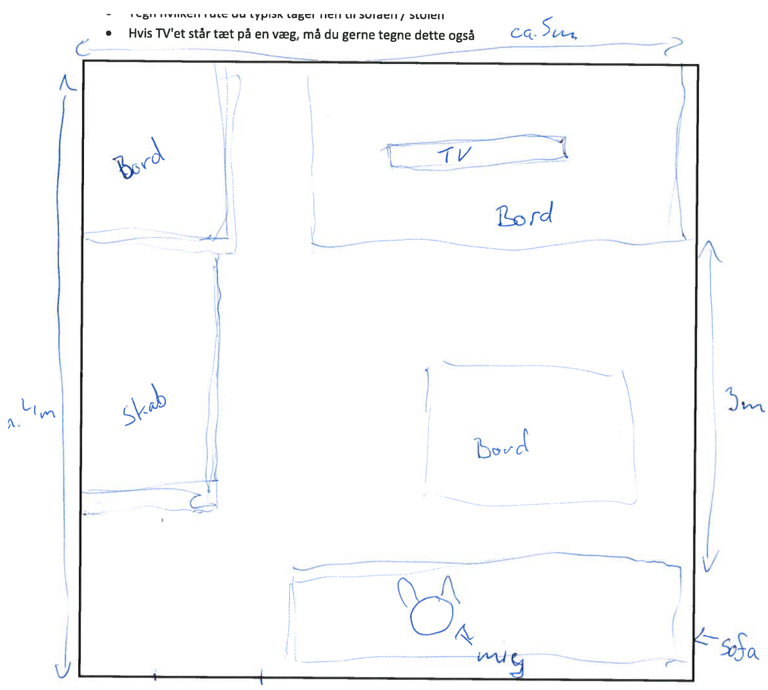


Figure B.9: Drawing from user 12.

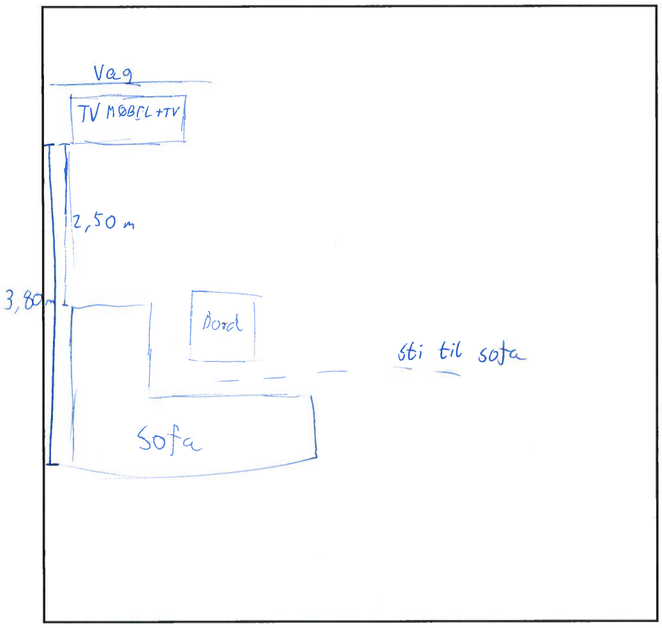


Figure B.10: Drawing from user 13.

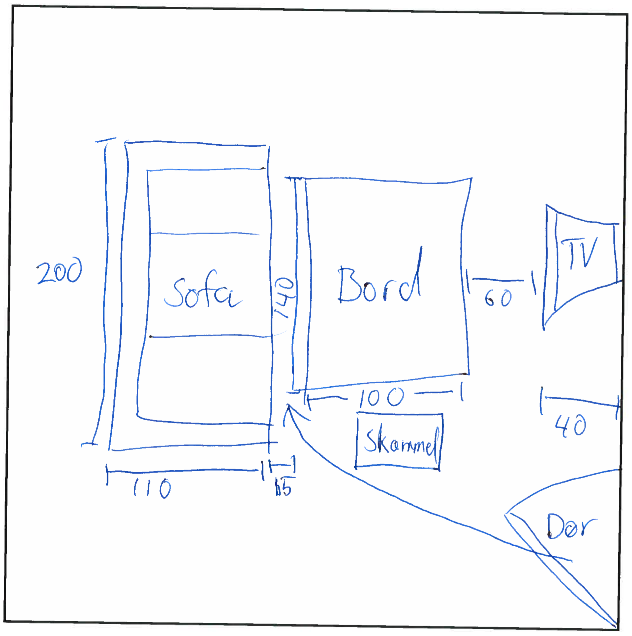


Figure B.11: Drawing from user 14.

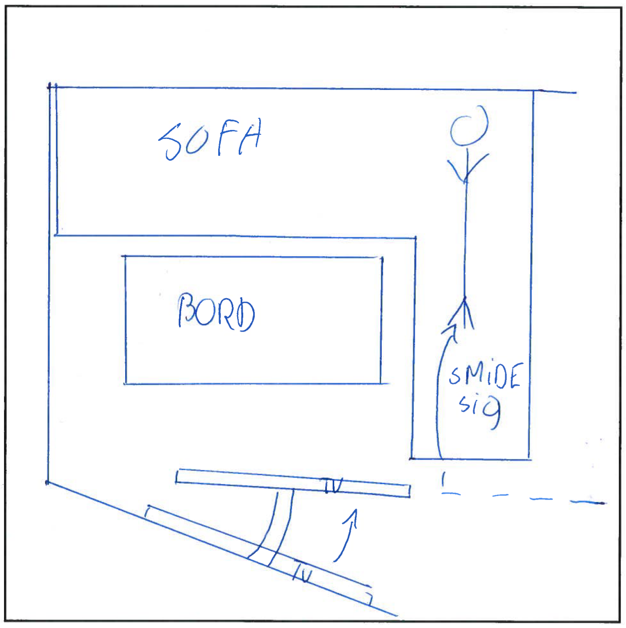


Figure B.12: Drawing from user 15.

APPENDIX C

EXPERIMENT EXECUTION

This chapter shows the papers, that were used under the execution of the experiment. Here the introduction to the test subjects can be seen. Furthermore the execution order of the tasks, can also be seen.

Manuscript

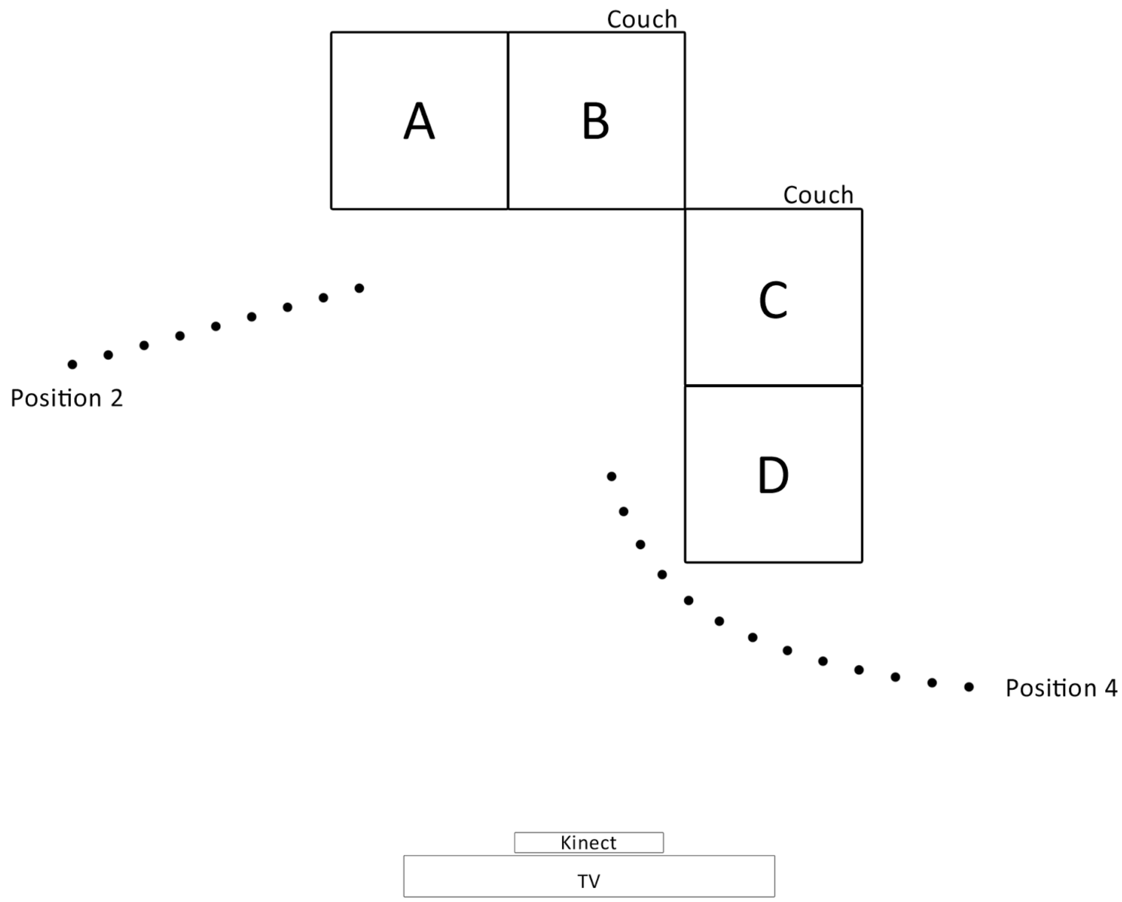
1. Velkommen
2. Dette eksperiment vil tage ca. 30 minutter
3. Der vil være et spørgeskema, som du skal udfylde efter eksperimentet
4. Som du kan se er der en masse streger og krydser på gulvet.
    - Et kryds er lig med en startposition
    - De stiplede linjer er vejen du skal følge.
        - Disse stiplede linjer er der for at indikerer den sti du skal tage hen til sofaen, men det er ikke noget der skal følges slavisk, det skal blot ses som en guideline
5. Vi fortæller hvilken
    - Startposition du skal starte fra
    - Farve streg du skal følge
    - Plads du skal sætte dig på
6. Ved nogle af opgaverne vil vi bede dig blive siddende, og foretage nogle simple bevægelser
7. Der er 4 forskellige setups, hvert setup har et antal opgaver tilknyttet
    - Der er 13 opgaver i alt, 12 af opgaverne skal udføres 5 gange

Calibration

The user must stand in a fitting distance away from the Kinect, so that the Kinect it is able to view his entire body, vertically as well as horizontally.
The user will then be asked to:

1. Slowly raise the arms to the side and subsequently raise them above the head, then take the arms down
2. Kick the feet so that the entire leg may move

Test 2 - Corner Couch

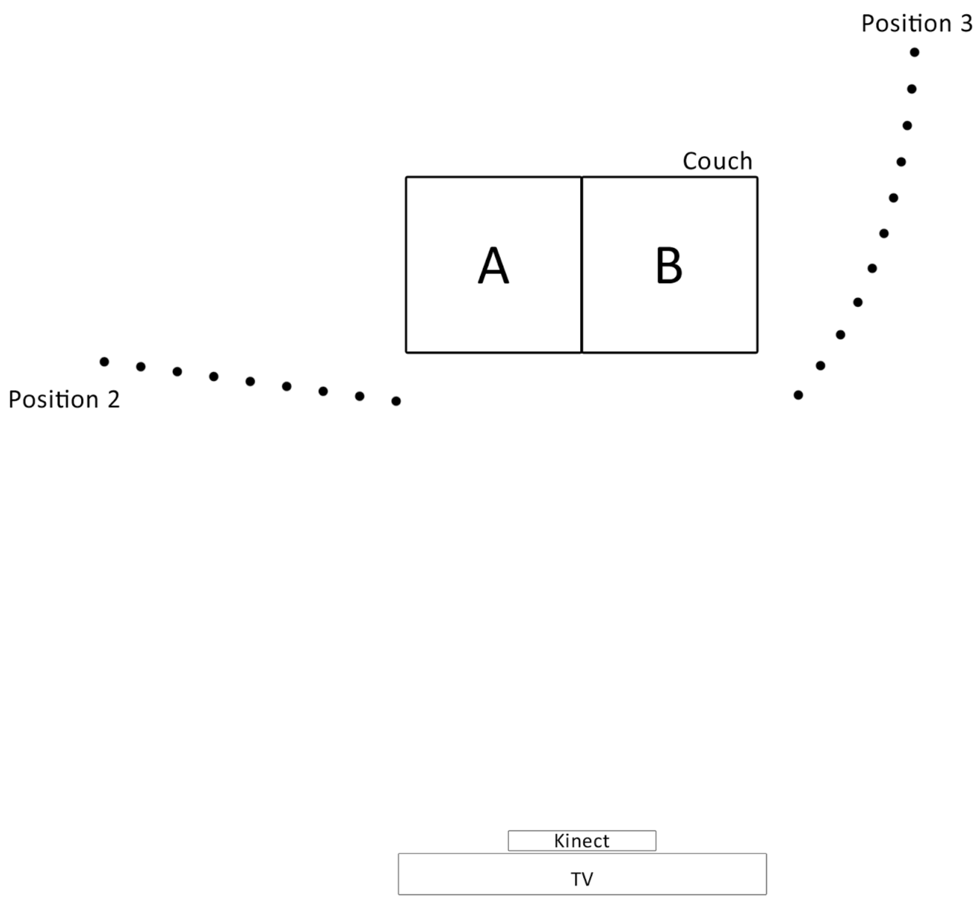


Times repeated	Walk task	Sit task
1, 2	2 -> B	B
1, 2	4 -> C	C
1, 2	4 -> B	
1, 2	2 -> C	
3, 4, 5	2 -> B	
3, 4, 5	4 -> B	
3, 4, 5	2 -> C	
3, 4, 5	4 -> C	

Sit tasks:

1. Lean slightly forward
2. Stretch your arms above your head
3. Simulate grabbing an object in front of you
4. Simulate grabbing an object to your side, such that you reach over the side of the sofa (If you seated right in the sofa, reach out to your left side)

Test 3 - Single Souch

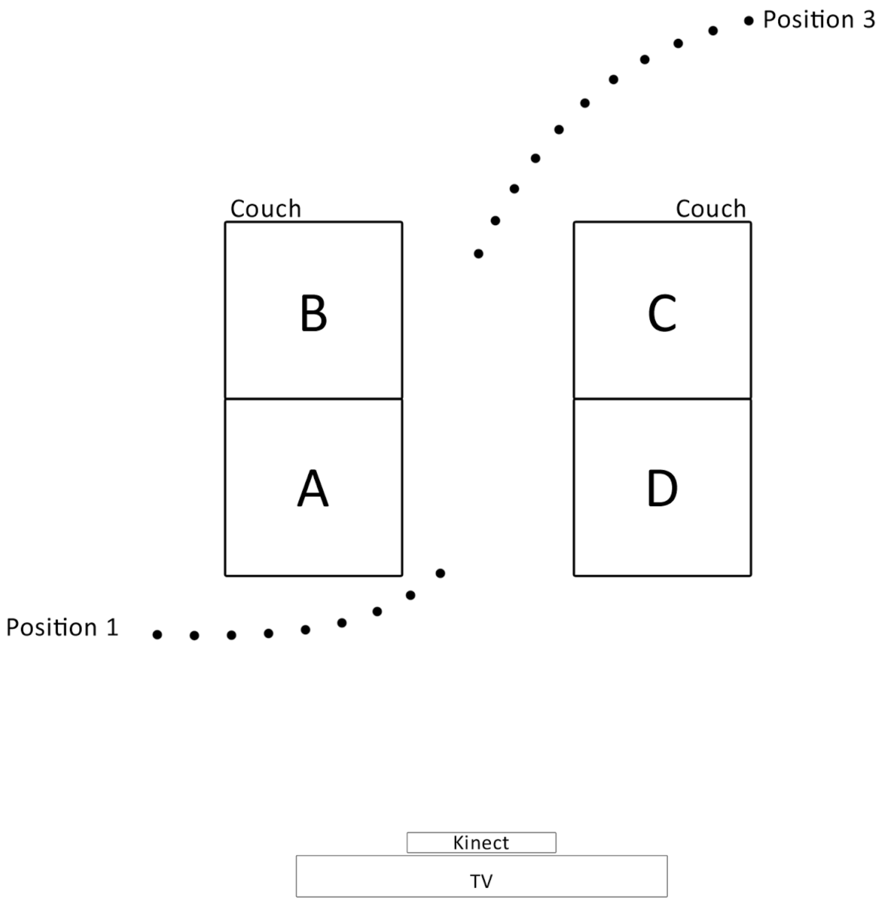


Times repeated	Walk task	Sit task
1, 2	2 -> A	A
1, 2	2 -> B	B
1, 2	3 -> B	
1, 2	3 -> A	
3, 4, 5	2 -> B	
3, 4, 5	2 -> A	
3, 4, 5	3 -> B	
3, 4, 5	3 -> A	

Sit tasks:

1. Lean slightly forward
2. Stretch your arms above your head
3. Simulate grabbing an object in front of you
4. Simulate grabbing an object to your side, such that you reach over the side of the sofa (If you seated right in the sofa, reach out to your left side)

Test 4 - Perpendicular to the Kinect



Times repeated	Walk task	Sit task
1, 2	3 -> D	D
1, 2	1 -> B	B
1, 2	3 -> B	
1, 2	1 -> D	
3, 4, 5	1 -> B	
3, 4, 5	3 -> D	
3, 4, 5	3 -> B	
3, 4, 5	1 -> D	

Sit tasks:

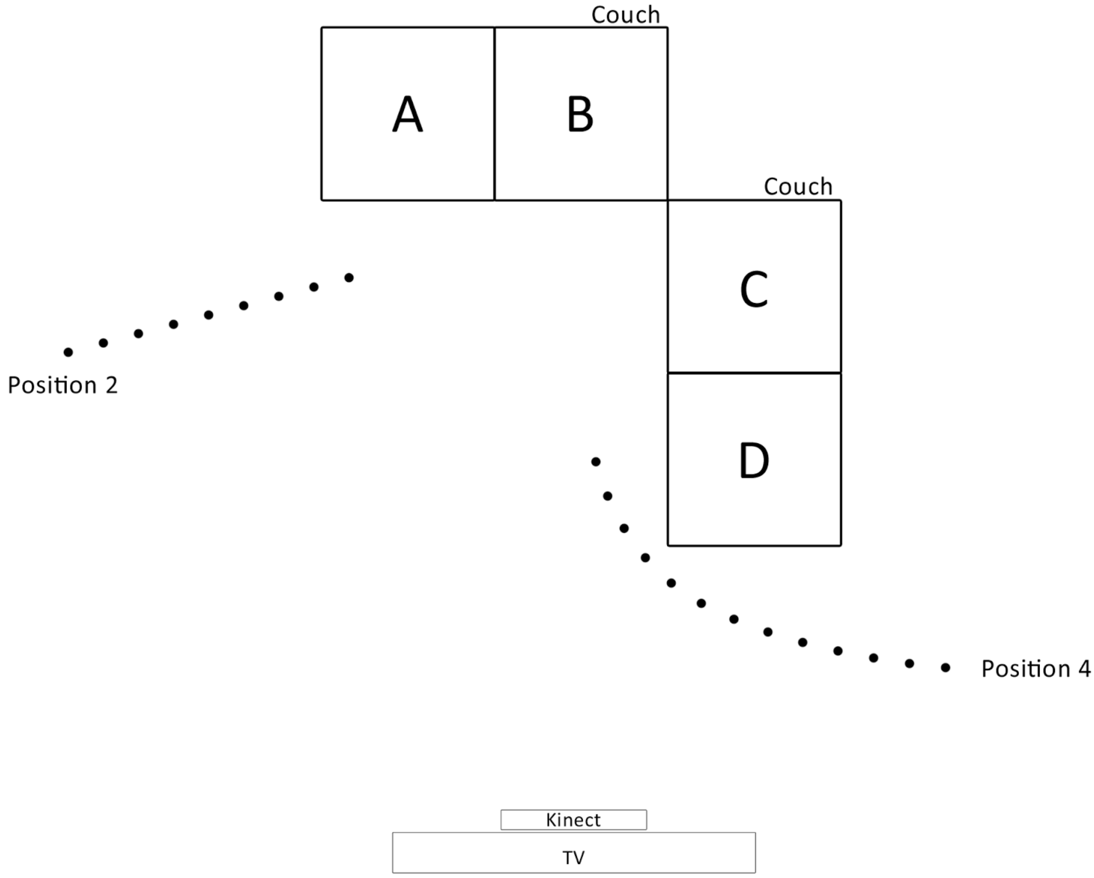
1. Lean slightly forward
2. Stretch your arms above your head
3. Simulate grabbing an object in front of you
4. Simulate grabbing an object to your side, such that you reach over the side of the sofa (If you seated right in the sofa, reach out to your left side)

APPENDIX D

PILOT EXECUTION

This chapter shows the papers, that were used under the execution of the pilot experiment. Here the introduction to the test subjects can be seen. Furthermore the execution order of the tasks, can also be seen.

Test 2 corner couch

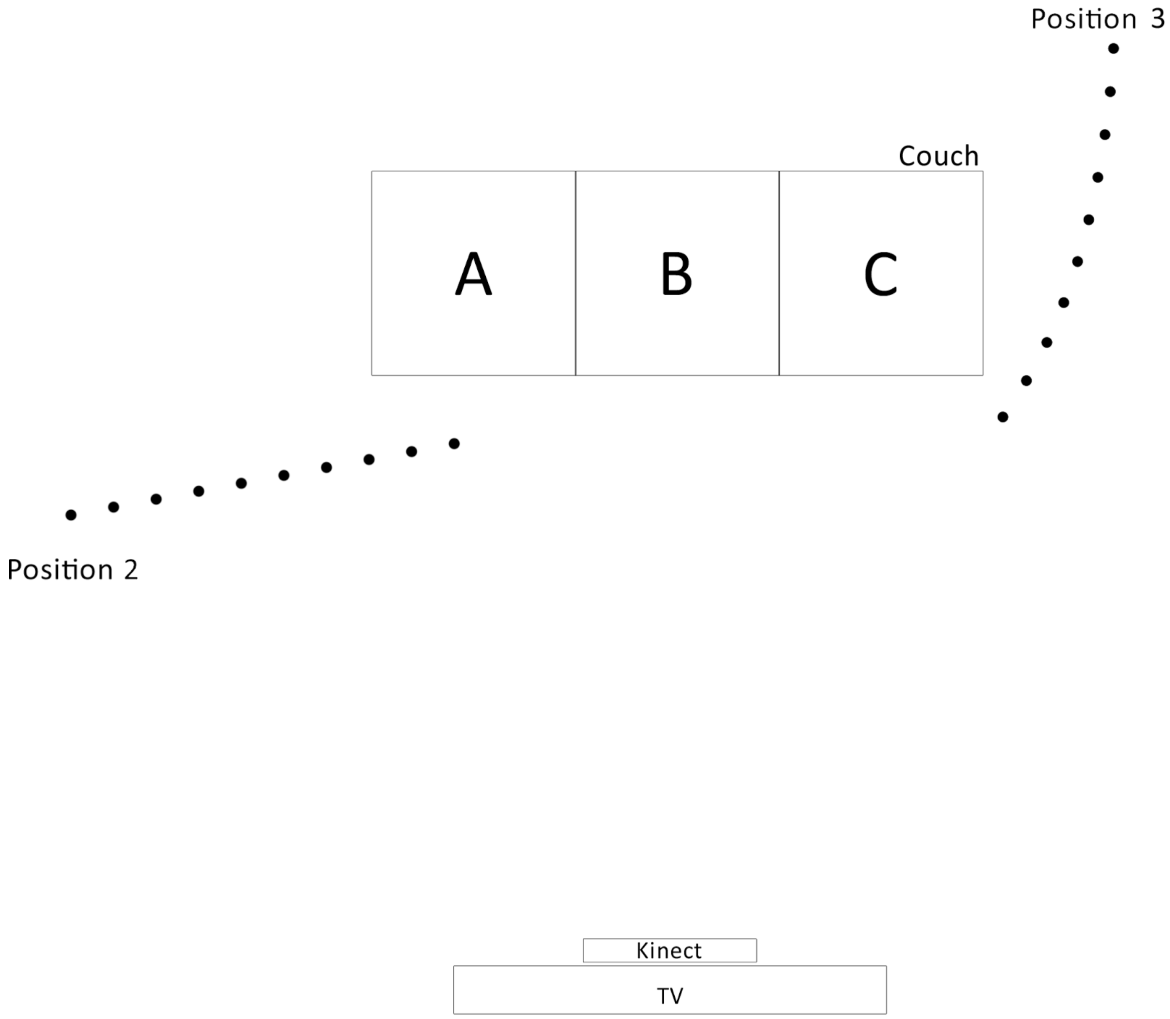


Times repeated	Walk task	Sit task
2x	2 -> A	A
2x	2 -> B	B
2x	4 -> D	D
2x	4 -> B	
2x	2 -> C	C
2x	4 -> A	
2x	2 -> D	
2x	4 -> C	
3x	2 -> B	
3x	2 -> A	
3x	4 -> B	
3x	4 -> D	
3x	4 -> A	
3x	2 -> D	
3x	2 -> C	
3x	4 -> C	

Sit tasks:

1. Lean slightly forward
2. Stretch your arms above your head
3. Simulate grabbing an object in front of you
4. Simulate grabbing an object to your side, such that you reach over the side of the sofa (If you seated right in the sofa, reach out to your left side)

Test 3 single couch

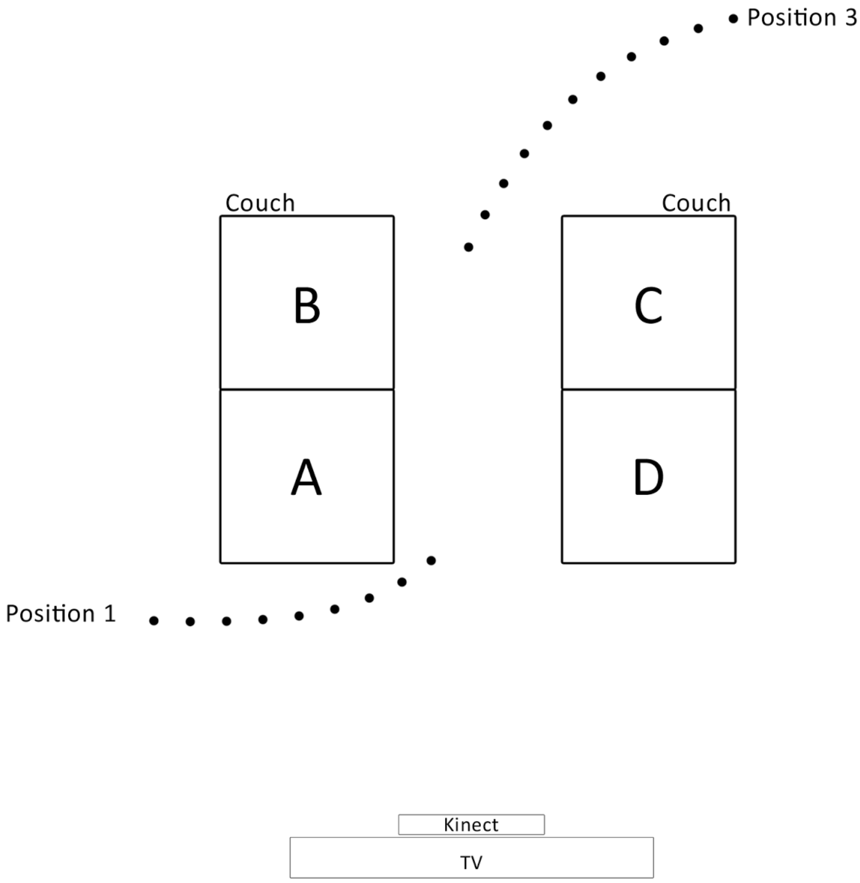


Times repeated	Walk task	Sit task
2x	2 -> A	A
2x	2 -> B	B
2x	3 -> B	
2x	2 -> C	C
2x	3 -> A	
2x	3 -> C	
3x	2 -> B	
3x	2 -> A	
3x	3 -> B	
3x	3 -> C	
3x	2 -> C	
3x	3 -> A	

Sit tasks:

1. Lean slightly forward
2. Stretch your arms above your head
3. Simulate grabbing an object in front of you
4. Simulate grabbing an object to your side, such that you reach over the side of the sofa (If you seated right in the sofa, reach out to your left side)

Test 4 perpendicular to kinect



Times repeated	Walk task	Sit task
2x	1 -> A	A
2x	3 -> D	D
2x	1 -> C	C
2x	1 -> B	B
2x	3 -> B	
2x	3 -> A	
2x	1 -> D	
2x	3 -> C	
3x	1 -> B	
3x	3 -> D	
3x	3 -> B	
3x	1 -> A	
3x	3 -> A	
3x	1 -> D	
3x	1 -> C	
3x	3 -> C	

Sit tasks:

1. Lean slightly forward
2. Stretch your arms above your head
3. Simulate grabbing an object in front of you
4. Simulate grabbing an object to your side, such that you reach over the side of the sofa (If you seated right in the sofa, reach out to your left side)

APPENDIX E
RAW DATA SAMPLE

This appendix shows a data sample of calculated skeletons.

Raw Skeleton Data (all values in cm)										
Torso	**Low-Arm L**	**Low-Arm R**	**Shin L**	**Shin R**	**Shoulders**	**Thigh L**	**Thigh R**	**Up-Arm L**	**Up-Arm R**	**Waist**
55,6									25,1	3,1
38,1	32,8									3,9
61,5		30,7	19,5	38,9	41,7	41,1	40,8		29,5	4,2
96,1					33,6					4,4
75,3										4,5
51,1	27,8									4,5
71,4		23,3							22,5	5,0
90,1	21,6									5,1
				34,9	51,4		32,1		23,6	5,2
				33,9			33,4		19,0	5,4
94,0	20,9									5,4
88,7	31,1	22,8	38,5		55,1	48,3	39,2	32,1	31,6	5,7
79,2	23,7	23,6	38,0	40,0	46,4	50,3	43,2	33,6	31,6	5,8
50,7									24,4	6,0
55,1										6,1
58,9		24,2								6,1
68,2	30,0	29,0							32,8	6,3
50,6					35,7				24,9	6,4
96,2					32,2			24,6		6,4
55,1						41,7				6,4

Figure E.1: Sample of raw data.