4-2016

Gesture based non-obstacle interaction on mobile computing devices for dirty working environment

William B. Huynh
Purdue University

Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By William Huynh

Entitled

GESTURE BASED NON-OBSTACLE INTERACTION ON MOBILE COMPUTING DEVICES FOR DIRTY WORKING ENVIRONMENT

For the degree of Master of Science

Is approved by the final examining committee:

Dr. Yingjie Victor Chen
Chair

Dr. Tim McGraw

Ray Hassan

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Dr. Yingjie Victor Chen

Approved by: Dr. Mihaela Vorvoreanu 4/27/2016

Head of the Departmental Graduate Program Date

GESTURE BASED NON-OBSTACLE INTERACTION ON MOBILE COMPUTING DEVICES FOR DIRTY WORKING ENVIRONMENT

A Thesis

Submitted to the Faculty

of

Purdue University

by

William B. Huynh

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

May 2016

Purdue University

West Lafayette, Indiana

This thesis is dedicated to my mom and dad for the support and caring they have given me while I was working on my masters. To my siblings who encouraged me to keep going and to strive to do better.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

ABSTRACT

Huynh, William B. M.S., Purdue University, May 2016. Gesture Based Non-Obstacle Interaction on Mobile Computing Devices for Dirty Working Environment. Major Professor: Yingjie Victor Chen.

The dominant way of interacting with tablets, smartphones, or wearable devices are through touchscreen or touchpad, which requires the user to physically touch the device's screen. However, in certain situation, for example, a dirty working environment, touching is not ideal or feasible. This study examined a new method that allows for a non-touch interaction by using the devices' back camera along with simple gesture to simulate mouse clicking. Cameras are used to capture motion-based gestures coupled with object detection, achieving a non-touch interaction. With human subject evaluation, the researcher found that using the back camera on mobile devices for gesture based controls is not as efficient as touch. Efficiency was measured in speed and accuracy, speed was about three times slower than current touch method, accuracy varied depending on smoothness of a task but was about two times less accurate. However, in a dirty working environment, this method is still effective and intuitive to use.

Keyword: Mobile human interaction, gesture, mobile devices, dirty working enviroment

CHAPTER 1. INTRODUCTION

1.1 Introduction

Mobile computing devices are an integral part of today's society. As time moves forward wearable devices are made more readily available to the general populace; Google Glass for example. To interact with these wearable devices touchpads or touchscreens are not as efficient as on a mobile device, making them obsolete due to either inefficiency, clunky size, and/or lack of available space. This thesis presents a form of interaction for mobile devices that is a non-touch interaction. Which can be applied to wearable devices as an alternative to touchpads or touchscreens, and allowing the usage mobile computing devices in a dirty working environment.

1.2 Scope

This research addresses on how current forms of interactions with mobile and wearable devices are inefficient. Touchpads and touchscreens are the most readily used for these devices, though they are not the only available options. This study looks into using gesture based controls to simulate that of a traditional computer mouse.

1.3 Significance

Gesture based interactions can resolve issues compounded by touchscreens for new generations of mobile devices. This study presents a method that uses gesture based controls, that is be based off of a tradition computer mouse. Allowing for new methods to display and organize information. By using gestures, it creates a non-touch interaction that can be applied to mobile and wearable devices, such as a Surface Tablet or Google Glass.

1.4 Statement of Purpose

This study provides a new method of interacting with mobile and wearable devices. By utilizing a camera to capture motion-based gestures coupled with object detection already available on mobile devices, a new mode of user interaction is introduced that will directly benefit users.

1.5 Research Question

How can we apply gesture based interaction onto mobile and wearable devices, to achieve a non-touch interaction, with no obstacles in the way, and simulate the traditional mouse behavior on the interface while in a dirty environment?

1.6 Assumptions

The following are the assumptions of this study:

- Participants had basic hand-eye coordination ability.

- Participants had normal vision to read the content on Surface Tablet.
- Tests was done in a well-lit and quiet environment.
- Participants had basic skills and intelligence to finish the given tasks within the time limit (Thirty minutes).
- Participants were honest in answering the questions on post survey.

1.7 Limitations

The following are the limitations of this study:

- Hardware specifications (4th Generation Intel Core i5, 4 GB RAM, 128GB Storage, 5 MP Camera and 2160 x 1440-pixel resolution) of the Surface Tablet.
- Hardware Layout of Camera

1.8 Delimitations

The following are the delimitations of this study:

- For many mobile or wearable computing devices, they all have camera and display with basic computing powers to execute our solution. In this study, the researcher chooses to use Surface Tablet to implement a non-touch interaction method.
- Choose to simulate the basic set of mouse operation (move, and click). Although there are many extended ways of using the mouse, these are the basic operations that a mouse should perform. These operations can be combined to create more complicated interactions (e.g. drag and drop)

- This study only focuses in a two dimensional space for interaction. The researchers are not going to bring in depth information into the interaction.
- Human subjects: limited to young adults that are more readily able to adapt to such new technologies as the Google Glass. Specifically, participants were students classified as College Undergraduates, or Graduate students at Purdue University.

1.9 Definitions of Key Terms

- Object Detection: Uses computer vision algorithms to identify objects in a view or video.
- Eye Wearable Computing: Small electronic device that individuals can move about and interact freely, supported by their personal domain (Mann, 1997).
- Google Glass: Is an eye wearable computing device made by Google, using widget based UI.
- Non-Touch Interaction: A new interface technology that allows users to manipulate objects by using gestures, accelerometers, and camera.

1.10 Summary

The goal of this thesis is to create a non-touch interaction for mobile computing devices for dirty working environments. Created on the Surface Tablet using OpenCV to capture motion-based gestures coupled with object detection. The testing is done using college students at Purdue University to determine if the non-touch interaction is efficient.

This solution can be generalized to other devices that require non-touch interaction. In order to create a valid non-touch interaction a literature review was conducted.

CHAPTER 2. LITERATURE REVIEW

2.1 Introduction

Object detection is a form of computer vision that can be used for interaction, which allows for gesture based control. Looking at the current technology and the direction it is heading a trend can be seen with every new iteration: that gesture based controls are becoming more prominent and is being refined to better suit users. In the literature review the researcher expands and looks into related works in object detection and motion capture to expand on the current knowledge base.

Motion capture uses object detection to detect movements in objects or people so they are similar in functionality. In the first section, the researcher goes over the current methods which are being used for object detection and motion capture to establish a base for the research. The researcher then moves on to talking about gestures and how they come into play with object detection. In the following section the researcher talks about eye-hand coordination and what part it plays with object detection. Then the researcher moves on to talking about computer vision.

2.2 Methods of Object Detection and Motion Capture in Computer Vision

Currently object detection and motion capture are used in movies and gaming consoles. In movies there are some uses of motion capture to achieve facial emotions in a scene. Such object detection and motion capture for facial emotions require high computational power, expensive equipment, and complex setup. Simplified versions of object detection and motion capture have been done and applied to mobile devices; an example of use is augmented reality. Azuma, et al. (2001) wrote about the recent advancement in the augmented reality field. Due to the advancement in tracking and computing technology, mobile augmented reality has become more feasible. Vuforia is a platform used for augmented reality, using this platform makes it easy to setup and use augmented reality on mobile devices. Gervautz and Schmalstieg (2012) show an image of a tablet running an augmented reality game using the Vuforia platform for enhanced learning.

As more research is being done, there are several methods which are possible to apply to this research. For the Surface Tablet, the approach taken is by using OpenCV object detection to determine what the user gesture commands are (Shrivastava, 2013). To support why the researcher chose to use the object detection method, the different methods available will be expanded upon. As technology becomes more advanced computational power is less of a problem, however it comes at the cost of a higher price. The approach the researcher will like to take is one that isn't costly and mobile.

2.2.1 Instrumental Tracking

Currently the most accurate way to capture motion is by using external equipment, which uses high end sensors attached to them. Examples of external equipment would be gloves with sensors attached to them and full body suits used for movies. However, these methods are expensive to use and take time to setup. Looking at the movie industry, there are users putting on full body suits that have sensors embedded in them for accurate motion capture. There are also methods in which one can isolate and only capture hand movements, but the gloves used to track them are still expensive and difficult to use (Shrivastava, 2013).

Instrumental tracking, although accurate, does not fit the criteria set out by the researcher. This method isn't viable due to portability issues because it requires external power, expensive equipment, and long setup time.

2.2.2 Marker Based Tracking

The marker based motion capture uses LED light that is attached to the fingertip of a glove which is harnessed by the camera to detect hand gestures. By using LED lights on the glove, this method is robust to different light settings (Park & Yoon, 2006).

The use of the LED lights makes this an invalid solution. This method does not fit the criteria of mobility; requiring the user to put on gloves which will need an external power source. Also depending on where the wearable devices' camera is located it might not be ideal to use LED lights (e.g. Google Glass located on the user's face). Though the gloves with LEDs is a cheaper solution than the previous methods above, it is still expensive none the less (Park & Yoon, 2006).

2.2.3 Color Marker Tracking

Wang and Popović (2009) applied a method which used color markers to track the user input. By using an inexpensive glove with colored patterns printed onto it, Wang and Popović were able to track the user motion. By using the colors on the glove they could use the nearest neighbor algorithm, which referenced the colors on the glove to approximate hand pose. The method they used only required a single camera; however, they found that it didn't achieve the accuracy which they had hoped for.

Color marker tracking used in analytics for sports, in this case baseball, was used for testing. The player would wear a glove with color markers on the joints of the finger. As the player swings their arm they were able to capture the motion and have an accurate result. They could take the motion data recorded and show the player how they swung from any angle to improve the player's pitching technique to help them visualize how they currently performed (Theobalt et al., 2004). Dorner (1994) uses the same method to capture sign language alphabet, however the tracking speed wasn't capable of real-time performance.

For this experiment a simpler version of this method was used. Instead of using a glove, stickers are placed on the user's fingertips. Wang and Popović used a glove since they were tracking more complex positions, in this experiment the researcher tracks only two fingers making it easier to track.

2.2.4 Bare-Hand Tracking

Bare-hand tracking is an ongoing research endeavor to simplify the technology and the process and still achieve the accuracy needed. The current process of achieving bare-handed tracking is to decipher edges and silhouettes, then decipher the pose or position of the hand. This method is highly computational and taxing because it requires complex algorithms to solve for the edges (Tzionas, et al., 2014).

2.2.4.1 Multiple Degrees of Freedom

Due to the complex algorithms, the speed at which the researcher can track motion slows down. To handle the speed issue, the researcher would need to create a library of image sequences which can be referred to when tracking. However, the amount of memory it would take to hold such a library is too large, therefore the researcher would need to generate the pose on the spot which would ultimately decrease the speed (Stenger et al., 2006). The reason why the library for bare-hand tracking is so large is due to the numerous degrees of freedom the hands have (Sudderth et. al., 2004). If the researcher decreases the number of degrees of freedom taken into account, they will not be able to account for all the different deformations our hand is capable of (Dewaele et al., 2006).

2.2.4.2 Limited Degrees of Freedom

There have been some successes with bare-hand tracking, but it comes at a cost of scope and degree of freedom. Schlattman and Klein (2007) devised a bare hand tracking system that had 10 degrees of freedom. Out of the 10 degrees of freedom, 6 of them were used for hand translation and rotation, and 4 were used for gestures. So by decreasing the degrees of freedom they were able to decrease the computational power needed.

Another solution is by Lee, Ghyme, Park, and Wohn (1998) who used a different method for motion detection to allow people to navigate through a virtual environment. They gave the users a set number of actions. The users were then tasked to move around in a 3d environment. By limiting the amount of commands to use, they are able to restrain the motions they needed to keep track of.

Dhawale, Masoodian, and Rogers (2006) took a 3d input and applied it to a 2d interaction. They applied the 3d input and 2d interaction to Microsoft solitaire and picture. To capture the 3d input they used a single camera-based video input. For movement in the Z direction they measure the size of the hand, and by using hand gestures they would invoke certain commands. For example, in solitaire when a fist was made it was considered as a click to drag the card to a certain location before unclicking when the hand goes flat. This represents another example where the users are given set commands to use and are limited in the degrees of freedom.

2.2.5 Object Detection

Object detection in computer vision is the processing and analyzation of an image. When processing and analyzing an image, edge detection is something that can be difficult. Viola and Jones (2001) proposed a method which would speed up this process by using simple features. The three main contributions they had are Integral Imaging, AdaBoost, and discarding background region in an image. Integral Imaging allows features used by detectors to be computed quickly. AdaBoost collects critical visual features from the image to create an efficient classifier. The third contribution uses classifiers to quickly discard the background so that they may compute the object region quicker. By doing this they decreased computation time while still achieving accuracy. Expanding upon Viola and Jones' study, Lienhart and Maydt (2002) introduced rotation of features by an efficient 45-degree angle into the learning framework. Also they optimized the classifier to improve performance, improving the overall performance by 23.8%.

2.2.6 Conclusion

The last few papers on object detection showed that the computation required can be decreased with decreasing degrees of freedom. Therefore, by combining this method with object detection which uses image processing, then decreasing the degrees of freedom, and limiting the tracking to a two dimensional space, it will decrease the computational power needed and complexity of the algorithm. When limiting the degrees of freedom for a user the researcher will only be looking at two finger gestures. Since the Surface Tablet will be focused on detecting only two fingers and tracking its location relevant to the tablet itself it will only be tracking in a two dimensional space. This will keep the design simple and easy to execute.

Taking into account some of the other methods mentioned that dealt with three dimensional space, it does not fit the scope of this project. The current scope of this project does not involve tracking in a three dimensional space due to time constraints. Though for future research it is possible to expand upon tracking in three dimensional space.

2.3 Gestures

There are many forms of gestures; one form of which came about when object detection was introduced. In the recent years the gaming industry has been looking into new forms of gameplay. These new forms required users to make certain gestures to play the game. Current consoles such as the Nintendo Wii U, requires the users to play by swinging the Wii remote. As the player swings the remote in a certain fashion, which is considered a gesture, that gesture is then mirrored on the screen for gameplay. Games like Wii Fit, Just Dance, and many other games recognize certain gestures for an interactive gameplay experience. The Wii remote uses infrared sensors to interact with the game.

2.3.1 Gesture for Console

The Wii controller is not limited to the use of Nintendo console. Schlömer, et al. (2008) used the Wii remote with bluetooth technology to conduct research on gesture recognition for an application they developed in Java. By using an open-source community they were able to extract the data sent from the Wii remote and read it into their application. The results showed they were able to recognize gestures by using a small sample size.

The Xbox One Kinect was also a key point for motion capture used for gaming. This device utilizes motion sensing input via range camera technology to track the user's full body movement and interpret their gestures (Tam & Li, 2012).

2.3.2 Gesture for Mobile

Mobile devices are an area in which motion capture is starting to be more prevalent, such as in augmented reality which is available on a majority of portable devices with a camera. Winkler, Hutflesz, Holzmann, and Rukzio (2012) attached a small projector to their mobile phones and were able to play bowling using the floors and walls. Another form of interaction could be gestures located on the back side of phones. By placing sensors on the backside of the phone users can use gestures to do certain commands (Wolf, McGee-Lennon, and Brewster 2012).

When creating the method of which users could do gestures on the backside of the phone Wolf, McGee-Lennon, and Brewster (2012) took into account the human anatomy. They also pointed out that there are no design guidelines for gestural control. At the current rate of growth for technology, gesture based interactions are becoming more prominent in society. Due to the fact that gesture interactions are growing, gesture based controls need to be more ergonomically effective.

Some of the common gestures that have been implemented for mobile devices include pinching with two fingers for zooming in/out. As well as pinching with four or five fingers to enter the multitask screen. Another example is swiping from the bottom or top of the screen to bring up certain menu's or task bars. These are only some of the types of gesture that are integrated in mobile devices that we have already seen

2.3.3 Conclusion

Technology is growing in a way where current interactions, will need to be expanded upon to fit the needs of wearable devices. Motion capture is a growing field, leading to gesture based interaction with new consoles or devices. The Google Glass for example is a new form of information display for the user; however, it doesn't give the user a new form of interaction. By applying gesture based interaction to the Google Glass it allows for more varieties of information displays, user interface systems, interactive applications, and general interactions to be available at their fingertips. Take augmented reality for example where the user no longer has to hold a phone to an image with the encoded game, but instead just look at the image. Other examples include opening the user's hand for an interactive gameplay and make for a new form of user interface.

2.4 Eye-Hand Coordination

When defining the eye-hand (or hand-eye) coordination, it is defining the coordinated control from the eye to our hand. In other words, it is the process in which our eye communicates to our hand guiding it in a 3 dimensional space to grab an object. Fitts's Law is a model of human movement used for human-to-computer interaction. An example of Fitts's Law is when using a computer, a user is using a mouse which is assigned to a pointer on the screen, which can then be used for on screen interactions. Lazzari, Mottet, and Vercher (2009) concluded that "eye-hand movements in a rhythmical Fitts's task are dynamically synchronized to produce the best behavioral performance." Meaning, eye-hand coordination is important to achieve the best result.

Johansson, Westling, Bäckström, and Flanagan (2001) said, "The gaze supports hand movement by planning and marking key positions to which the fingertips or grasped object(s) are subsequently directed." In other words, being able to see the target, the eyes will guide the hands to a key position of the object. Without the use of the eyes it will take more tries due to decreased accuracy to the target as the hands lack directions given by the eyes.

The articles above are saying the eyes are providing spatial information to the hands to move in the manner necessary to achieve a goal. When using motion capture, the researcher expects the physical movements to be relative to the movements on the screen in real-time. By applying motion capture to mobile or wearable devices, movements made by the users will be displayed relative to the screen to achieve the best experience possible.

2.5 Feature Detection in Computer Vision

In computer vision a concept of feature detection refers to computing abstractions of image information. Selecting good features properly to track in the physical world is difficult. Feature selection criterion is optimal for it is based on how trackers work and feature monitoring method detect occlusion. To make feature tracking more accurate using the features' dissimilarity which includes linear warping and translation to track the object's movements from frame-to-frame (Shi, & Tomasi, 1994).

There is a great diversity in applications using edge detection, but they share a common set of requirements. These requirements yield an abstract edge detection problem, the solution of which can be applied in any of the original problem. A common criterion in evaluating edge detector performance are the following: low error rate and edge points need to be localized well (Canny, 1986).

Yuen, Princen, Illingworth, and Kittler (1990) investigated a number of circle detection methods which are based on variations of the Hough Transform Method, which include the standard HT and Fast HT, and compared the performance of the two methods as well as illustrated the properties such as accuracy, reliability, computational efficiency and storage requirements. They found complex images caused difficulties and that GHTG method performed the best due to robustness.

2.6 Conclusion

The key aspects that are mentioned are the different varieties of methods to record motion capture. In conclusion with the research materials that were obtained detailing the various methods for motion capture, the previous stance of using the color markers method is suitable for our purposes. It is the least computationally taxing of the methods as well as being cost efficient and mobile. This method also enables eye-hand coordination in which the researcher shows the significance of relational response. By using Fitts' Law to show that the correlated response from the hand to what is seen proved to produce the best performance.

Motion capture can then be used to detect certain gestures for new forms of interaction. For example, if the user were to face their hands to the side and swipe to the right or left, the display will change between applications. For mobile devices this is a form of interaction applied to new technology, making it a novel invention. Scope wise our method will allow the users to move their hands left, right, up, and down in a two dimensional space while controlling the Surface Tablet. This allows users to have more information available at their fingertips.

Wearable devices are new forms of technology that is trying to break into the market. It is an innovative way that allows for new interactions to look at information. Current methods require users to use a touch pad attached to the wearable device, but for this research paper the researcher looks into motion capture. By using motion capture users can use gestures which will correlate with commands, for non-touch interaction. This research will be a basis which will be used to further studies done on mobile and wearable devices interactive interface.

CHAPTER 3. METHODOLOGY

In order to measure the effectiveness and efficiency of the new form of interaction, the time and failed attempts were recorded for the completion of the set task given was evaluated. Participants are given three tasks repeated twice in which they are given at most five minutes per task. The independent variable of the study is the task given to the participants to complete. The post-test is given to determine if the participant had a good experience with the new form of interaction. The dependent variable of the study is the time taken to complete each of the set tasks.

The hypotheses for the study follow:

H0: The user feels that the non-touch interaction is acceptable because either of the following: a) it is intuitive and natural, b) the interaction can finish the task given, c) easy to learn, d) if gesture based control was as efficient as touch in terms of speed and accuracy.

Ha: The user feels that the non-touch interaction is not acceptable because either of the following: a) it is non-intuitive or natural, b) the interaction cannot finish the task given, c) not easy to learn, d) if gesture based control was not as efficient than touch in terms of speed and accuracy.

3.1 Recruiting

The recruiting and testing process happened on Purdue University Campus during the Spring of 2016. A mass email was sent to the anyone in the College of Technology, with cash incentive offered for participation. The researcher was looking for at least 30 participants, ranging from the age of 18-30 years old. Participants are not expected to have any prior knowledge on the subject matter or experience with the Surface Tablet.

3.2 Measurement

To measure efficiency of gestures compared to that of touch, the speed and accuracy was recorded for each task. Speed was recorded by the time (seconds) spent to accomplish a given task. Accuracy was measured in the amount of attempts taken to accomplish the set task. The post-survey used Likert scale to measure the participants experience of the gesture based control (e.g. ease of use). There are seven questions that use a Likert scale, first two questions involve a time frame so they were give a range of time (0 -1 hr., 2-3 hr., 4-5 hr., or 6 plus hours). The rest of the questions measure experience on gesture based controls (strongly disagree, disagree, neither, agree, and strongly agree).

3.3 Analyze the Data Collected

This study uses a combination of paired t-test and ANOVA test to analyze the data collected. Paired t-test is used to compare two population of means, one being the time taken using touch, and second being time take for gestures. Paired t-test gives a p-value in which was used to conclude whether or not the gesture based control are as efficient as touch. One-way ANOVA is used on the data collected from the post-survey, to compare more than two group at once. ANOVA gives an f-value instead which is used to conclude if the variances are equal.

CHAPTER 4. DEVELOPMENT AND EVALUATION

4.1 Design

The researcher looked at how the use of object detection computer vision algorithms can be used for the Surface Tablet to achieve a non-touch interaction. After which the researcher uses an object detection algorithm to detect certain gestures. When looking for gestures that the users could use for the non-touch, the researcher had a few iterations. Originally users were going to be able to use their whole hand to interface with the Surface Tablet, however the amount of degrees of freedom used was too high in computation. The next iteration was used three fingers, however that felt too unnatural to hold three fingers up be it thumb or ring finger as the third. This narrowed it down to two fingers which had a natural feel and was low in computation. From there it was decided to use the front of the hand instead of the back of the hand for higher ease of use (Figure 4.1).

Figure 4.1 Back Hand and Front Hand

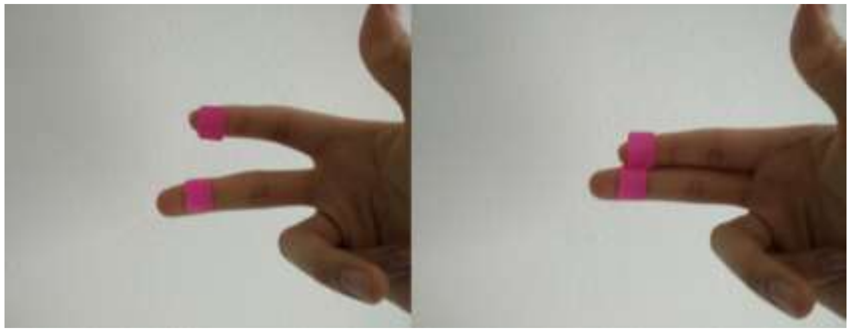

Figure 4.2 Gesture Controls

Using only two fingers, the users are given two gestures in which they can perform as mentioned above. One gesture being a scissor shape in which the user's cursor was located in between the two fingers. The next gesture was placing both those fingers together to simulate a click, which can then be used to also drag/drop (Figure 4.2). A prototype application for these gestures was developed for the Surface Tablet, so that the researcher may evaluate the new form of interaction, developed in C++ and OpenCV.

4.2 Development and Evaluations

4.2.1 Development of the Interaction

The development of the tasks and gesture based controls for this research is done on Visual Studios 2013 in C++ with OpenCV's library. OpenCV library takes the back camera feed on the tablet, and converts the image from a BGR (Blue, Green, and Red) to HSV (Hue, Saturation, and Value) value. Predefined HSV values are then applied such that it tracks the pink color marker, which needs to be changed depending on lighting. Once the color markers can be found, the image is converted to gray scale where canny edge can be applied. From the gray scale image contours can be found, if more than the max numbers of contours are found the HSV values may need to be changed. For the gestures based controls two contours need to be found to move without clicking, and one contour being clicked (Figure 3.2). From there C++ is used to find the middle of the contours found and OpenCV to draw the cursor. C++ is also used to draw the window and track time it takes for a given task.

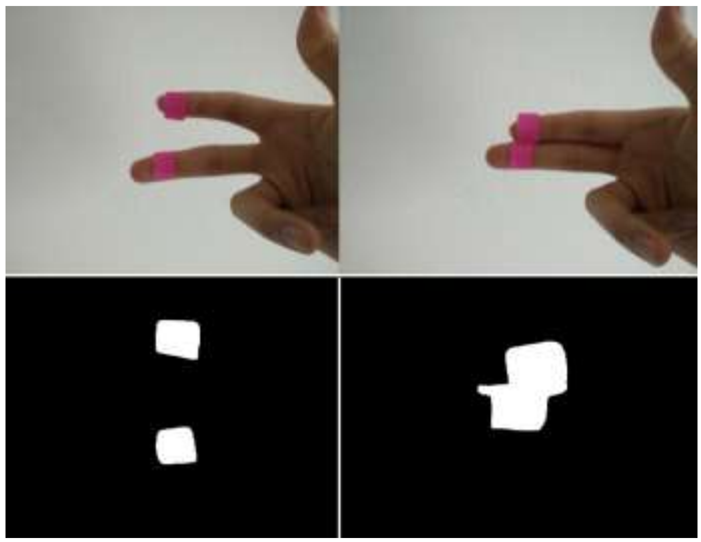

Figure 4.2 Gesture after grayscale

4.2.2 Types of Task

The research conducted is a quantitative study, there are three types of task the researcher will be measuring, each task was repeated twice. The three types of tasks are the following a calibration task (click), and two of which are picking something up and moving it along two kinds of paths (drag and drop), a jagged path and smooth path (Figure 4.2.2). For the click task the user selects a marker on the screen that was highlighted at random and change after it has been clicked. To complete this task, the user needed to click twenty of the markers, the time between each click was recorded except the first. The first click was not recorded since there was not a prior click to record from. For the move task the user follows a path in which they can fall off so the user must be careful not to fall off. The timer does not reset for each time the participants falls off.

Each participant is given three tasks on the Surface Tablet. The task each participant starts with is the calibration, jagged path, and then smooth path. The time taken to complete each task will be recorded. Each task will take no more than 3 - 5 minutes to complete. After completing each task once using touch they will then move to using gesture based controls for the tasks.

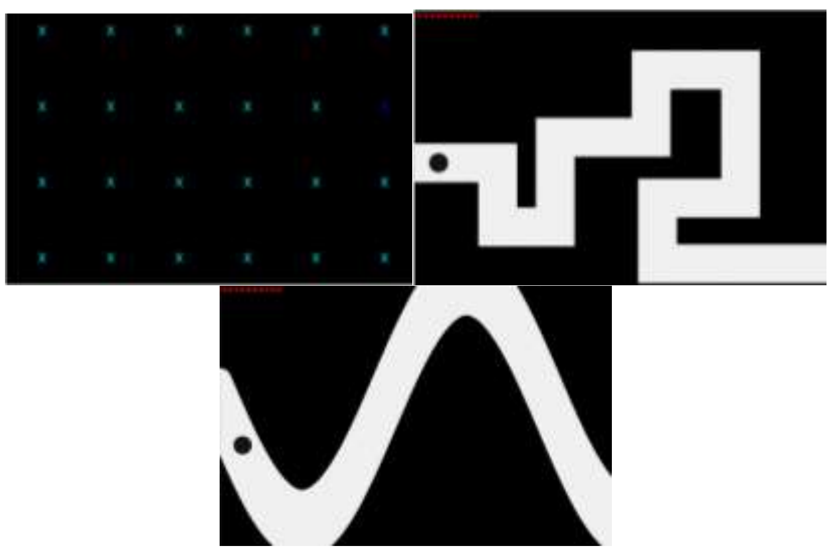

Figure 4.2.2 Task 1 top left, Task 2 top right, Task 3 bottom center

4.2.3 Recruiting the Participants

Thirty-one students were recruited to participate in the test. However, one of the data points collected was not done correctly and was removed from the test results.

Among the thirty students, nineteen were male and eleven females from an age range of eighteen to thirty. Participants included both undergraduates and graduate students a majority coming from the College of Technology.

4.2.4 Testing Process

Before the participants start the test they are given a consent form in which the researcher will go over with the participants. Such that they know what to expect during the test and what is expected of them. After they have signed the consent form they may then proceed on with the test.

Participant will perform each of the given task twice, and the scores were recorded each round. They will perform each task once as touch before performing the task a second time as gestures. The reason for why the participants are to complete each task twice is because the first time through they will be using touch to set a time base for comparison to the gesture based controls. Each participant is given a twenty to thirty-minute time frame total to complete each task twice. The researcher can spectate but cannot interfere with the mobile device to keep the user moving along. This will also allow the researcher to record the time taken to complete each task.

After the participant has completed each of the tasks they are given a post-test. The questions use a Likert scale. The questions will ask about their experience with the Surface Tablet, how they rate the new form of interaction felt (moving, clicking, and drag and drop), and allow them to rate if they felt tired having their hands up in front of them for the interaction. The post-test takes about five minutes to complete and participants will be given a comment area in which they can write their thoughts, however that will not be used in the evaluation of this thesis.

4.2.5 Testing Environment

The testing was done on Purdue University campus, on a voluntary basis using convenience sampling. The sample size will be about 30 participants around the age of 18-30 years of age. Testing will be done in a quiet office, and closed door in Knoy Hall of Technology. The room used for testing was lit like a normal room, and such that the camera can record/recognize the fingers. The researcher will use a two-way ANOVA test to analyze the time recorded and post-test to determine a statistical significance.

4.3 Summary

This study will test whether or not the new form of interaction on mobile or wearable devices is an efficient method. The study is conducted at Purdue University with a convenience sampling of 30 participants. Each participant will have completed three tasks twice in the time allotted to them. The time recorded to complete each of the task is then used to analyze for efficiency.

CHAPTER 5. DATA ANALYSIS

This chapter describes the data collected in the study. The data was analyzed using a one-way ANOVA and paired T-tests. Data collection and general results of the raw data will be discussed, followed by the results from the analyzed data from the questionnaires on ease of use and efficiency.

5.1 General Results

According to the survey the age range in which most participants fell into were 20-25 with no significant difference between ages (Figure 5.1). Participants spent an average time of 6 or more hours on the computer, and with a median of 2-3 hours on a tablet or mobile device daily. There was about a 19:11 male to female ratio that participated in the test.

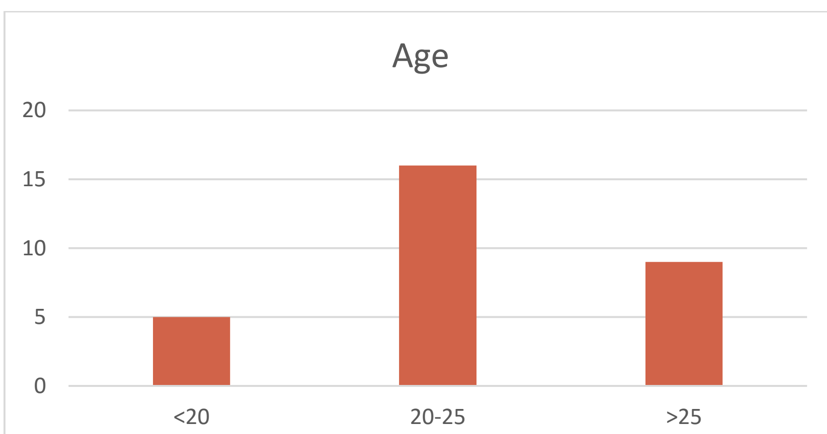

Figure 5.1 Age ANOVA Test

5.2 Time Spent on Computers or Tablets

As mentioned above the average time spent on the computer were 6 or more hours a day, though does more time spent equal better results. Using ANOVA test the data showed that time spent on the computer did not make a difference (Figure 5.2.1). Doing the same test for tablets we received the same results that there was no significant difference between those who spend more time on tablets than those who spent less time (Figure 5.2.2).

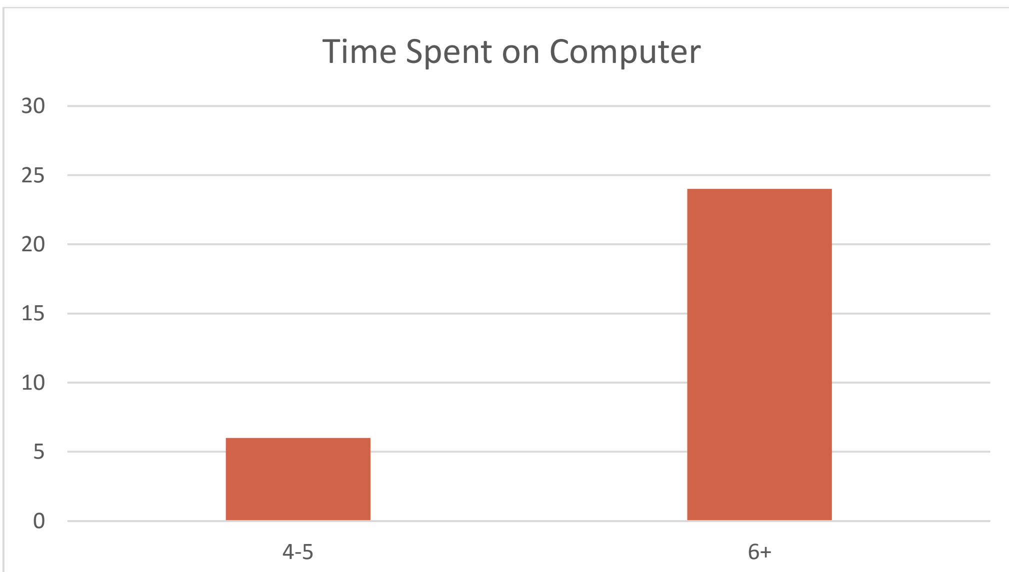

Figure 5.2.1 ANOVA Test for Time Spent on Computers

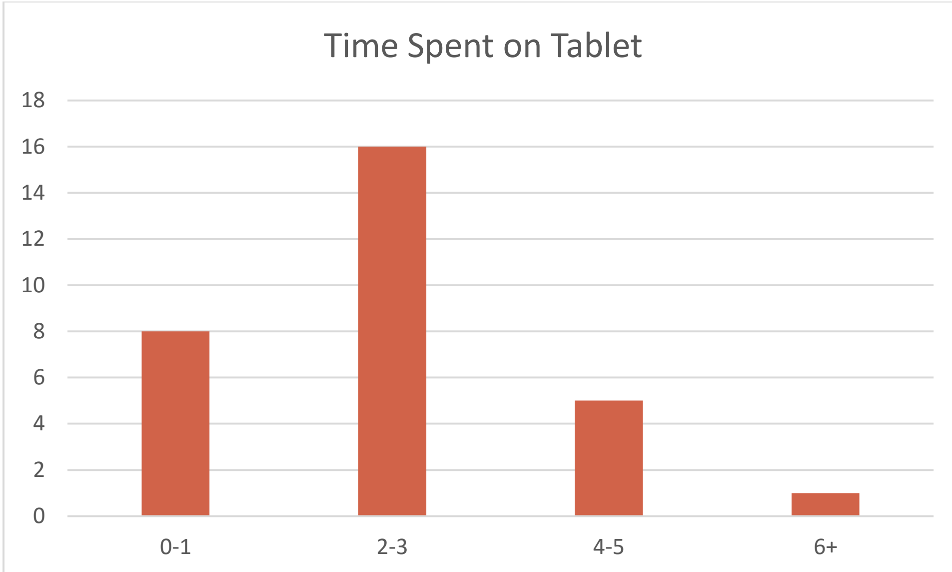

Figure 5.2.2 ANOVA Test for Time Spent on Tablets

5.3 Intuitive and Natural

To evaluate whether or not the gestures were intuitive and natural in the post survey we had two questions one asking was the gesture intuitive, and the other asking does the gesture feel natural and comfortable. The data showed that the average response to the if the gesture was intuitive or not was a 3.733 out of 5 on a Likert scale, with a high frequency in 4's therefore agreeing that the gestures were intuitive (Figure 5.3.1). Though it was intuitive the question on if the gestures felt natural and comfortable showed an average of 2.433 out of 5, and a high frequency of 2's. Since there was a high frequency of 2's it shows that the gestures though intuitive were not comfortable to use (Figure 5.3.2).

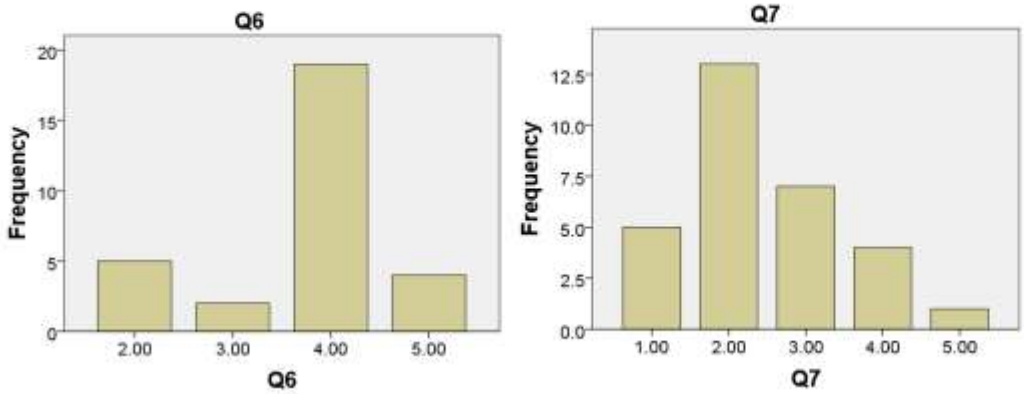

Figure 5.3.1 Bar chart from response to questionnaires, left bar chart is the response on whether the response was intuitive and right is on if the gestures were natural and comfortable

Which brings up if the gestures were easy to use, the response as shown in figure 5.3.3 was that the gestures were neither easy or difficult to use with a mean of 2.667 out of 5. Another question asked was if the gestures were available would they use it. A mean score of 2.7337 so possible use. This brought up the question of if this was on a smaller device would it be better. The average response given was 2.733 out of 5 which meant the participants believed it would not necessarily be better (Figure 5.3.3).

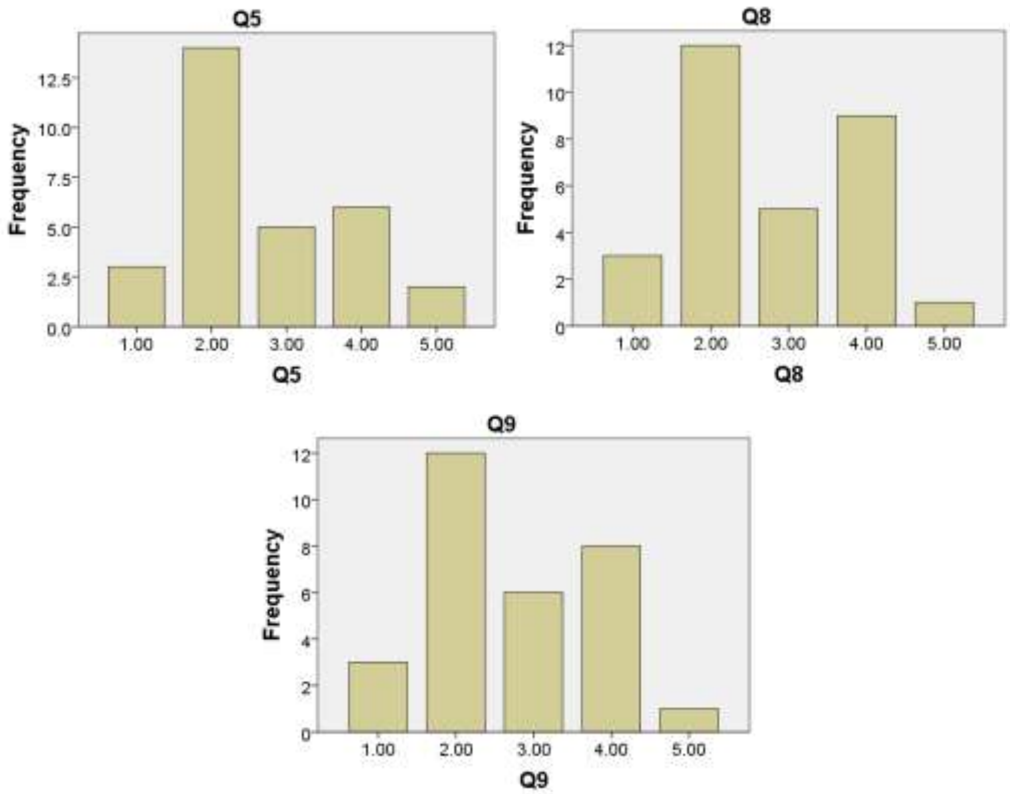

Figure 5.3.3 Top left bar chart shows ease of use, top right is on if available would they use it, and bottom center bar chart is whether it was better on a smaller device

5.4 Time Taken on Touch and Gesture

Looking at the time data collected for the three task, the researcher found a few outliers by calculating the lower and upper limits (First or third quartile ± IQR*1.5). The researcher removed the outliers and found that touch was significantly faster and more efficient. Task one's mean touch time from one point to the next was 1.0845 seconds as for the mean for gestures was 2.9342 seconds. A two sample t-test shows that touch and gestures on task one has a t-value of -30.72 therefore rejecting the null hypothesis that gestures are as efficient as touch. The second task given had similar results with a time difference of 19.4855 seconds touch with a 95.2615 seconds gestures, and a t-value of -

9.632 (Figure 5.4.1). With gestures having a greater chance of going outside of the path and restarting than touch (Figure 5.4.2). The third task was no different than the second task the mean times being 15.1596 seconds for touch, a 106.6480 seconds for gestures, and a t-value of -6.120 (Figure 5.4.1).

Paired Samples Statistics

		Mean (sec)	N	Std. Deviation	Std. Error Mean
Pair 1	T1	1.0845	523	.22752	.00995
	G1	2.9342	523	1.36459	.05967
Pair 2	T2	19.4855	28	8.76283	1.65602
	G2	95.2615	28	42.01369	7.93984
Pair 3	T3	15.1596	30	4.96444	.90638
	G3	106.6480	30	81.79684	14.93399

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	T1 & G1	523	.029	.515
Pair 2	T2 & G2	28	.148	.452
Pair 3	T3 & G3	30	.015	.938

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean (sec)	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper			
Pair 1	T1 - G1	-1.84975	1.37701	.06021	-1.96804	-1.73146	-30.720	522	.000
Pair 2	T2 - G2	-75.77598	41.62670	7.86671	-91.91713	-59.63483	-9.632	27	.000
Pair 3	T3 - G3	-91.48846	81.87360	14.94801	-122.06057	-60.91636	-6.120	29	.000

Figure 5.4.1 T-Test between touch and gesture controls

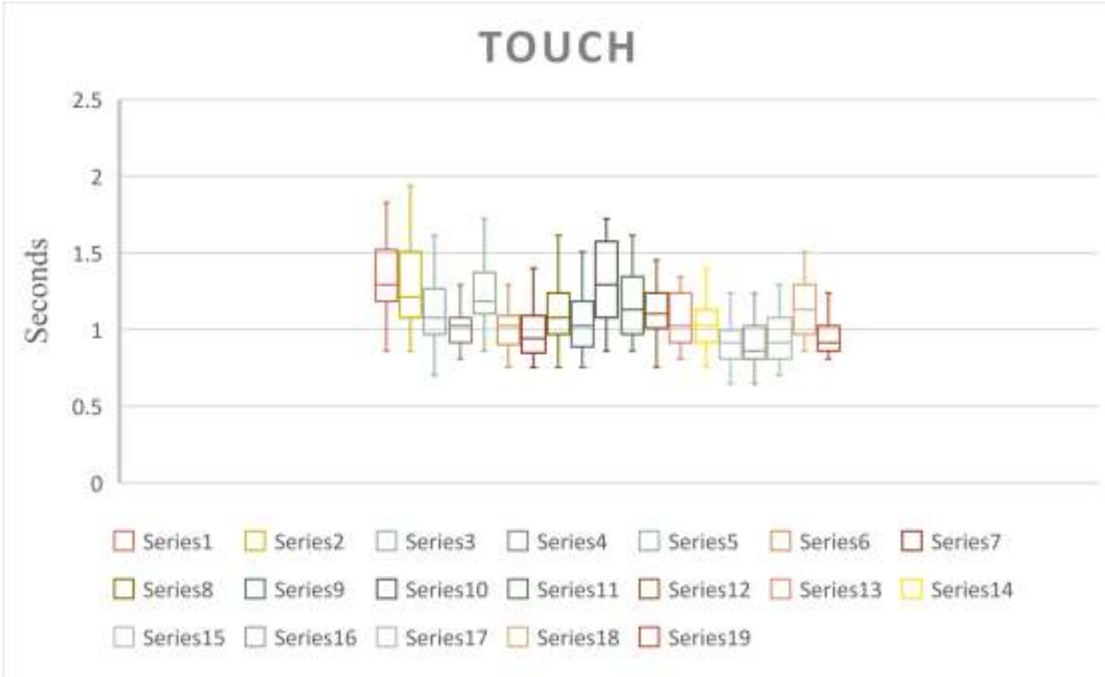

Figure 5.4.2 Touch Box and Whisker Chart

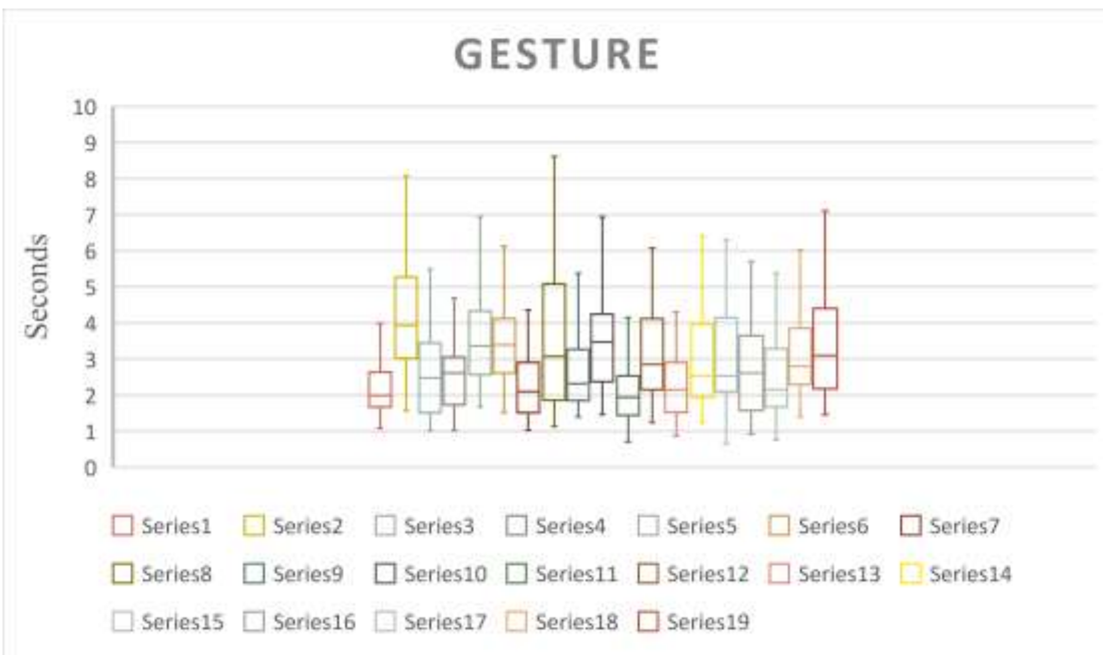

Figure 5.4.3 Touch Box and Whisker Chart

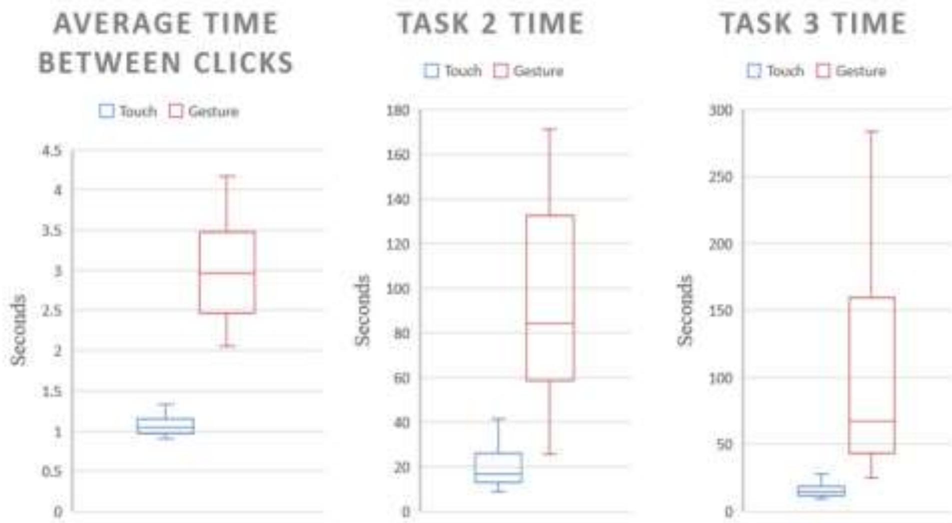

Figure 5.4.4 Box Chart of Times

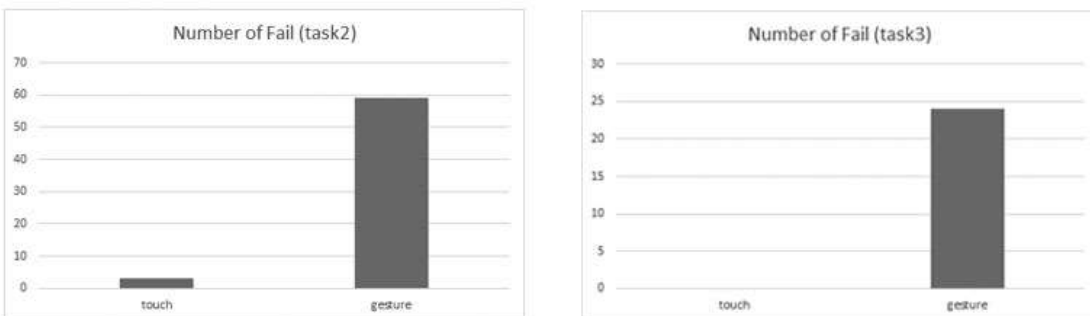

Figure 5.4.5 Failed attempts on jagged and smooth path task

A comment was from a participant that if we had used the front facing camera instead of the back facing camera results maybe better. The reason being was so they could create a visual window or bounds box which in turn could allow for better accuracy and speed. Though the decision to use the back camera was such that the monitor would not be impaired.

CHAPTER 6. CONCLUSION AND FUTURE WORK

This chapter provides a discussion of the results of the data analysis as well as recommendations for future work.

6.1 Conclusion

We failed to reject that the non-touch interaction is intuitive for the data provides insufficient evidence. However, the data provides sufficient data to reject that the gestures felt natural. During the post survey one of the questions asked was disadvantages and advantages of gesture based controls compared to touch. Comments made about the disadvantages were back camera felt unnatural, would feel more natural if front camera could have been used. A few other notable comments were that it would be tiresome so limited time of use, precision is dependent on distance from the camera which meant tablet had to be closer, and difficult to keep a constant steady hand. Some notable advantages mentioned were gestures are more flexible and interactive, users don't have to worry about finger size to screen size, doesn't block vision of the screen, and possible 3d applications.

Through the data collected we fail to reject the gestures were easy to use, but reject that gesture based control was equal to or more efficient as touch. As shown in figure 4.4.1 we have t-value of -7.445 and a p-value of 0.0001. A comment made during the test was though precision and accuracy got better as the hands got further it was uncomfortable and usually meant the screen got closer. The suggestions to switch to the front camera was then made such that the accuracy and precision could be made without having the screen need to be closer to the face. It also allowed for the participants to create an imagery of a window for a bounding box, making it easier to control and less likely to move out of screen.

Applications that participants thought that gesture based control could be useful for fell under a few categories: social networking, games, and virtual reality applications. Some devices that were mentioned were TV, Google Glass, and in general devices with a camera. For TVs the suggestion was the ability to use hand gestures to select, play, or pause movies in place of the remote.

6.2 Future Works

From comments made by participants and through evaluations some future improvements identified were the following. Switching from back camera to front camera though it doesn't sound like much from the comments given it can change the results quite a bit. Being as they have visual confirmations similar to that of touch, arm and hand placement can be more comfortable, as well as being further away allows for better precision and accuracy. Another improvement is different detection algorithm to get rid of markers which can be inconvenient, and tracking in a three dimensional space. These points are worth investigating.

6.3 Final Thoughts

Though the results proved that gesture based controls were not as efficient as touch, it should not dissuade the use of gestures. Gestures provide more forms of creative interactions including that of a three dimensional space. Opening the door for where more research can be done on computer vision, and the interactions with new technology.

LIST OF REFERENCES

LIST OF REFERENCES

Canny, J. (1986). A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6), 679-698.

Dewaele, G., Devernay, F., Horaud, R., & Forbes, F. (2006). The alignment between 3-d data and articulated shapes with bending surfaces. In *Computer Vision–ECCV 2006* (pp. 578-591). Springer Berlin Heidelberg.

Dhawale, P., Masoodian, M., & Rogers, B. (2006, July). Bare-hand 3D gesture input to interactive systems. In *Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI* (pp. 25-32). ACM.

Dorner, B. (1994). *Chasing the colour glove: Visual hand tracking* (Doctoral dissertation, Simon Fraser University).

Glass gestures. (2014, January 1). Retrieved December 3, 2014, from https://support.google.com/glass/answer/3064184?hl=en

Google Developers. (n.d.). Retrieved November 29, 2014, from https://developers.google.com/glass/design/ui

Hodson, H. (2013, November 13). Hacked Google Glass recognizes finger gestures. Retrieved December 3, 2014, from http://www.newscientist.com/article/mg22029434.300-hacked-google-glass-recognises-finger-gestures.html#.VH8CNTHF-ZN

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye–hand coordination in object manipulation. *the Journal of Neuroscience*, *21*(17), 6917-6932.

Lazzari, S., Mottet, D., & Vercher, J. L. (2009). Eye-hand coordination in rhythmical pointing. *Journal of motor behavior*, *41*(4), 294-304.

Lee, C., Ghyme, S., Park, C., & Wohn, K. (1998, November). The control of avatar motion using hand gesture. In *Proceedings of the ACM symposium on Virtual reality software and technology* (pp. 59-65). ACM.

Leinhart, R., & Maydt, J., (2002). An Extended Set of Haar-like Features for Rapid Object Detection. In *International Conference on Image Processing* (pp 900-903). IEEE.

Park, J., & Yoon, Y. L. (2006, November). Led-glove based interactions in multi-modal displays for teleconferencing. In *Artificial Reality and Telexistence--Workshops, 2006. ICAT'06. 16th International Conference on Image Processing* (pp. 395-399). IEEE.

Schlattman, M., & Klein, R. (2007, November). Simultaneous 4 gestures 6 dof real-time two-hand tracking without any markers. In *Proceedings of the 2007 ACM symposium on Virtual reality software and technology* (pp. 39-42). ACM.

Schlömer, T., Poppinga, B., Henze, N., & Boll, S. (2008, February). Gesture recognition with a Wii controller. In *Proceedings of the 2nd international conference on Tangible and embedded interaction* (pp. 11-14). ACM.

Shi, J., & Tomasi, C. (1994, June). Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on* (pp. 593-600). IEEE.

Stenger, B., Thayananthan, A., Torr, P. H., & Cipolla, R. (2006). Model-based hand tracking using a hierarchical bayesian filter. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *28*(9), 1372-1384.

Sudderth, E. B., Mandel, M. I., Freeman, W. T., & Willsky, A. S. (2004). Distributed occlusion reasoning for tracking with nonparametric belief propagation. In *Advances in Neural Information Processing Systems* (pp. 1369-1376).

Tam, V., & Li, L. S. (2012, August). Integrating the kinect camera, gesture recognition and mobile devices for interactive discussion. In *Teaching, Assessment and Learning for Engineering (TALE), 2012 IEEE International Conference on* (pp. H4C-11). IEEE.

Theobalt, C., Albrecht, I., Haber, J., Magnor, M., & Seidel, H. P. (2004, August). Pitching a baseball: tracking high-speed motion with multi-exposure images. In*ACM Transactions on Graphics (TOG)* (Vol. 23, No. 3, pp. 540-547). ACM.

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-511). IEEE.

Wang, R. Y., & Popović, J. (2009, July). Real-time hand-tracking with a color glove. In *ACM Transactions on Graphics (TOG)* (Vol. 28, No. 3, p. 63). ACM.

Welcome to opencv documentation! (n.d.). Retrieved December 4, 2014, from http://docs.opencv.org/

Winkler, C., Hutflesz, P., Holzmann, C., & Rukzio, E. (2012, September). Wall Play: a novel wall/floor interaction concept for mobile projected gaming. In*Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services companion* (pp. 119-124). ACM.

Wolf, K., McGee-Lennon, M., & Brewster, S. (2012, September). A study of on-device gestures. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services companion* (pp. 11-16). ACM.

Woollaston, V. (2014, March 7). Samsung wants to turn your fingers into a KEYBOARD: Thumbs could control virtual keys on your hand when wearing Galaxy Glass. Retrieved December 3, 2014, from http://www.dailymail.co.uk/sciencetech/article-2575805/Samsung-wants-turn-fingers-KEYBOARD-Thumbs-control-virtual-keys-hand-wearing-Galaxy-Glass.html

Yuen, H. K., Princen, J., Illingworth, J., & Kittler, J. (1990). Comparative study of Hough transform methods for circle finding. *Image and Vision Computing*,*8*(1), 71-77.